

На правах рукописи

РЫКОВА Татьяна Владимировна

**АНАЛИЗ ПОКАЗАТЕЛЕЙ ЭФФЕКТИВНОСТИ РАСПРЕДЕЛЕНИЯ
РЕСУРСОВ В МОБИЛЬНЫХ СЕТЯХ С ПОМОЩЬЮ ДВУХФАЗНЫХ
СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ**

05.13.17 – теоретические основы информатики

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата физико-математических наук

Москва – 2021

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

Научный руководитель: доктор технических наук, профессор, заведующий кафедрой прикладной информатики и теории вероятностей Российского университета дружбы народов (РУДН)
Самуйлов Константин Евгеньевич.

Официальные оппоненты: доктор физико-математических наук, доцент, заведующий кафедрой информационных технологий Российского университета дружбы народов (РУДН)
Орлов Юрий Николаевич,

доктор физико-математических наук, профессор, заведующий кафедрой прикладной математики Федерального государственного бюджетного образовательного учреждения высшего образования «Вологодский государственный университет»
Зейфман Александр Израилевич,

кандидат физико-математических наук, старший научный сотрудник Федерального государственного бюджетного учреждения науки «Институт проблем управления им. В.А. Трапезникова Российской академии наук» (ИПУ РАН)
Семенова Ольга Валерьевна.

Защита состоится « 24 » декабря 2021 г. в 15 часов 00 минут на заседании диссертационного совета ПДС 0200.001 на базе Российского университета дружбы народов, расположенного по адресу: г. Москва, ул. Орджоникидзе, д. 3, ауд. 219.

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) или на официальном сайте диссоветов РУДН по адресу: <http://dissovet.rudn.ru/>.

Автореферат разослан « ____ » _____ 2021 г.

Ученый секретарь диссертационного совета
ПДС 0200.001, доцент



А.В. Демидова

Общая характеристика работы

Актуальность темы исследования

Рост количества пользователей мобильных сетей и их повышенные потребности в ресурсах беспроводной сети приводят к стремительному развитию телекоммуникационных технологий. 5G — поколение мобильных сетей, работающих в соответствии со стандартами телекоммуникаций, следующих за технологией LTE (Long-Term Evolution), предлагающих еще более высокие скорости передачи за счет использования более широкого спектра частот и повышения спектральной эффективности, а также снижение задержки передачи данных. Одним из методов улучшения спектральной эффективности является переход к гетерогенной сети, суть которой заключается в наличии в сети наряду с традиционными базовыми станциями ретрансляционных станций. Однако, без применения алгоритмов эффективного распределения радио ресурсов между станциями, использование гетерогенных сетей не даст ожидаемого результата, учитывая ограниченность радио ресурсов в сети и возникновение ошибок в результате интерференции.

К повышению пропускной способности в сетях с динамически изменяющимся каналом также относится применение принципов межуровневой оптимизации (CLA, Cross-Layer Adaptation), которые позволяют за счет протокольного взаимодействия обеспечить оптимизированную передачу данных. В частности, исследования, связанные с адаптацией видео контента в зависимости от состояния качества канала и других факторов, являются одним из важных направлений исследований для CLA.

Несмотря на высокий потенциал мобильных сетей пятого поколения, согласно последнему релизу стандарта 5G, до сих пор существует ряд нерешенных задач по обеспечению требуемой низкой задержки, что подчеркивает необходимость поиска новых протокольных решений на пути к услугам сверхнадежных коммуникаций с низкой задержкой (URLLC, Ultra Reliable and Low Latency Communications). Существует несколько стратегий по уменьшению задержки для сервисов URLLC, например, за счет схемы раннего предсказания (e-HARQ, early-Hybrid Automatic Repeat reQuest) с применением искусственного интеллекта. Отметим, что математические модели для перечисленных выше задач практически отсутствуют.

Анализ источников, рекомендаций и стандартов международных организаций, таких как 3GPP (3rd Generation Partnership Project), IEEE (Institute of Electrical and Electronics Engineers), ETSI (European Telecommunications Standards Institute), показал, что необходимы комплексные модели, которые адекватно описывали бы особенности алгоритмов управления доступом и распределением ресурсов сетей 5G.

Особенностью настоящей диссертации является создание нового метода анализа показателей эффективности мобильных сетей с помощью двухфазных моделей массового обслуживания в дискретном времени, позволяющего дать рекомендации по улучшению известных протоколов. Ввиду вышеупомянутого тема работы является актуальной.

Степень разработанности темы

Для анализа вероятностно-временных характеристик (VBX) различных решений в мобильной сети, таких как гетерогенная сеть, CLA и ранняя адаптация e-NARQ, рациональным является применение аналитических моделей в дискретном времени, которые позволяют учитывать как дискретный характер передаваемых данных, так и существенно дискретный характер функционирования реальных сетей; при построении и анализе таких моделей используются методы теории массового обслуживания и математической теории телетрафика. Следует подчеркнуть, что наиболее приемлемыми для исследования мобильных сетей и алгоритмов распределения ресурсов в них являются многофазные СМО вследствие естественной многоэтапности процессов передачи информации. Существенный вклад в исследования таких систем внесли следующие ученые: Г.П. Башарин, П.П. Бочаров, В.М. Вишневский, Ю.В. Гайдамака, Б.С. Гольдштейн, А.И. Громов, И.А. Гудкова, В.А. Ефимушкин, А.И. Зейфман, А.Е. Кучерявый, Е.А. Кучерявый, С.П. Моисеева, Д.А. Молчанов, А.А. Назаров, В.А. Наумов, Ю.Н. Орлов, А.В. Печинкин, А.П. Пшеничников, О.Н. Ромашкова, К.Е. Самуйлов, Б.А. Севастьянов, О.В. Семенова, С.Н. Степанов, А.Д. Харкевич, И.И. Цитович, Bohge M., Capozzi F., Iversen B., Kelly F.P., Kobayashi H., Rappaport S., Shariat M., Berardinelli G., Frederiksen F. и др.

