

На правах рукописи

Дараселия Анастасия Валерьевна

**МОДЕЛИ И АНАЛИЗ ПОКАЗАТЕЛЕЙ ЭФФЕКТИВНОСТИ
МЕХАНИЗМОВ ВЫГРУЗКИ ТРАФИКА В ГЕТОРОГЕННЫХ
БЕСПРОВОДНЫХ СЕТЯХ**

Специальность 1.2.3. Теоретическая информатика, кибернетика

Автореферат
диссертации на соискание ученой степени
кандидата физико-математических наук

Москва – 2023

Работа выполнена на кафедре прикладной информатики и теории вероятностей Российского университета дружбы народов.

Научный руководитель: кандидат физико-математических наук,
доцент, доцент кафедры прикладной информатики
и теории вероятностей Российского университета
дружбы народов
Сопин Эдуард Сергеевич

Официальные оппоненты: **Зейфман Александр Израилевич**, доктор физико-математических наук, профессор, заведующий кафедрой прикладной математики Федерального государственного бюджетного образовательного учреждения высшего образования «Вологодский государственный университет»
Степанов Сергей Николаевич, доктор технических наук, профессор, заведующий кафедрой сетей связи и систем коммутации Ордена Трудового Красного Знамени федерального государственного бюджетного образовательного учреждения высшего образования «Московского технического университета связи и информатики»
Горбунова Анастасия Владимировна, кандидат физико-математических наук, старший научный сотрудник федерального государственного бюджетного учреждения науки Института проблем управления им. В.А. Трапезникова РАН

Защита диссертации состоится 17 февраля 2023 г. 14:00 на заседании диссертационного совета ПДС 0200.006 при Российском университете дружбы народов по адресу: г. Москва, ул. Орджоникидзе, д. 3, ауд. 214.

С диссертацией можно ознакомиться в Научной библиотеке Российского университета дружбы народов по адресу: 117198, г. Москва, ул. Миклухо-Маклая, д. 6. (Отзывы на автореферат просьба направлять по указанному адресу.)

Автореферат разослан «___» _____ 2023 г.

Ученый секретарь
диссертационного совета
канд. физ.-мат. наук, доц.



А.В. Демидова

Общая характеристика работы

Актуальность темы исследования. Концепция выгрузки в инфокоммуникационных сетях становится все более актуальной. Выделяют два основных направления реализации этой концепции. Первым из них является выгрузка мобильных вычислений в распределенную систему туманно-облачных вычислений. Туманные вычисления предполагают размещение вычислительных узлов на границе сети, что является хорошим решением для критичных ко времени отклика ресурсоемких мобильных приложений. Система туманно-облачных вычислений позволяет мобильным устройствам выгрузить некоторые задачи и, следовательно, увеличить время автономной работы. Кроме того, в отличие от классических облачных решений, у туманных вычислений меньшее время отклика благодаря своей близости к конечному пользователю. Однако узлы туманных вычислений не обладают ресурсами облачных дата-центров, поэтому легко перегружаются. В связи с этим задача разработки моделей выгрузки мобильных вычислений в туманно-облачную инфраструктуру для анализа времени отклика является актуальной.

Вторым крупным направлением является выгрузка трафика из лицензированного спектра частот, традиционно используемого мобильными операторами, в нелицензированный для увеличения емкости сети. Впервые такая технология выгрузки была стандартизирована в рамках сетей связи четвертого поколения и недавно была включена в стандарты пятого поколения. Таким образом задача разработки моделей для анализа дополнительной емкости сети, возникающей вследствие использования нелицензированного диапазона частот, является актуальной.

Степень разработанности темы. Анализ вероятностных характеристик выгрузки трафика и мобильных вычислений в гетерогенных беспроводных сетях проведен с помощью аппарата теории вероятностей, теории массового обслуживания, теории случайных процессов, теории телетрафика. К российским ученым, исследователям, внесшим большой вклад в эти области, относятся Г.П. Башарин, П.П. Бочаров, В.М. Вишневский, Ю.В. Гайдамака, А.Н. Дудин, А.И.

Зейфман, А.Е. Кучерявый, Е.А. Кучерявый, А.Н. Моисеев, С.П. Моисеева, Д.А. Молчанов, А.А. Назаров, В.А. Наумов, А.П. Пшеничников, В.В. Рыков, К.Е. Самуйлов, С.Н. Степанов, И.И. Цитович, С.Я. Шоргин и др., а к зарубежным – J.G. Andrews, Luis M Correia, M. Dohler, E. Gelenbe, V.B. Iversen, F.P. Kelly, L. Kleinrock, K.W. Ross и др.

Цели и задачи исследования

Целью диссертационной работы является разработка и анализ моделей для расчета показателей эффективности механизмов выгрузки задач мобильных вычислений и выгрузки трафика в гетерогенных беспроводных сетях.

Для достижения цели в диссертационной работе решаются следующие **задачи**:

1. Построение в виде системы массового обслуживания модели выгрузки в систему туманно-облачных вычислений задач мобильных вычислений, разработка метода расчета функции распределения времени отклика с учетом неоднородности задач по объему вычислений и данных.
2. Построение и анализ двухпараметрической модели в виде ресурсной системы массового обслуживания, описывающей обслуживание трафика в лицензированном диапазоне частот и дискретной цепи Маркова, описывающей процедуру случайного доступа при выгрузке в нелицензированный диапазон частот. Разработка метода расчета распределения скорости передачи в нелицензированном диапазоне.

Научная новизна диссертационной работы:

1. В однопараметрической модели выгрузки задач с мобильных устройств, в отличие от известных, была учтена неоднородность задач по объему необходимых вычислений. Объем вычислений, от которого зависит время обслуживания, полагается случайной величиной с заданным распределением.
2. В модели выгрузки задач мобильных вычислений с двухпараметрическим критерием выгрузки, помимо неоднородности объема вычислений, были

учтены неоднородность объема данных и случайная задержка при выгрузке в узел туманных вычислений.

3. В модели ресурсной СМО, описывающей передачу трафика в лицензированном диапазоне частот, получено распределение требований сброшенных заявок к ресурсу системы. Это позволило вычислить распределение скорости передачи данных в нелицензированных частотах.

Теоретическая и практическая значимость работы. Полученные в диссертационной работе результаты могут быть использованы проектными телекоммуникационными компаниями, операторами сетей связи при планировании сетей радиодоступа для предоставления требуемого качества услуг.