Цели и задачи исследования

Целью диссертационной работы является построение вероятностных моделей в виде двухфазных систем массового обслуживания сложной структуры в дискретном времени для анализа показателей эффективности распределения ресурсов беспроводных сетей. Для достижения цели в диссертационной работе решаются **следующие задачи.**

1. Разработка модели двухфазной СМО в дискретном времени с распределением приборов между фазами для решения задачи повышения пропускной способности беспроводной гетерогенной сети с ретрансляторами данных.

2. Разработка модели двухфазной СМО в дискретном времени для анализа среднего времени пребывания заявки в системе и вероятности потери заявки при решении задачи межуровневой оптимизации – повышения пропускной способности сети и качества восприятия видео потока.

3. Разработка модели двухфазной СМО в дискретном времени, моделирующей механизм предсказания повторной передачи и механизм обратной связи для решения задачи снижения задержки передачи пакета на пользовательскую станцию мобильной сети. Формализация задачи оптимизации, результаты решения которой могут быть использованы как исходные данные в алгоритме предсказания повторной передачи.

Научная новизна работы

1. Для решения задачи повышения пропускной способности соты беспроводной гетерогенной сети предложена модель многопоточной двухфазной СМО в дискретном времени с групповым поступающим потоком и второй фазой сложной структуры, состоящей из параллельных СМО конечной емкости. Отличительной особенностью модели является распределение на каждом такте дискретного времени множества приборов между системами первой и второй фаз.

2. Для решения задачи межуровневой оптимизации – повышения пропускной способности сети и качества восприятия видео потока на пользовательской станции – предложена модель двухфазной СМО в дискретном времени, в которой в отличие от известных входящий поток и обслуживание на фазе 1 управляются цепью Маркова, на фазе 2 применяется обслуживание «с опустошением».

3. Для решения задачи снижения задержки передачи пакета на пользовательскую станцию предложена модель двухфазной СМО в дискретном времени, которая в отличие от известных моделей учитывает результаты итеративного моделирования механизма предсказания повторной передачи и механизма обратной связи.

Теоретическая и практическая значимость работы

Полученные результаты в диссертационной работе могут быть использованы операторами сетей связи при планировании сетей радиодоступа для предоставления услуг требуемого качества. Разработанные математические модели могут быть использованы профильными подразделениями университетов и институтов высшего образования в учебной деятельности, научно-исследовательскими и проектными институтами в практических разработках при расчете и планировании мобильных сетей для получения быстрой оценки показателей эффективности алгоритмов распределения ресурсов в беспроводных сетях без больших затрат времени на имитационное моделирование сети и без применения дорогих экспериментальных установок.

Результаты работы включены в исследования по гранту 19-07-00933 А – «Стохастические модели и задачи оптимизации для разработки информационных технологий виртуализации и управления ресурсами в беспроводных мультисервисных сетях» и внедрены в учебный

процесс – в научно-образовательные курсы «Модели для анализа качества сетей подвижной связи» и «Анализ производительности сетей сотовой подвижной связи», читаемые студентам бакалавриата направлений подготовки 02.03.02 «Фундаментальная информатика и информационные технологии» в РУДН.

Методология и методы исследования

В диссертации применяются методы теории массового обслуживания, теории вероятностей, теории марковских случайных процессов, математической теории телетрафика, теории матриц и имитационного моделирования.

Положения, выносимые на защиту

1. Для решения задачи повышения пропускной способности беспроводной гетерогенной сети с ретрансляторами данных применима предложенная в диссертации модель двухфазной СМО в дискретном времени с распределением приборов между фазами. Для расчета показателей эффективности СМО предложен алгоритм пропорционального распределения фиксированного числа приборов с ограничениями между буферами первой и второй фазы.

2. Для анализа среднего времени пребывания заявки в системе и вероятности потери заявки при решении задачи межуровневой оптимизации – повышения пропускной способности сети и качества восприятия видео потока – применима предложенная в диссертации модель двухфазной СМО в дискретном времени, получены формулы для расчета стационарного распределения и характеристик.

3. Для решения задачи снижения задержки передачи пакета на пользовательскую станцию мобильной сети применима модель двухфазной СМО в дискретном времени, моделирующая механизм предсказания повторной передачи и механизм обратной связи, и формализована задача оптимизации, результаты решения которой могут быть использованы как исходные данные в алгоритме предсказания повторной передачи.

Степень достоверности и апробация результатов

О достоверности результатов диссертации свидетельствует сравнительный анализ расчетов для построенных моделей технических систем с соответствующими вычислительными экспериментами, проведенными на базе близких к реальным исходных данных. Более того, степень достоверности основных результатов, полученных в диссертации, подтверждается их апробацией на научных конференциях и семинарах:

– VII и VIII международные конференции «Finnish-Russian University Cooperation in Telecommunications (FRUCT)» (Санкт-Петербург, 2010 г., Лаппеенранта, 2010 г.).

– Международная конференция «Consumer Communications and Networking Conference» IEEE CCNC (Лас-Вегас, США, 2011).

– V всероссийская конференция (с международным участием) «Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем» ИТММ (Москва, 2011 г.).

– XIV и XVII международные конференции «Distributed Computer and Communication Networks (DCCN)» (Москва, 2011 г., 2013 г.).

– XII Всероссийское совещание по проблемам управления (ВСПУ), (Москва, 2014 г.).