Разработанные математические модели позволяют провести анализ показателей эффективности механизмов выгрузки задач мобильных вычислений и выгрузки трафика в гетерогенных беспроводных сетях.

Результаты работы включены в исследования по грантам РФФИ №19-07-00933 “Стохастические модели и задачи оптимизации для разработки информационных технологий виртуализации и управления ресурсами в беспроводных мультисервисных сетях”, № 19-07-00919 “Исследование задач оптимизации топологии инфокоммуникационных сетей и разработка семейства точных и приближенных алгоритмов решения на основе декомпозиции их математических моделей”, № 20-07-01052 “Вероятностный анализ эффективности механизмов выгрузки задач интернета вещей и мобильных вычислений в туманно-облачную систему беспроводных сетей 5G”.

Методология и методы исследования. В диссертации применяются методы теории массового обслуживания, теории вероятностей, теории случайных процессов и математической теории телетрафика.

Положения, выносимые на защиту.

1. Разработанная модель механизма выгрузки задач с мобильных устройств в туманно-облачную инфраструктуру по пороговому значению объема вычислений позволяет рассчитывать функцию распределения (ФР) времени

отклика. Модель учитывает неоднородность поступающих задач по объему необходимых вычислений.

2. Модель двухпараметрического механизма выгрузки мобильных вычислений в туманно-облачную инфраструктуру учитывает неоднородность задач по объему вычислений и по объему передаваемых данных, а также задержку передачи на беспроводном участке сети. Модель позволяет рассчитывать дискретное распределение времени отклика и математическое ожидание потребления энергии мобильными устройствами.
3. Анализ показателей эффективности выгрузки трафика в нелицензированный диапазон частот беспроводной сети осуществляется при помощи модели, состоящей из двух компонентов. Модель в виде ресурсной системы массового обслуживания (СМО) описывает обслуживание сессий в лицензированном диапазоне и позволяет вычислять долю выгружаемого трафика, а также распределение требований к ресурсу выгружаемых сессий. Модель в виде дискретной цепи Маркова описывает процедуру случайного доступа в нелицензированном диапазоне и позволяет вычислять распределение скорости передачи.

Степень достоверности и апробация результатов

Достоверность и обоснованность полученных результатов следует из строгих математических доказательств, а также из сравнения результатов, полученных с помощью аналитических методов.

Апробация результатов. Основные результаты, изложенные в диссертации, докладывались на научных конференциях и семинарах:

- международная конференция «International Congress on Ultra-Modern Telecommunications and Control Systems and Workshops (ICUMT)» (Россия, Москва, ноябрь 2018);
- международная конференция имени А.Ф. Терпугова «Информационные технологии и математическое моделирование, ИТММ» (Томск, сентябрь 2020 г., декабрь 2021 г.);

- международная конференция «IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)» (Лондон 2020 г.);
- международная конференция по проводным и беспроводным сетям и системам следующего поколения (NEW2AN) (Санкт-Петербург, август 2018 г., август 2019 г.);
- всероссийская конференция с международным участием «Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем (ИТТММ)» (Москва, апрель 2019 г., апрель 2020 г., апрель 2021 г.);
- международная конференция «12th International Workshop on Applied Problems in Theory of Probabilities and Mathematical Statistics (APTP+MS 2018, Summer Session)» (Португалия, Лиссабон, октябрь 2018г.).

Основные результаты опубликованы в статьях ведущих научных журналах – Lecture Notes in Computer Science, IEEE Transactions on Vehicular Technology, Информатика и ее применения, а также в трудах международных конференций, индексируемых WoS (Web of Science) и Scopus.

Реализация результатов работы. Основные научные достижения, полученные в диссертации, использованы в исследовательских мероприятиях РУДН, в исследованиях по грантам РФФИ, в проекте «5-100» повышения конкурентоспособности ведущих российских университетов среди ведущих мировых научно-образовательных центров.

Публикации. Основные результаты по теме диссертационного исследования изложены в 8 печатных изданиях [1-8], в том числе 2 работы в журналах из списка ВАК и 8 в журналах, индексируемых в базах Scopus и WoS.

Содержание работы

Во введении обоснована актуальность темы диссертации, сформированы цели и задачи исследований, представлено содержание основных результатов диссертации, дана характеристика результатов по главам, изложена научная новизна и практическая ценность.

В первой главе рассматриваются основные принципы выгрузки задач мобильных вычислений и трафика в беспроводных сетях, проводится общий обзор технологий и методов анализа, описываются базовые модели выгрузки.

В разделе 1.1 изложены основные особенности выгрузки мобильных вычислений и трафика в беспроводных сетях, приведен краткий обзор литературы в направлении исследования.

В разделе 1.2 построена базовая модель времени отклика с однопараметрическим критерием выгрузки мобильных вычислений в систему туманно-облачных вычислений. Рассмотрена схема выгрузки мобильных вычислений, которая состоит из узла туманных вычислений и удаленной системы облачных вычислений. Получены формулы ФР времени обработки на мобильных устройствах, узлах туманных и облачных вычислений и среднего времени отклика мобильного устройства.

В разделе 1.3 построена базовая модель для анализа эффективности выгрузки трафика в нелицензированный диапазон. Получены формулы вероятности успешной передачи устройств NR, средней скорости передачи, достигаемой устройством NR в нелицензированном диапазоне, и вероятность снижения скорости ниже требуемого порога устройства NR.

Вторая глава посвящена построению и анализу модели выгрузки задач мобильных вычислений в туманно-облачную систему.

В разделе 2.1 разработана расширенная модель с двухпараметрическим критерием выгрузки, который учитывает, как вычислительную сложность через объем вычислений, так и объем данных, которые должны быть переданы в случае выгрузки на узел туманных вычислений. Кроме того, в отличие от базовой модели, также вводится узел беспроводной сети.

Рассматривается распределенная вычислительная система, состоящая из мобильных устройств, узла туманных вычислений и узла облачных вычислений. Мобильные устройства запускают приложения, требующие значительного количества вычислительных ресурсов. Для каждой задачи устройство принимает решение о выгрузке в узел туманных вычислений или обслуживании локально.

Емкость узла туманных вычислений ограничена. Система может быть описана в терминах СМО (Рис. 1).

Пусть существует M устройств, каждое из которых генерирует пуассоновский поток задач с интенсивностью $\lambda_i, 1 \leq i \leq M$. Каждая задача характеризуется требуемым объемом вычислений и объемом данных, которые необходимо передать в случае выгрузки. Пусть объем вычислений и объем данных являются независимыми случайными величинами (с.в.), имеющими гамма-распределение с параметрами k_l и $\delta_l, l=1,2$, с плотностями $w_i(x)$ и $s_i(x)$ и с ФР $W_i(x)$ и $S_i(x)$, соответственно. Мобильные устройства локально обрабатывают задачи в порядке поступления с постоянной скоростью обработки $\mu_i, 1 \leq i \leq M$.

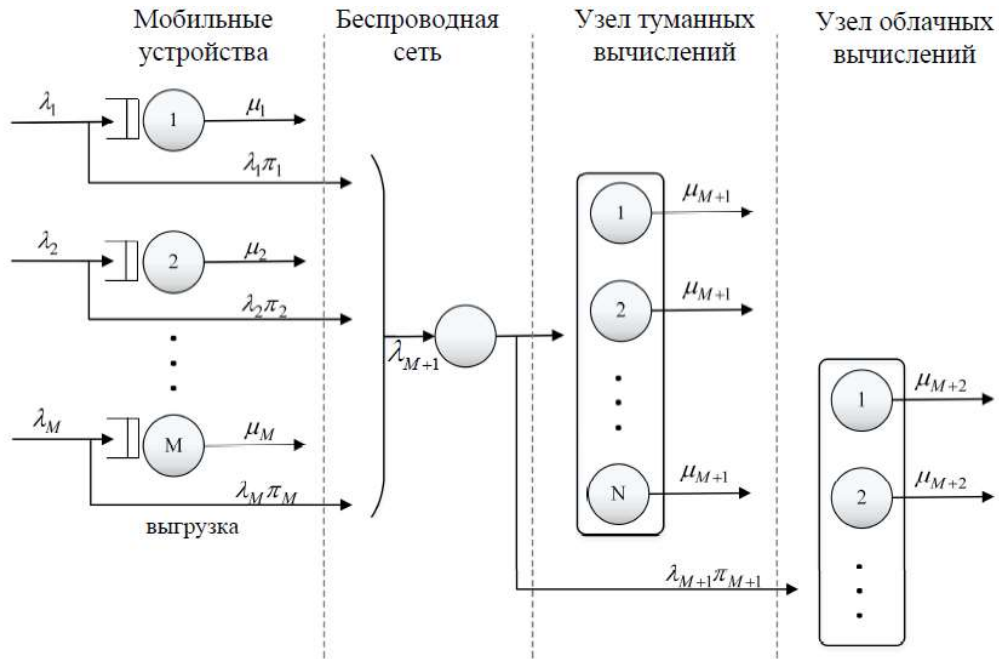


Рис. 1. Схема модели выгрузки мобильных вычислений в терминах СМО.

Предлагаемый механизм выгрузки подразумевает выгрузку задач, являющихся «тяжелыми» с точки зрения объема вычислений и «легкими» с точки зрения объема данных. Разделение на «тяжелые» и «легкие» задачи выполняется пороговыми значениями w^* для объема вычислений и s^* для объема данных. Следовательно, вероятность выгрузки π_i на i -м устройстве имеет вид

$$\pi_i = \int_{w^*}^{\infty} w_i(x) dx \int_0^{s^*} s_i(y) dy = (1 - W_i(w^*)) S_i(s^*). \quad (1)$$

При локальной обработке задачи время отклика равно времени обработки на мобильном устройстве. При выгрузке задачи на узел туманных вычислений общее время отклика является суммой времени передачи задачи в узел туманных вычислений через беспроводную сеть, времени обработки в узле туманных вычислений и времени передачи обратно на мобильное устройство. Если узел туманных вычислений перегружен, задача перенаправляется в удаленный узел облачных вычислений, а время обработки в узле туманных вычислений заменяется временем передачи между узлами туманных и облачных вычислений, временем обработки в узле облачных вычислений и временем передачи назад к узлу туманных вычислений.

Пусть беспроводная сеть обеспечивает общую скорость передачи R , которая используется для передачи данных заявок в порядке поступления, тогда время передачи равно отношению объема данных задачи к скорости передачи R . А время передачи между узлами туманных и облачных вычислений считаем постоянным.

На узел туманных вычислений поступает поток выгружаемых из мобильных устройств задач, являющийся по сути просеянным пуассоновским потоком с интенсивностью λ_{M+1} . Узел туманных вычислений предоставляет вычислительные ресурсы с помощью виртуальных машин (ВМ), каждая из которых имеет постоянную скорость обработки μ_{M+1} . Общее количество ВМ в узле туманных вычислений равно N . Постоянная скорость обработки μ_{M+2} ВМ в узле облачных вычислений больше, чем μ_{M+1} , и предполагается, что количество вычислительных ресурсов, т.е. ВМ, в узле облачных вычислений достаточно велико, чтобы его нельзя было перегрузить.

В разделе 2.2 проведен анализ распределения времени отклика в условиях выгрузки. В модели проводится анализ задержек на каждом из узлов, которые затем используются для вывода распределения времени отклика. Процесс обслуживания в i -м мобильном устройстве моделируется в терминах СМО с интенсивностью поступления входящих заявок λ_i , $\sum_{i=1}^M \lambda_i = \lambda$.

Распределение объема вычислений локально обслуживаемых заявок имеет следующую ФР:

$$V_i(x) = \begin{cases} \frac{W_i(x)}{1 - \pi_i}, & x \leq w^*, \\ \frac{W_i(w^*) + (W_i(x) - W_i(w^*))(1 - S_i(s^*))}{1 - \pi_i}, & x > w^*. \end{cases} \quad (2)$$

ФР времени обработки заявки на мобильном устройстве выражается через ФР (2):
 $T_i(x) = V_i(\mu_i x)$.

Если задача выгружена в распределенную вычислительную инфраструктуру, она сначала передается через беспроводную сеть в узел туманных вычислений. ФР размера передаваемого файла выгруженной задачи имеет вид

$$G_i(x) = \begin{cases} \frac{1}{\pi_i}(1 - W_i(w^*))S_i(x), & x \leq s^*, \\ 1, & x > s^*. \end{cases} \quad (3)$$

ФР времени обслуживания в беспроводной сети задается по формуле
 $D_i(x) = G_i(Rx)$.