– Научный межвузовский семинар «Современные телекоммуникации и математическая теория телетрафика» (Москва, 2014 г.).

– XXXII International Seminar on Stability Problems for Stochastic Models (Трондхейм, Норвегия, 2014 г.).

– XX International Conference on Next Generation Wired/Wireless Advanced Networks and Systems (NEW2AN, Санкт-Петербург, 2020 г.).

– Семинар кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов (Москва, 2021 г.).

Результаты главы 2 диссертационной работы, представленные автором в заявке «Комплекс моделей, методик и программных средств оптимизации ресурсов в сетях LTE» на Конкурс инноваций и инновационных проектов 2013/2014, проводившийся Международной академией связи при поддержке Московского технического университета связи и информатики и Общественного совета при Федеральном агентстве связи получил диплом лауреата первой степени в номинации «Конкурс концептуальных идей, методик, рекомендаций».

Степень достоверности также подтверждается научными публикациями. Основные результаты по теме диссертационного исследования изложены в 7 печатных работах, из которых [1, 2] – в ведущих рецензируемых изданиях, рекомендованных ВАК РФ, [3-6] – в сборниках материалов конференций, представленных в зарубежных научных изданиях, входящих в Scopus, [7] – российский патент.

Диссертационная работа состоит из введения, трех глав, заключения и списка литературы из 166 наименований на русском и английском языках. Научная работа изложена на 119 страницах. Текст работы иллюстрируется 35 рисунками и 4 таблицами.

Основное содержание работы

Во введении обоснована актуальность темы диссертации, определены цели и задачи исследований, сформулирована теоретическая и практическая ценность работы, представлены выносимые на защиту научные результаты.

В главе 1 ставятся задачи моделирования и оценки показателей эффективности алгоритмов распределения ресурсов, проводится построение и

анализ модели беспроводной гетерогенной сети для решения задачи повышения пропускной способности.

В *разделе 1.1* обсуждаются вопросы распределения ресурсов для передачи информации в мобильных сетях, показывается актуальность разработки моделей функционирования соты беспроводной сети в виде двухфазных СМО в дискретном времени.

В *разделе 1.2* построена и анализируется модель соты гетерогенной беспроводной сети, состоящей из одной базовой станции (БС) и $K, K < \infty$, ретрансляционных (РС), совместно разделяющих частотно-временные ресурсы соты, в виде двухфазной СМО в дискретном времени сложной структуры с буферным накопителем (БН) первой фазы (в БС) конечной емкости, БН второй фазы (в каждой РС) также конечной емкости, разделяемым на каждом такте между БС и РС ограниченным множеством приборов и неординарным потоком заявок на первую фазу. Кадр нисходящего канала разбит на S ресурсных блоков (РБ), рассматриваемых в качестве приборов. Заявка k -го типа (k -заявка) при $k = 0$ должна быть передана пользователю, находящемуся в зоне обслуживания БС, и при $k = 1, 2, \dots, K$ – пользователю в зоне обслуживания k -й РС (далее – РС $_k$). Поступающие на фазу 1 новые заявки и на K СМО фазы 2 заявки со стороны фазы 1 буферизуются в БН фазы 1 и БН этих СМО фазы 2, соответственно. Емкость БН БС (далее – БН $_0$) равна $r_0, r_0 < \infty$, и емкость БН РС $_k$ (далее – БН $_k$) равна $r_k, r_k < \infty, k = 1, 2, \dots, K$. Функционирование системы рассматривается в дискретном времени с тактом постоянной длины h .

В модели рассматривается геометрический групповой поступающий поток $Geom^G$. Период времени между поступлениями групп имеет геометрическое распределение со средним $\frac{1}{a}$. Заявка из поступившей группы принадлежит типу k (k -заявка) с вероятностью $c_k, k = 0, \dots, K, c. = 1$. Здесь точка вместо индекса означает полную сумму переменной по этому индексу. С физической точки зрения будем считать, что 0-заявка соответствует пакету, предназначенному для передачи напрямую UE данной БС, а k -заявка – для передачи в направлении РС $_k, k = 1, 2, \dots, K$. Таким образом, поступающий на СМО поток является $(K + 1)$ -мерным групповым геометрическим. В общем случае в СМО обслуживание заявок может описываться случайным законом распределения, например, геометрическим. Однако в условиях сложной структуры системы при неординарном неоднородном входящем потоке далее в целях упрощения рассматривается детерминированный закон обслуживания заявок с длительностью, равной одному такту, что эквивалентно времени передачи пакета, не превышающему длительности одного РБ. Мнемоническое обозначение для введенной СМО с K системами на второй фазе имеет вид: $Geom_{K+1}^G | D = 1 | s_0 | r_1 < \infty \rightarrow (D = 1 | s_k | r_2 < \infty)^K$.

В разделе 1.3. приведены четыре различных алгоритма распределения ресурсов, включая детерминированные и пропорциональные схемы. Предложен алгоритм пропорционального распределения ресурсов с ограничениями между СМО первой и второй фазы, особенностью которого является выделение базовой станцией на обслуживание k -заявок группы РБ объема, не приводящего к потерям на буферных накопителях РС. Для учета распределения РБ между БС и РС в сети на шаге n введен вектор $\mathbf{s}^n := (s_0^n, s_1^n, \dots, s_{2K}^n)^T$, определяющий распределение приборов, значения которого зависят от состояния $\mathbf{x}$ системы на такте. Пусть $\lfloor y \rfloor$ ($\lceil y \rceil$) - означает округление в сторону наименьшего (наибольшего) целого числа y .