В узле туманных вычислений находится N ВМ для обработки выгруженных задач, поэтому процесс обслуживания может быть смоделирован в виде системы, где заблокированные задачи перенаправляются на узел облачных вычислений. Интенсивность поступления задач такая же, как для беспроводной сети - λ_{M+1} . ФР объема вычислений для задач, поступающих с i -го устройства на узел туманных вычислений определяется как

$$V_{M+1,i}(x) = \begin{cases} 0, & x \leq w^*, \\ \frac{1}{\pi_i}(W_i(x) - W_i(w^*))S_i(s^*), & x > w^*. \end{cases} \quad (4)$$

Тогда ФР времени обработки задач в этом узле равна $T_{M+1,i}(x) = V_{M+1,i}(\mu_{M+1}x)$.

Обозначим β_{M+1} - среднее время обработки задачи в узле туманных вычислений, которое можно получить из ФР $T_{M+1,i}(x)$. Вероятность перегрузки, т.е.

вероятность того, что задача будет перенаправлена в узел облачных вычислений вычисляется по формуле Эрланга:

$$\pi_{M+1} = \frac{(\lambda_{M+1}\beta_{M+1})^N}{N!} \left(\sum_{k=0}^N \frac{(\lambda_{M+1}\beta_{M+1})^k}{k!} \right)^{-1}. \quad (5)$$

ФР времени обработки на узле облачных вычислений равна $T_{M+2,i}(x) = V_{M+1,i}(\mu_{M+2}x)$. После обработки в узле облачных вычислений задача возвращается на мобильное устройство с постоянной задержкой передачи между узлами туманных и облачных вычислений и беспроводной сетью. Задача считается обработанной.

Утверждение 1. ПЛС времени обработки задачи в i -м мобильном устройстве имеет вид

$$\tilde{T}_i(s) = \frac{\mu_i}{1 - \pi_i} \left[\frac{1}{(\delta_1 s + \mu_i)^{k_1}} - e^{-\left(\frac{sw^* + w^*}{\mu_i + \delta_1}\right)} \sum_{n=0}^{k_1-1} \frac{(w^*/\mu_i)^n}{n! \delta_1^n (\delta_1 s + \mu_i)^{k_1-n}} \left(1 - e^{-\frac{s^*}{\delta_2} \sum_{n=0}^{k_2-1} \frac{(s^*/\delta_2)^n}{n!}} \right) \right]. \quad (6)$$

ПЛС распределения времени пребывания на i -м устройстве имеет вид

$$\phi_i(s) = \tilde{T}_i(s) \frac{s(1 - \rho_i)}{s - \lambda_i + \lambda_i \tilde{T}_i(s)}, \quad (7)$$

где ρ_i - предложенная нагрузка на i -м мобильном устройстве.

Утверждение 2.

ПЛС распределения времени передачи имеет вид

$$\tilde{D}_i(s) = \frac{R}{1 - \pi_i} \left[\frac{1}{(\delta_2 s + R)^{k_2}} - e^{-\left(\frac{ss^* + s^*}{R + \delta_2}\right)} \sum_{n=0}^{k_2-1} \frac{(s^*/R)^n}{n! \delta_2^n (\delta_2 s + R)^{k_2-n}} \right] \sum_{m=0}^{k_1-1} \frac{(w^*/\delta_1)^m}{m!} e^{-\left(\frac{w^*}{\delta_1}\right)}, \quad (8)$$

где R – скорость передачи в беспроводной сети.

ПЛС $\phi_i(s)$ времени пребывания в беспроводной сети равно

$$\phi_i(s) = \tilde{D}_i(s) \frac{s(1 - \rho_\bullet)}{s - \lambda_{M+1} + \lambda_{M+1} \tilde{D}_i(s)}, \quad (9)$$

где ρ_\bullet - предложенная нагрузка в беспроводной сети, равная произведению λ_{M+1} на средний размер передаваемых данных θ_i , деленное на скорость R .

Утверждение 3. ПЛС распределения времени обслуживания на узле туманных вычислений имеет вид

$$\tilde{T}_{M+1,i}(s) = \frac{\mu_{M+1}}{\pi_i} \left[e^{-\left(\frac{sw^*}{\mu_{M+1}} + \frac{w^*}{\delta_1}\right)} \sum_{n=0}^{k_1-1} \frac{(w^*/\mu_{M+1})^n}{n! \delta_1^n (\delta_1 s + \mu_{M+1})^{k_1-n}} \left(1 - e^{-\frac{s}{\delta_2} \sum_{n=0}^{k_2-1} \frac{(s^*/\delta_2)^n}{n!}} \right) \right]. \quad (10)$$

Утверждение 4. ПЛС распределения времени обслуживания на узле облачных вычислений имеет вид

$$\tilde{T}_{M+2,i}(s) = \frac{\mu_{M+2}}{\pi_i} \left[e^{-\left(\frac{sw^*}{\mu_{M+2}} + \frac{w^*}{\delta_1}\right)} \sum_{n=0}^{k_1-1} \frac{(w^*/\mu_{M+2})^n}{n! \delta_1^n (\delta_1 s + \mu_{M+2})^{k_1-n}} \left(1 - e^{-\frac{s}{\delta_2} \sum_{n=0}^{k_2-1} \frac{(s^*/\delta_2)^n}{n!}} \right) \right]. \quad (11)$$

ПЛС времени передачи между узлом туманных вычислений и узлом облачных вычислений задается выражением $e^{-\Delta_2 s}$, где Δ_2 - постоянное время передачи между узлом туманных и узлом облачных вычислений.

Следствие 1. ПЛС времени отклика имеет вид

$$\tilde{\beta}_i(s) = (1 - \pi_i) \phi_i(s) + \pi_i (1 - \pi_{M+1}) \tilde{T}_{M+1,i}(s) \phi_i^2(s) + \pi_i \pi_{M+1} \tilde{T}_{M+2,i}(s) \phi_i^2(s) e^{-\Delta_2 s}. \quad (12)$$

где Δ_2 - время передачи между узлами туманных и облачных вычислений.