Алгоритм 1. Итерационный алгоритм пропорционального распределения ресурсов с ограничениями между СМО первой и второй фазы для расчета показателей эффективности двухфазной СМО в дискретном времени состоит из следующих шагов:

Шаг 1. Входные параметры:

$$n = 0; \tilde{S}_{(n)} = \tilde{S}, c_k^{(n)} = c_k, s_k^{(n)} = 0, k = 0, 1, \dots, K.$$

$$\text{Здесь } s_{K+k} = \min(x_k, \tilde{S}_k), \tilde{S}_k = S - \sum_{i=0}^{k-1} s_{K+i}, s_K = 0, k = 1, \dots, K, \tilde{S} = S - \sum_{i=1}^K s_{K+i}.$$

Шаг 2. Если $\tilde{S}_{(n)} > 0$ переходим к шагу 3. В противном случае происходит выход из алгоритма.

Шаг 3. $n = n + 1$,

$$s_k^{(n)} = \min \left(\sum_{m=0}^{n-1} \left\lfloor \frac{\lfloor x_0 c_k^{(m)} \rfloor \tilde{S}_{(m)}}{x_0} \right\rfloor, r_k - (x_k - s_{K+k}) \right), k = 1, 2, \dots, K,$$

$$s_0^{(n)} = \sum_{m=0}^{n-1} \left\lfloor \frac{\lfloor x_0 c_0^{(m)} \rfloor \tilde{S}_{(m)}}{x_0} \right\rfloor,$$

$$\tilde{S}_{(n)} = \tilde{S}_{(n-1)} - \sum_{k=0}^K (s_k^{(n)} - s_k^{(n-1)}),$$

$$c_k^{(n)} = \frac{c_k^{(n-1)} (1 - \delta(s_k^{(n)}, r_k - (x_k - s_{K+k})))}{c_{\cdot}^{(n)}},$$

где δ – символ Кронекера,

$$c_{\cdot}^{(n)} = \sum_{k=1}^K \left(c_k^{(n-1)} \left(1 - \delta(s_k^{(n)}, r_k - (x_k - s_{K+k})) \right) \right) + c_0^{(n-1)}.$$

Шаг 4. Чтобы избежать заикливания алгоритма необходимо проверить условие $\tilde{S}_{(n)} \neq \tilde{S}_{(n-1)}$; если оно выполняется, переходим к шагу 2. В противном

случае, все оставшиеся приборы $\tilde{S}_{(n)}$ назначаются на обслуживание заявок 0-типа: $s_0^{(n)} = s_0^{(n)} + \tilde{S}_{(n)}$, и происходит завершение алгоритма.

В разделе 1.4 получена система уравнений равновесия (СУР) для описанной СМО, поведение которой описывается однородной цепью Маркова (ЦМ) ξ_n по моментам $nh + 0, n \geq 0$ над пространством состояний:

$$X = \{x = (x_0, x_1, \dots, x_K): x_k = 0, \dots, r_k, k = 0, \dots, K\}, |X| = \prod_{k=0}^K (r_k + 1), \quad (1)$$

где x_k – число k -заявок, находящихся в БН $_k$. Вследствие непростой структуры исследуемой СМО СУР имеет крайне сложное представление; для ее получения в компактной записи в разделе введен ряд обозначений, используемых также ниже.

Утверждение 1. Система уравнений равновесия для цепи Маркова $\xi_n, n \geq 0$ имеет вид:

$$\begin{aligned} a[0] &= \bar{a} \sum_{\Omega_0} c_0^{x_0} [x], \\ 1 - \sum_{\Omega_1} c_0^{s_0^{\min}} \prod_{k=1}^K c_k^{s_k^{\min} + \delta(x_k, r_k)(1+n_1 \dots + n_k)} \tilde{a}[x] &= \\ \sum_{\Omega_2} c_0^{s_{0,q}^{\min}} \prod_{k=1}^K c_k^{s_k^{\min} - q_k + \delta(x_k, r_k)(1+n_{1,q} \dots + n_{k,q})} \tilde{a}_q [x + \sum_{k=0}^K q_k e_k], \end{aligned} \quad (2)$$

Стационарное распределение вероятностей $[x], x \in X$, находится из (2) и нормировочного условия $\sum_{x \in X} [x] = 1$.

Для решения СУР в работе используется итерационный метод; его применение рационально в данном случае вследствие сильной разреженности матрицы переходных вероятностей ЦМ $\xi_n, n \geq 0$ и сложности получения матрично-рекуррентного решения из-за ее многомерности.

В разделе также получены формулы для основных ВВХ, например, среднее число потерянных заявок на СМО $_k, k = 0, 1, \dots, K$, за такт есть

$$P_0 = \sum_{g=r_0-\tilde{r}_0+1}^{\infty} g a_g \sum_X [x], \quad (3)$$

$$P_k = \sum_{\tilde{s}_k=r_k-x_k+s_k^{\min}}^{n_k} \tilde{s}_k c_k^{\tilde{s}_k} \sum_{s:s_k > r_k-x_k+s_k^{\min}} [x], k = 1, \dots, K, \quad (4)$$

а среднее число потерянных заявок в системе за такт есть $P = P_0$.

В разделе 1.5 проводится численный анализ для четырех алгоритмов распределения РБ в соте беспроводной сети, который показал снижение среднего числа потерянных заявок для случая предложенного алгоритма распределения ресурсов благодаря адаптации к изменениям нагрузки трафика.

В главе 2 предлагается и исследуется двухфазная СМО в дискретном времени с управляемыми цепью Маркова геометрическим поступающим потоком и ординарным обслуживанием на фазе 1, и обслуживанием по геометрическому закону с опустошением на фазе 2.

В разделе 2.1 формулируется задача распределения ресурсов в беспроводной сети на основе межуровневой оптимизации при передаче видео.

В разделе 2.2 описывается и исследуется предложенная двухфазная СМО

$Geom_s | Geom_s | 1 | r_1 < \infty \rightarrow Geom^E(q_2) | 1 | r_2 < \infty$, где на первой фазе моделируется процесс передачи видео на базе технологии DASH (Dynamic Adaptive Streaming over HTTP) от БС к одному пользователю (UE, User Equipment), а на фазе 2 рассматривается процесс видео декодирования на терминале. В каждом временном интервале пользователь производит измерение качества канала, и передает его значение в виде сообщения (CQI, Channel Quality Indicator) на БС. В свою очередь, согласно технологии DASH UE контролирует адаптацию видео контента за счет выбора следующего сегмента на основе измеренного состояния канала и передачи DASH запроса на сервер по протоколу HTTP. Учитывая, что оба сообщения, DASH запрос и CQI, содержат коррелированную информацию о состоянии канала, будем рассматривать в модели передачу только CQI, подразумевая при этом, что значение DASH запроса может быть легко найдено из полученного значения индикатора.

В каждом временном такте значение состояния качества канала $s, s = 1, \dots, S$, где S – общее число его возможных значений, доступно на БС и на сервере HTTP. Изменение переменной s в общем случае моделируется с помощью графа вероятностей переходов, s_{ij} – есть вероятность перехода s из состояния i в j . Межуровневая адаптация обеспечивается за счет изменения интенсивности поступающего потока a_s и обслуживания b_1^s в зависимости от состояния канала s в данном такте.

В разделе 2.3 показан метод нахождения СУР для двухфазной модели с помощью декомпозиции системы и отдельного анализа фаз, при этом при анализе второй фазы используются результаты, полученные при анализе первой фазы, что позволяет снизить вычислительную сложность задачи. Функционирование первой фазы описывается однородной ЦМ $\zeta_n^1 = (\xi_n, \psi_n)$ по моментам $nh + 0, n \geq 0$, над пространством состояний

$$X^1 = \bigcup_{q_1=0}^{r_1} X_{q_1}^1, X_{q_1}^1 = \{(q_1, s), s = 1, \dots, S\}. \quad (5)$$

Функционирование второй фазы описывается однородной ЦМ $\zeta_n^2 = (\eta_n)$ по моментам $nh + 0, n \geq 0$, над пространством состояний:

$$X^2 = \{(q_2), q_2 = 0, \dots, r_2 - 1\}. \quad (6)$$

Утверждение 2.

1. Для двухфазной СМО, моделирующей процессы передачи видео и его декодирования на терминале, стационарное распределение вероятностей $\mathbf{p}^T = (\mathbf{p}_0^T, \mathbf{p}_1^T, \dots, \mathbf{p}_{r_1}^T)$, $\mathbf{p}_{q_1}^T = (p_{q_1 1}, p_{q_1 2}, \dots, p_{q_1 S})$ для первой фазы имеет следующее рекуррентное представление:

$$\mathbf{p}_m^T = \mathbf{p}_{m-1}^T \mathbf{W}_{m-1}, m = 1, \dots, r_1, \quad (7)$$

$$\mathbf{W}_{m-1} = -\tilde{\mathbf{A}}_{m-1m} \left(\tilde{\mathbf{A}}_{mm} + \mathbf{W}_m \tilde{\mathbf{A}}_{m+1m} \right)^{-1}, m = 1, \dots, r_1 - 1,$$

$$\mathbf{W}_{r_1-1} = -\tilde{\mathbf{A}}_{r_1-1r_1} \left(\tilde{\mathbf{A}}_{r_1r_1} \right)^{-1}, \quad (8)$$

где $\tilde{\mathbf{A}}$ – трехдиагональная клеточная матрица порядка $r_i + 1$ и вектор $\mathbf{p}_0$ определяется из системы уравнений $\mathbf{p}_0^T \tilde{\mathbf{W}} = \mathbf{e}_S^T$. Матрица $\tilde{\mathbf{W}}$ образуется заменой последнего столбца матрицы $\mathbf{W} = \tilde{\mathbf{A}}_{00} + \mathbf{W}_0 \tilde{\mathbf{A}}_{10}$ на вектор $\sum_{q_1=-1}^{r_1-1} (\prod_{m=0}^{q_1} \mathbf{W}_m) \mathbf{1}$, причем $\prod_{m=0}^{-1} \mathbf{W}_m = \mathbf{I}$, а $\mathbf{e}_S^T = (0, \dots, 0, 1)$.

2. Стационарное распределение вероятностей $\mathbf{g}^T = (g_0, g_1, \dots, g_{r_2-1})$ для второй фазы имеет следующий вид:

$$g_m = \prod_{i=1}^m \bar{b}_2^i g_0, m = 1, \dots, r_2 - 1, g_0 = [1 + \sum_{i=1}^{r_2-1} \prod_{j=1}^i \bar{b}_2^j]^{-1}. \quad (9)$$

3. Стационарное распределение вероятностей $\hat{\mathbf{p}}$ системы в целом имеет мультипликативный вид:

$$\hat{p}_{q_1 q_2 s} = p_{q_1 s} g_{q_2}, q_1 = 0, \dots, r_1, q_2 = 0, \dots, r_2 - 1, s = 1, \dots, S, \quad (10)$$

и вычисляется последовательно на основе полученных распределений $\mathbf{p}$ и $\mathbf{g}$ для первой и второй фазы, соответственно.