Среднее потребление энергии при локальной обработке задачи на i -м устройстве задается формулой $E_{1,i} = P_{1,i} t_i$, где $P_{1,i}$ – потребляемая мощность во время обработки задачи i -м устройством. Среднее потребление энергии i -го мобильного устройства во время выгрузки задачи составляет $E_{2,i} = P_{2,i} \frac{\theta_i}{R}$, где $P_{2,i}$ – потребляемая мощность при передаче с i -го устройства. Тогда среднее потребление энергии i -м устройством составляет $E_i = (1 - \pi_i) E_{1,i} + \pi_i E_{2,i}$.

Среднее потребление энергии для задачи с произвольного мобильного устройства определяется следующим образом:

$$E = \left(\sum_{j=1}^M \lambda_j \right)^{-1} \sum_{i=1}^M \lambda_i E_i. \quad (13)$$

В разделе 2.3 приведены результаты численного анализа порогов критерия выгрузки. Рассматривается система с $M = 20$ однородными мобильными

устройствами, которые запускают одни и те же приложения, поэтому распределения объема вычислений и размера данных задач также одинаковы. Узлы туманных вычислений могут запускать максимум $N = 8$ виртуальных машин.

На рисунке 2 показана вероятность $B(t^*)$, что время отклика меньше порога t^* для разных пороговых значений объема вычислений W^* . Как видно из графика, при небольших значениях порога объёма данных, $S^* = 0.2$, ничего не выгружается и почти все обрабатывается локально, из-за чего мобильные устройства оказываются перегруженными и с увеличением W^* нагрузка на устройства продолжает расти. При $S^*=0.3$ часть задач выгружается на узел туманных вычислений, из-за чего нагрузка на устройства ниже. С увеличением порога объема данных, $S^*=0.4$, при небольших W^* туманный узел оказывается перегружен, с увеличением порога W^* происходит балансировка нагрузки, и в точке $W^*=1.4$ происходит оптимальное перераспределение нагрузки с точки зрения $B(t^*)$.

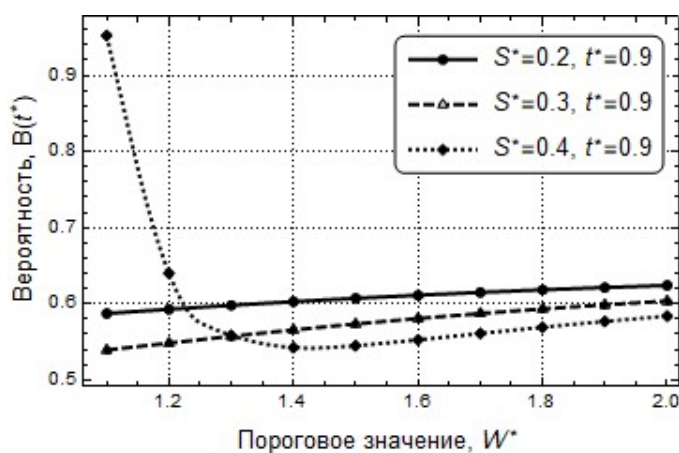


Рис. 2. Вероятность того, что время отклика меньше порога t^* от порогового значения объёма вычислений W^* .

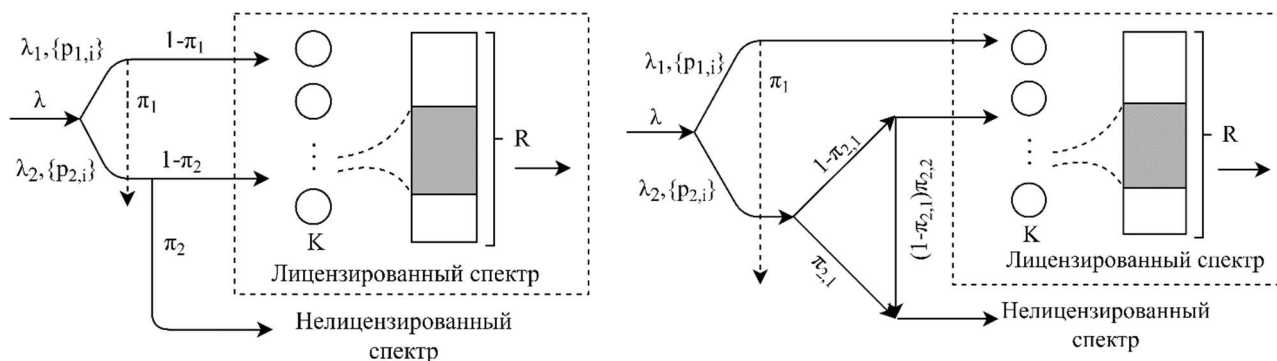
В третьей главе построена и проанализирована двухкомпонентная модель для анализа показателей эффективности выгрузки трафика в нелицензированный диапазон на примере сосуществования сетей New Radio (NR) и WiGig, использующих миллиметровый спектр частот, в котором возникают прерывания связи из-за блокировки пути прямого распространения сигнала.

В разделе 3.1 представлена модель выгрузки трафика с использованием ресурсных СМО, предложены 3 механизма выгрузки трафика на нелицензированный диапазон.

Рассматриваются следующие схемы выгрузки.

- *Базовая выгрузка* (рис. 3.а). Устройства пытаются подключиться к ближайшей базовой станции (БС) и использовать лицензированный диапазон. Если у БС недостаточно ресурсов для обслуживания сессии, она перенаправляется на нелицензированный диапазон.
- *«Тяжелая» выгрузка.* (рис.3.б). Если объем требуемых ресурсов больше порогового значения R_1 , то сессия сразу направляется в нелицензированный диапазон. В противном случае сессия сначала пытается зарезервировать ресурсы в лицензированном диапазоне БС и только при их недостатке выгружается в нелицензированный диапазон.
- *«Легкая» выгрузка.* Если объем требуемых ресурсов меньше порогового значения R_2 , то сессия сразу направляется в нелицензированный диапазон. В противном случае сессия сначала отправляется в лицензированный диапазон БС и только в случае нехватки ресурсов, выгружается в нелицензированный диапазон.

Все три стратегии выгрузки анализируются с использованием одинаковой структуры, состоящей из двух основных компонентов: (а) СМО со случайными требованиями к ресурсам и (б) модель случайного доступа к нелицензированной общей среде.



а) базовая стратегия

б) стратегия «тяжелой» выгрузки

Рис. 3. Модель ресурсной системы массового обслуживания.

Для моделирования процесса обслуживания заявки в лицензированном диапазоне была использована ресурсная СМО с $K < \infty$ приборами и $R < \infty$ единицами ресурса. Учитывая разницу между радиусами покрытия БС в лицензированном и нелицензированном диапазонах, вводятся два типа заявок. Заявки первого типа могут обслуживаться только в лицензированном диапазоне. Когда ресурсов для обработки этой заявки недостаточно, заявка теряется. Заявки второго типа могут быть выгружены в нелицензированный диапазон. В систему поступают заявки двух типов, поступление заявок описывается пуассоновским процессом с интенсивностью поступления λ_1 для первого типа и λ_2 для второго. Распределение времени обработки заявок экспоненциальное с интенсивностью μ .

Для процесса обслуживания каждой заявки требуется прибор и случайное число ресурсов, $0 \leq r \leq R$. Пусть $\{p_{l,j}\}_{j \geq 0}$, $l=1,2$, - распределение требований заявок типа l к ресурсу. Ресурсная СМО с двумя потоками может быть проанализирована как система с одним агрегированным потоком следующим образом:

$$\tilde{p}_{1,j} = \frac{\rho_1}{\rho} p_{1,j} + \frac{\rho_2}{\rho} p_{2,j}, \quad (14)$$

где предложенная нагрузка составляет $\rho = \rho_1 + \rho_2$, $\rho_i = \lambda_i / \mu$, $i=1,2$.

Поступающая заявка принимается, если в системе имеется достаточное количество свободных ресурсов. В противном случае, заявка первого типа теряется, а заявка второго типа перенаправляется на нелицензированный диапазон. По завершении обслуживания заявка покидает систему, освобождая все занятые ресурсы. Поведение системы описывается случайным процессом $X(t) = (\xi(t), \gamma(t))$, где $\xi(t)$ — количество заявок в системе, а $\gamma(t) = (\gamma_1(t), \gamma_2(t), \dots, \gamma_{\xi(t)}(t))$, $\gamma_i(t)$ — вектор количества ресурсов, выделенных для i -й заявки в момент времени t .

Следствие 2. Вероятность π_1 потери заявки первого типа и вероятность π_2 выгрузки заявки второго типа имеет вид

$$\pi_i = 1 - G^{-1}(K, R) \sum_{j=0}^R p_{i,j} G(K-1, R-j), \quad i=1,2, \quad (15)$$

где функция $G(k, r)$ вычисляется рекуррентно следующим образом:

$$G(k, r) = G(k-1, r) + \frac{\tilde{\rho}}{k} \sum_{j=0}^r \tilde{p}_{1,j}, \quad 2 \leq k \leq N, 0 \leq r \leq R, \quad (16)$$

с начальными условиями

$$G(0, r) = 1, r \geq 0, \quad (17)$$

$$G(1, r) = 1 + \tilde{\rho} \sum_{0 \leq j \leq r} \tilde{p}_{1,j}, \quad 0 \leq r \leq R. \quad (18)$$

Утверждение 5. Распределение требований к ресурсам выгружаемых заявок для базовой стратегии имеет вид

$$\tilde{p}_{2,j} = \frac{1}{\pi_2} p_{2,j} \frac{G(K, R) - G(K-1, R-j)}{G(K, R)}, \quad 0 \leq j \leq R. \quad (19)$$

В отличие от базовой модели, вероятность выгрузки заявки второго типа для стратегии «тяжелой» выгрузки (рис. 3.б) равна сумме вероятности $\pi_{2,1}$ того, что заявка была "тяжелой", и вероятности $\pi_{2,2}$ того, что "легкая" заявка не может быть обработана на лицензированном диапазоне и поэтому выгружается на нелицензированный, т.е. $\pi_2 = \pi_{2,1} + (1 - \pi_{2,1})\pi_{2,2}$. Для стратегии «тяжелой» выгрузки вероятность того, что заявка является «тяжелой», вычисляется как $\pi_{2,1} = 1 - \sum_{j=0}^{R_1} p_{2,j}$. Для стратегии «легкой» выгрузки вероятность того, что заявка является «легкой», определяется выражением $\pi_{2,1} = \sum_{j=0}^{R_2} p_{2,j}$. Для стратегий «тяжелой» и «легкой» выгрузки вероятность того, что заявка второго типа потребует j ресурсов в лицензированной полосе, определяются выражением $(1 - \pi_{2,1})^{-1} p_{2,j}$, где $j \in [0, R_1]$ для «тяжелой» выгрузки и $j \in [0, R_2]$ для «легкой». Вероятность того, что заявка, изначально направленная на основе своего «веса», не может быть обслужена в лицензированном диапазоне и выгружается в нелицензированный для «тяжелой» выгрузки, может быть рассчитана по формуле

$$\pi_{2,2} = 1 - (1 - \pi_{2,1})^{-1} G^{-1}(K, R) \sum_{j=0}^{R_1} G(K-1, R-j) p_{2,j}. \quad (20)$$

Для «легкой» выгрузки формула аналогична (20) для интервала $j \in [0, R_2]$.

Распределение требований к ресурсам заявок в агрегированном потоке при стратегиях «тяжелой» и «легкой» выгрузки вычисляется по формуле

$$\tilde{p}_{1,j} = \frac{\rho_1}{\rho^*} p_{1,j} + \frac{\rho_2^*}{\rho^*} (1 - \pi_{2,1})^{-1} p_{2,j}, \quad (21)$$

где предложенная нагрузка $\rho^* = \rho_1 + \rho_2^*$, $\rho_2^* = \lambda_1 (1 - \pi_{2,1}) / \mu$.

Утверждение 6. Вероятность того, что заявка, выгруженная в нелицензированный диапазон, потребует j ресурсов для стратегии «тяжелой» выгрузки, определяется выражением

$$\tilde{p}_{2,j} = \begin{cases} \frac{1 - \pi_{2,1}}{\pi_2} p_{2,j} \frac{G(K, R) - G(K-1, R-j)}{G(K, R)}, & j \leq R_1, \\ \frac{1}{\pi_2} p_{2,j}, & j > R_1. \end{cases} \quad (22)$$

Утверждение 7. Вероятность того, что заявка, выгруженная в нелицензированный диапазон, потребует j ресурсов для стратегии «легкой» выгрузки, определяется выражением

$$\tilde{p}_{2,j} = \begin{cases} \frac{1 - \pi_{2,1}}{\pi_2} p_{2,j} \frac{G(K, R) - G(K-1, R-j)}{G(K, R)}, & j \geq R_2, \\ \frac{1}{\pi_2} p_{2,j}, & j < R_2. \end{cases} \quad (23)$$

Перейдем к оценке эффективности выгрузки на нелицензионном спектре.