В разделе 2.4 получены выражения для ВВХ, отдельно для каждой из фаз и для системы в целом. Для проведения численных экспериментов был введен вектор вероятностей обслуживания, который определяется функциональной зависимостью от параметра γ , $\mathbf{b}_1 = \mathbf{x}^\gamma$, где $\mathbf{x}^T = (x_1, x_2, \dots, x_S)$ – вектор равномерно распределенных значений, $0 < x_i < 1, x_i < x_{i+1}, i = 1, \dots, S - 1$ и γ – параметр, управляющий формой кривой $\mathbf{b}_1$.

В экспериментальной части был проведен анализ показателей эффективности распределения ресурсов: вероятности потери заявок на первой фазе в состоянии s для двух сценариев местности (городская, сельская), среднего

времени пребывания заявки в СМО в состоянии s сообщения CQI, среднего числа заявок, покинувших систему из-за истечения срока их действия.

В главе 3 проведены построение и анализ двухфазной СМО в дискретном времени, в которой на первой фазе моделируется процесс предсказания повторной передачи на базе e-HARQ с возможностью ретрансляции в случае квитации NACK (Non-ACKnowledgement), а на второй фазе рассматривается процесс обработки сообщения терминалом пользователя на базе HARQ.

В разделе 3.1 формулируется задача ранней адаптации канала в беспроводной сети на основе предсказания повторной передачи e-HARQ.

В разделе 3.2 исследуется двухфазная СМО в дискретном времени с тактом h постоянной длины, равным длительности одного OFDM символа, что позволяет учесть дискретный характер физических процессов передачи сообщения в сети 5G. Будем полагать, что пакеты имеют одинаковую длину, равную длительности N OFDM символов, или микрослотов. В модели используется понятие заявки, имеющей физическое значение пакета. Заявка поступает в систему с вероятностью a , $0 < a \leq 1$, и, пока она находится в СМО, поступление новой заявки не происходит.

Если предиктор e-HARQ принял положительное решение ACK на первой фазе, что соответствует событию $\{eHARQ = ACK\}$, то заявка переходит на вторую фазу с вероятностью $b_1(l)$, где l – текущее число попыток передачи данного пакета на UE. В противном случае, если решение предиктора e-HARQ, отправленное на БС – NACK, т.е. $\{eHARQ = NACK\}$, заявка будет ретранслирована с дополнительной вероятностью $\bar{b}_1(l)$. Однако, обработка предыдущей заявки, которой предиктор e-HARQ вынес отрицательное заключение, продолжается на второй фазе с вероятностью, равной 1.

Вероятности $b_2(l)$ и $b_3(l)$ успешного обслуживания на второй фазе, где l – текущее число попыток передачи данной заявки на UE, могут быть представлены в следующем виде:

$$b_2(l) := P(\{HARQ = ACK\}|\{eHARQ = ACK\}), \quad (11)$$

$$b_3(l) := P(\{HARQ = ACK\}|\{eHARQ = NACK\}), \quad (12)$$

где $\{HARQ = ACK\}$ – событие, при котором механизм обратной связи на второй фазе принимает положительное решение ACK.

Если отправка заявки на UE превышает максимальное число микрослотов T_N , где N – длина заявки, или максимальное число возможных передач $T_c = \left\lfloor \frac{T_N}{N} \right\rfloor$, то заявка считается потерянной и покидает систему.

В разделе 3.3 представлены СУР и основные ВВХ функционирования данной СМО. Поведение СМО описывается однородной ЦМ ξ_n по моментам $nh + 0, n \geq 0$, над пространством состояний $X = X_m^* \cup X_t$, где X_m^* определяет множество основных состояний системы, в которых происходят активные события:

$$X_m^* = \left\{ (0,0,0,0)^*, (1,1,1,t)^*, (s,l,v,t)^* : s \in \{1,2,3\}, l = \left\lfloor \frac{v}{2} \right\rfloor + u(s,v), \dots, v-1, \right. \\ \left. v = 2, \dots, T_c, t \leq T_N \right\}, \quad (13)$$

где v – номер временного слота и $s = 0,1,2,3$ – параметр, соответствующий процедуре обработки на фазах, причем положение единицы в бинарной записи соответствует состоянию функционирования данной фазы, и, соответственно, e- HARQ и/или HARQ механизмов, и t – номер микрослота, равного одному OFDM символу.

Множество переходных состояний X_t , позволяющих попасть в X_m^* определяется следующий образом:

$$X_t = \{ \{ (s-1, l-1, v-1, t-j) : s = 1 \}, \{ \gamma(l,v)(s, l, v-1, t-i) : s \in \{2,3\} \}, \{ \delta(l,v)(s+2, l-1, v-1, t-j) : s \in \{2,3\} \}, l, v, t \in X_m^*, i = 1, \dots, N-d-1, j = 1, \dots, N+d-1 \}, \quad (14)$$

где $0 \leq d < N$ – временной сдвиг относительно окончания передачи заявки, измеряемый в OFDM символах, и

$$\gamma(l,v) = \begin{cases} 1, \text{ если } \{l < v-1\} \cup \{v = 2\}, \\ 0, \text{ в противном случае,} \end{cases} \quad (15)$$

$$\delta(l,v) = \begin{cases} 1, \text{ если } \left\{ l > \left\lfloor \frac{v}{2} \right\rfloor, v = 2n \right\} \cup \{v = 2n+1\}, n = 1, 2, \dots \\ 0, \text{ в противном случае.} \end{cases} \quad (16)$$

Стационарное распределение вероятностей $[s, l, v, t]$, $(s, l, v, t) \in X$, находится из полученной СУР и нормировочного условия $\sum_{(s,l,v,t) \in X} [s, l, v, t] = 1$ решением алгебраических уравнений. Данная аналитическая модель позволяет получить формулы на базе стационарного распределения для FN (False Negative) и FP (False Positive) вероятностей:

$$P_{FN} = \sum_{v=2}^{T_c} \sum_{l=\frac{v}{2}}^{v-1} b_3(l) [3, l, v, t], \quad (17)$$

$$P_{FP} = \sum_{v=2}^{T_c} \sum_{l=\frac{v}{2}}^{v-1} \bar{b}_2(l) [2, l, v, t]. \quad (18)$$