Пусть $\zeta(n, m)$ — вероятность коллизии при n устройств NR и m устройств WiGig, а ψ — вероятность блокировки. Тогда вероятность успеха при попытке передачи может быть выражена как

$$\theta = \theta(n, m) = (1 - \zeta(n, m))(1 - \psi). \quad (24)$$

Поведение системы можно описать цепью Маркова $\{X_k, k \geq 0\}$, где $X_k = i$, $i = 0, 1, \dots, T$ и обозначает попытку передачи, когда значение таймера обратного

отсчета находится в интервале $[0, 2^i W - 1]$, и W — минимальное значение конкурентного окна. Граф переходов цепи Маркова представлен на рис.4.

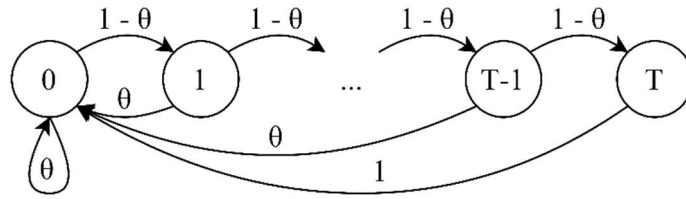


Рис. 4. Граф переходов между состояниями цепи Маркова, показывающих число неудачных попыток передачи.

Утверждение 8. Вероятность $\pi_i^*(n, m)$, $i = 1, 2$, что устройство NR или WiGig совершит попытку передачи в заданном временном такте, и вероятность коллизии $\zeta(n, m)$ являются решением системы уравнений

$$\begin{cases} \pi_i^*(n, m) = \left[\frac{\theta(n, m)W(1 - 2^{T_i+1}(1 - \theta(n, m))^{T_i+1})}{2(1 - (1 - \theta(n, m))^{T_i+1})(2\theta(n, m) - 1)} + \frac{1}{2} \right]^{-1}, & i = 1, 2, \\ \zeta(n, m) = (1 - \psi)(1 - \pi_1^*(n, m))^n (1 - \pi_2^*(n, m))^m. \end{cases} \quad (25)$$

Утверждение 9. Вероятность успешной передачи устройств NR имеет вид

$$P_1^* = \sum_{i=1}^{\infty} \frac{(\rho_1^*)^n}{n!} e^{-\rho_1^*} \sum_{j=0}^{\infty} \frac{(\rho_2^*)^m}{m!} e^{-\rho_2^*} \pi_1^*(n, m) \theta(n, m), \quad (26)$$

где ρ_1^* - нагрузка, вызванная устройствами NR, ρ_2^* - нагрузка, вызванная устройствами WiGig,

Пусть ν — с.в. скорости передачи данных в нелицензированном диапазоне частот, тогда справедливо следующее следствие.

Следствие 3. С.в. скорости ν в нелицензированном диапазоне частот является линейной функцией от с.в. спектральной эффективности η с распределением $\{\tilde{p}_{2,j}\}$:

$$\nu = P_1^* B \eta. \quad (27)$$

Следствие 4. Средняя скорость передачи, достигаемая устройством NR в нелицензированном диапазоне, составляет

$$Mv = \sum_{j=0}^R \tilde{p}_{2,j} \Pi_1^* B s_j. \quad (28)$$

Следствие 5. Вероятность снижения скорости устройства NR ниже требуемого порога имеет вид

$$Q = \pi_2 P\{v < R_{\min}\} = \pi_2 \sum_{j: \Pi_1^* B s_j < R_{\min}} \tilde{p}_{2,j}. \quad (29)$$

В разделе 5.a проведен сравнительный анализ показателей эффективности выгрузки для предложенных трех стратегий. Рассматривалась система с $K=10$ приборов, минимальной скоростью $R_{\min}=50$ Мбит/с и $T=10$ повторных передач.

На Рис. 5.б. показана скорость передачи данных, достигаемая устройством в нелицензируемом диапазоне для всех трех рассматриваемых стратегий. Наибольший прирост скорости достигается при стратегии «легкой» выгрузки в основном для легких нагрузок. При увеличении нагрузки происходит рост вероятности передачи, а вместе с ним рост скорости передачи, так как появляется больше сессий на нелицензированной полосе WiGig за счет переполнения NR, и нелицензированная полоса еще не перегружена. После этого наступает насыщение, и вероятность успешной передачи и скорость передачи начинают падать. Падение сначала медленное, так как коллизии разрешаются более-менее равномерно, и наконец случайный доступ на WiGig не выдерживает нагрузки и параметры начинают резко падать. Наличие такого режима – особенность случайного доступа.

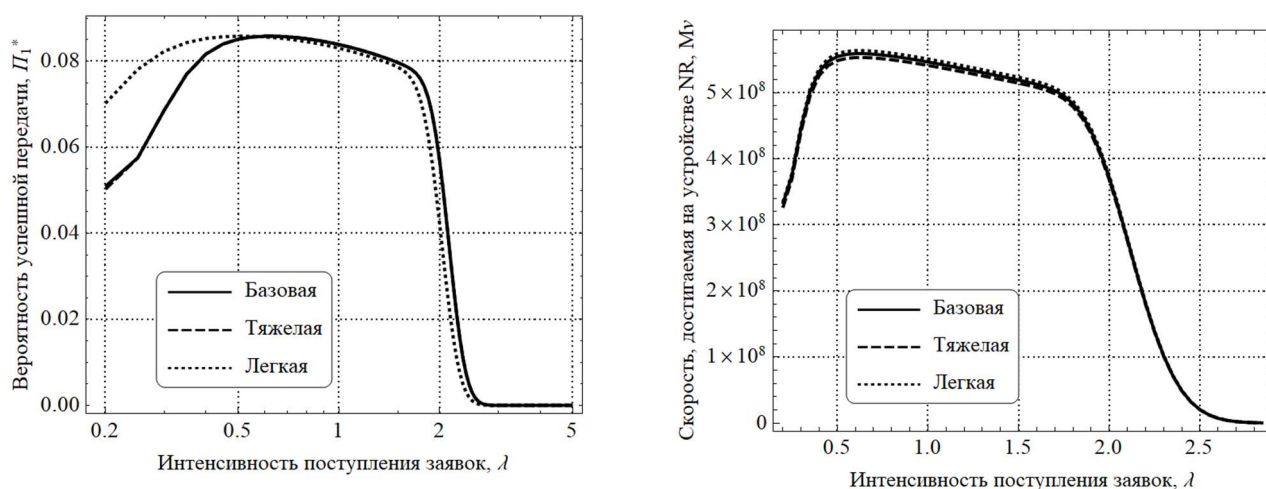


Рис. 5. Вероятность успешной передачи и скорость передачи данных, достигаемая устройством в нелицензируемом диапазоне.