В разделе 3.4 проводится численный анализ показателей эффективности данной модели. На первом этапе получены вероятности переходов $b_1(l)$, $b_2(l)$ и $b_3(l)$ за счет моделирования канального уровня сети 5G, а также разработан имитационный комплекс для проверки корректности аналитической модели,

основанный на принципах дискретно-событийного моделирования. Проводится обширный сравнительный анализ ВВХ с различными значениями отношения сигнал/шум (5.0 dB–12.0 dB), длины пакета N , равного 4,5,...,12 OFDM символам, длинами кодов с соответствующими сдвигами $d = 0,2, \dots, 10$ и пороговыми значениями параметра $\theta = 0, \dots, 9$, используемого в процедуре прогнозирования.

Большинство известных схем предсказания повторной передачи используют фиксированное пороговое значение θ в своих алгоритмах в качестве способа классификации между принятием ACK/NACK решения, поэтому выбор данного значения θ существенным образом влияет на эффективность схемы предсказания.

Утверждение 3. Значение параметра θ может быть найдено путем решения задачи минимизации целевой функции затрат:

$$\min(w_1 P_{FN}(\theta) + w_2 T^*(\theta)), \quad (19)$$

с ограничениями $w_1 + w_2 = 1$ и $P_{FN}(\theta) < FN_{max}$, где w_1 и w_2 – весовые коэффициенты, и $T^* = \frac{T}{T_N}$ – среднее время пребывания в системе до успешной передачи, нормированное относительно числа микрослотов T_N .

Утверждение 4. Значения параметров N, d, θ для обеспечения эффективной передачи пакета могут быть найдены путем решения задачи максимизации выигрыша, получаемого при применении схемы с предсказанием в сравнении со схемой без предсказания:

$$\max G(T(N, d, \theta), T_{HARQ}(N, d, \theta)), \quad (20)$$

с ограничениями $P_{FN}(N, d, \theta) < FN_{max}$, $P_{FP}(N, d, \theta) < FP_{max}$.

Здесь $G(x, y) = \frac{|x-y|}{y} \cdot 100\%$, $FN_{max} = \max(P(\{\text{HARQ} = \text{ACK}\}|\{\text{eHARQ} = \text{NACK}\}))$, $FP_{max} = \max(P(\{\text{HARQ} = \text{NACK}\}|\{\text{eHARQ} = \text{ACK}\}))$.

Данные задачи решены методом прямого поиска. На рис.1 показан график выигрыша среднего времени пребывания в системе при применении схемы с предсказанием относительно системы без предсказания. Как видно из графика, ограничивая вероятность FP до 0.1, можно получить выигрыш в 20% при $FN_{max} = 0.1$, и выигрыш около 7.5% при $FN_{max} = 0.001$.

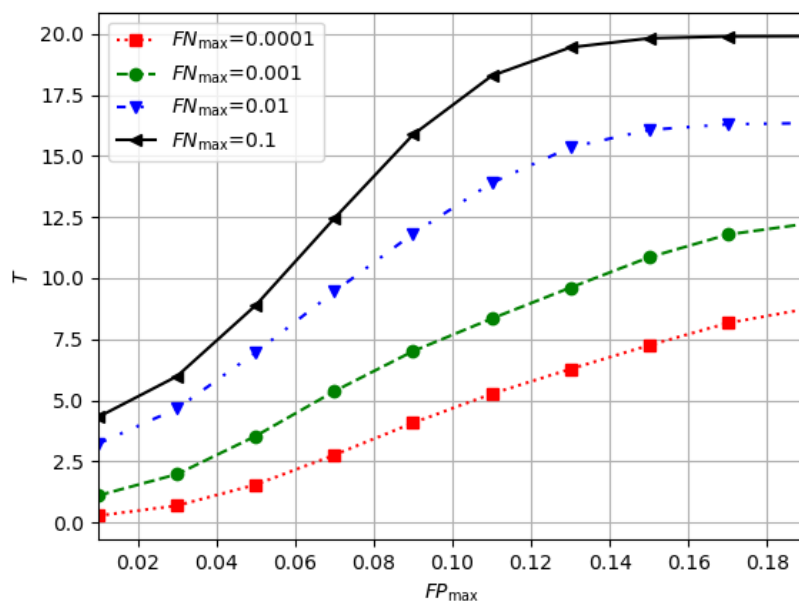


Рис.1 Выигрыш (в %) среднего времени пребывания в системе при применении схемы с предсказанием относительно системы без предсказания.

Заключение

Ниже приведены основные выводы и результаты диссертационной работы.

1. Разработана модель двухфазной СМО в дискретном времени для решения задачи повышения пропускной способности соты беспроводной гетерогенной сети, позволяющая учитывать различные алгоритмы распределения ресурсов между фазами. Разработан пропорциональный алгоритм распределения ресурсов с ограничениями. Получен метод вычисления основных вероятностных характеристик. Численный анализ показал снижение среднего числа потерянных заявок вследствие адаптации предложенного алгоритма распределения ресурсов к изменениям нагрузки трафика.

2. Разработана модель двухфазной СМО в дискретном времени для оценки показателей эффективности распределения ресурсов при решении задачи межуровневой оптимизации – повышения пропускной способности сети и качества восприятия видео потока на пользовательской станции. Получено матрично-рекуррентное решение для стационарного распределения цепи Маркова, описывающей функционирование модели.

3. Для решения задачи снижения задержки передачи пакета на пользовательскую станцию разработана модель двухфазной СМО в дискретном времени, позволяющая произвести оценку среднего времени пребывания заявки в системе и других показателей эффективности путем итеративного моделирования механизма предсказания повторной передачи и механизма

обратной связи. Разработана событийная имитационная модель двухфазной СМО для оценки точности аналитической модели.

4. Формализована и решена задача оптимизации длин пакетов, показателей качества предсказания, а также пороговых значений, при которых наблюдается наименьшая длительность обслуживания при условии, что ошибка предсказания не превышает допустимых значений. Задача решена методом прямого поиска. Результаты решения задачи оптимизации использованы как исходные данные в алгоритме предсказания повторной передачи, что дает возможность снизить задержку при передаче пакета пользователю.

В качестве перспективного направления дальнейшей разработки темы может быть рекомендовано исследование многофазной СМО с числом фаз более двух, с распределяемым между фазами множеством общих приборов в дискретном времени, с буферными накопителями конечной и бесконечной емкости на фазах. Данная СМО может использоваться, в частности, в качестве модели взаимодействия базовой и ретрансляционных станций в сети гетерогенной сети в целях анализа различных схем распределения ресурсных блоков между ними.

Список работ, опубликованных по теме диссертации

Статьи в рецензируемых изданиях, рекомендованных ВАК РФ:

1. Rykova T., Algorithmic calculation of stationary distribution for multiphase queue with common for phases servers in discrete time // T-Comm – Телекоммуникации и Транспорт, 2017. – №12. – С.71–76.
2. Rykova T. Towards the analysis of the performance measures of heterogeneous networks by means of two-phase queueing systems // Discrete and Continuous Models and Applied Computational Science, 2021. – №29 (3). – Pp. 242–250.

Публикации в сборниках материалов конференций, представленных в зарубежных научных изданиях, входящих в Scopus:

3. Rykova T. Marginal loss probabilities at phases for multiphase queue with common for phases servers in discrete time // In: Proc. of 2018 Systems of Signals Generating and Processing in the Field of on Board Communications, 2018. – Pp.1-10. – DOI 10.1109/SOSG.2018.8350629.
4. Rykova T., Göktepe B., Schierl T., Hellge C. Analytical Model of Early HARQ Feedback Prediction // In: Lecture Notes in Computer Science. – Springer International Publishing, 2020. – Pp.222–239. – DOI 10.1007/978-3-030-65729-1_20.
5. Göktepe B., Rykova T., Fehrenbach T., Schierl T., Hellge C. Feedback Prediction for Proactive HARQ in the Context of Industrial Internet of Things //

Proc. of the IEEE Globecom, Taipei, Taiwan, December 2020. arXiv: 2009.06301v1 [eess.SP] 14 Sep 2020. – 7 p. – DOI 10.1109/GLOBECOM42002.2020.9322302.

6. Rykova T., Göktepe B., Schierl T., Samuylov K., Hellge C. Analytical Model and Feedback Predictor Optimization for Combined Early-HARQ and HARQ // Mathematics 2021, 9, 2104. – DOI 10.3390/math9172104.

Прочие публикации:

7. Рыкова Т., Хелльге К., Санчес Я., Ширль Т., Хауштайн Т., Тиле Л., Вирт Т., Куррас М., Рашковский Л. Передача сигнала данных в системе беспроводной связи с уменьшенной сквозной задержкой // Заявка №2019124612/07(047981). Патентообладатель: ФРАУНХОФЕР-ГЕЗЕЛЛЬШАФТ ЦУР ФЕРДЕРУНГ ДЕР АНГЕВАНДТЕН ФОРШУНГ Е.Ф. – Дата подачи заявки: 02.08.2019 (Российская Федерация).

Рыкова Татьяна Владимировна (Россия)

Анализ показателей эффективности распределения ресурсов в мобильных сетях с помощью двухфазных систем массового обслуживания

В диссертации разработана и проанализирована модель двухфазной СМО в дискретном времени с групповым поступающим потоком и второй фазой сложной структуры, состоящей из параллельных СМО конечной емкости. Предложен алгоритм пропорционального распределения ресурсов с ограничениями между СМО первой и второй фаз. Для решения задачи межуровневой оптимизации – повышения пропускной способности сети и качества восприятия видео потока на пользовательской станции предложена модель двухфазной СМО в дискретном времени, в которой входящий поток и обслуживание на фазе 1 управляются цепью Маркова, на фазе 2 применяется обслуживание «с опустошением». Для решения задачи снижения задержки передачи пакета на пользовательскую станцию предложена модель двухфазной СМО в дискретном времени, которая учитывает итеративное моделирование механизма предсказания повторной передачи и механизма обратной связи. Проведен численный анализ показателей эффективности распределения ресурсов и решена задача оптимизации.

Rykova Tatiana Vladimirovna (Russia)

Performance measures analysis of resources allocation in mobile networks by means of two-phase tandem queues

In the thesis, a two-phase tandem queue with a group arrival flow of requests and a second phase of the complex structure that consists of parallel finite queues is proposed and investigated in discrete time. A proportional resource allocation algorithm with constraints on the queues in the system is derived. A two-phase tandem queue with Markov modulated geometric arrival and service processes at the first phase and exhaustive service process at the second phase is introduced and investigated in order to solve a cross-layer adaption problem by increasing throughput and quality of enhancement of the received video. A two-phase tandem queue in discrete time that evaluates the behavior of early-HARQ prediction and traditional HARQ approach is proposed in order to study the problem of latency reduction. The extensive numerical analysis, including imitation and a search of optimal parameters, is conducted.