Основные результаты работы

В заключении диссертационной работы сформулируем основные результаты и выводы полученных результатов по сравнению с результатами других исследований.

1. Разработана модель выгрузки задач с мобильных устройств по пороговому значению объема вычислений в туманно-облачную инфраструктуру. Модель в виде сети массового обслуживания содержит три типа узлов, моделирующие обработку задач на мобильных устройствах, узлах туманных и облачных вычислений. Получена формула для функции распределения времени отклика системы.
2. Разработана модель двухпараметрического механизма выгрузки задач мобильных вычислений в туманно-облачную инфраструктуру. Модель учитывает неоднородность задач по объему вычислений и размеру данных для выгрузки. В предположении, что объем вычислений и объем данных имеют гамма-распределение, получено преобразование Лапласа-Стилтьеса функции распределения времени отклика системы. Модель позволяет вычислять энергопотребление мобильных устройств в условиях выгрузки задач.
3. Разработана модель выгрузки трафика из лицензированного диапазона в нелицензированный спектр частот мобильной сети. Модель состоит из двух компонентов – ресурсная система массового обслуживания для анализа передачи трафика в лицензированном диапазоне и цепь Маркова, моделирующая механизм случайного доступа к среде передачи в нелицензированном диапазоне. В первой модели были получены дискретное распределение с.в. требований выгружаемых сессий к ресурсам СМО и вероятность выгрузки заявок в нелицензированный диапазон, во второй модели - дискретное распределение скорости передачи, достигаемой в нелицензированных частотах, с учетом вероятности успешной передачи и дискретного распределения первой модели в терминах ресурсных СМО.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

1. *Sopin, E.S., Daraseliya, A.V., Correia, L.M.* Performance Analysis of the Offloading Scheme in a Fog Computing System // 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). – IEEE. – 2018. – P. 1–5.
2. *Дараселия А.В., Сопин Э.С.* Comparative Analysis of the Mechanisms for Energy Efficiency Improving in Cloud Computing Systems // Lecture Notes in Computer Science. – Vol. 11118. – 2018. – P. 1–9.
3. *Daraseliy A.V., Sopin E.S., Rykov V.V.* On optimization of energy consumption in cloud computing system // 12th International Workshop on Applied Problems in Theory of Probabilities and Mathematical Statistics, Summer Session (APTP+MS'2018). – P. 23–31
4. *Daraseliya A., Korshykov M., Sopin E., Moltchanov D., Koucheryavy Y. and Samouylov K.* Handling Overflow Traffic in Millimeter Wave 5G NR Deployments using NR-U Technology // 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications. – 2020. – P. 1–7.
5. *M.V. Korshykov, A.V. Daraseliya, E.S. Sopin,* Development of Analytical Framework for Evaluation of LTE-LAA Probabilistic Metrics // Lecture Notes in Computer Science. – Vol.11660. – P. 318–328.
6. *Daraseliya A., Korshykov M., Sopin, E., Moltchanov D., Andreev S. and Samouylov K.* Coexistence Analysis of 5G NR Unlicensed and WiGig in Millimeter-Wave Spectrum // IEEE Transactions on Vehicular Technology. – Vol. 70. – №11. – P. 11721–11735.
7. *Дараселия А. В., Сопин Э.С., Молчанов Д.А., Самуйлов К. Е.* Анализ стратегии разгрузки базовых станций 5G NR с помощью технологии NR-U // Информатика и ее применения. –Т. 15. – Вып. 3. – С. 98–111.
8. *Daraseliya A.V., Sopin E.S.,* Optimization of mobile device energy consumption in a fog-based mobile computing offloading mechanism. Discrete and Continuous Models and Applied Computational Science. – Vol. 29. – № 1. – 2021. – P. 53-62.

Дараселия Анастасия Валерьевна (Россия)

Модели и анализ показателей эффективности механизмов выгрузки трафика в гетерогенных беспроводных сетях

Работа посвящена выгрузке в инфокоммуникационных сетях. Рассмотрены два основных направления реализации этой концепции. Первое посвящено выгрузке мобильных вычислений в распределенную систему туманно-облачных вычислений. Построена модель двухпараметрического механизма выгрузки, учитывающая неоднородность задач по объему вычислений и размеру данных для выгрузки. Получено распределение времени отклика системы, проведено вычисление энергопотребления мобильных устройств. Второе направление посвящено выгрузке трафика из лицензированного спектра частот, традиционно используемого мобильными операторами, в нелицензированный для увеличения емкости сети. Разработана модель выгрузки трафика, состоящая из двух компонентов – ресурсная СМО для анализа передачи трафика в лицензированном диапазоне и цепь Маркова, моделирующая механизм случайного доступа к среде передачи в нелицензированном диапазоне. Получены формулы для вероятности потери и распределение скорости передачи данных в нелицензированном спектре.

Daraseliiia Anastasiia (Russia)

Models and analysis of performance indicators of traffic offloading mechanisms in heterogeneous wireless networks

The paper studies the offloading in infocommunication networks. Two main directions for the implementation of this concept are considered. The first is dedicated to offloading mobile computing to a distributed fog-cloud computing system. We study the model of a two-parameter offloading mechanism, which takes into account the heterogeneity of tasks in terms of the processing volume and the data size for offloading. We obtained the distribution of the average system response time, and the energy consumption of mobile devices. The second direction is dedicated to offloading traffic from the licensed frequency spectrum, traditionally used by mobile operators, to the unlicensed one in order to increase network capacity. We study the traffic offloading model that has been developed that consists of two components: a resource QMS for analyzing traffic transmission in a licensed range and a Markov chain that models the mechanism of random access to a transmission medium in an unlicensed range. We obtained formulas for the loss probability and the distribution of the data rate attained by UE in the unlicensed spectrum.