

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ»

На правах рукописи

Пастушок Игорь Анатольевич

**ОПТИМИЗАЦИЯ ПЕРЕДАЧИ ВИДЕОДАНЫХ В
МОБИЛЬНЫХ СЕТЯХ**

Специальность 05.12.13 —
«Системы, сети и устройства телекоммуникаций»

Диссертация на соискание учёной степени
кандидата технических наук

Научный руководитель:
доктор технических наук, доцент
Тюрликов Андрей Михайлович

Санкт-Петербург — 2018

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	5
1 ОСОБЕННОСТИ ПЕРЕДАЧИ ВИДЕОДАНЫХ В ТЕЛЕКОММУНИКАЦИОННЫХ СЕТЯХ	10
1.1 Вводные замечания	10
1.2 Структура систем передачи видеоданных	12
1.3 Представление видеоданных при передаче через телекоммуникационные сети	14
1.4 Организация передачи видео через телекоммуникационные сети	16
1.5 Методики оценки качества передачи видеоконтента	28
1.6 О взаимосвязи объективных показателей производительности сети и оценки качества восприятия видео	35
1.7 Выводы по разделу	41
2 ПЕРЕДАЧА ВИДЕОДАНЫХ В БЕСПРОВОДНЫХ ЦЕНТРАЛИЗОВАННЫХ СЕТЯХ	43
2.1 Вводные замечания	43
2.2 Структура современных беспроводных централизованных сетей передачи информации	44
2.3 Общие положения аналитической модели	45
2.4 Модель беспроводного канала передачи информации	49
2.5 Распределение ресурсов беспроводного канала связи	52
2.6 Модель воспроизводящего устройства	57
2.7 Система допущений для модели передачи видеоданных	59
2.8 Система передачи видеоданных как система массового обслуживания	61
2.9 О взаимосвязи параметров системы передачи видеоданных	63
2.10 Выводы по разделу	67
3 ОПТИМИЗАЦИЯ ПЕРЕДАЧИ НЕАДАПТИВНЫХ ВИДЕОПОТОКОВ В БЕСПРОВОДНЫХ ЦЕНТРАЛИЗОВАННЫХ СЕТЯХ	69

3.1	Вводные замечания	69
3.2	Критерий качества восприятия неадаптивного видеопотока	70
3.3	Постановка оптимизационной задачи	72
3.4	Решение обобщенной задачи о непрерывном рюкзаке	74
3.5	Нижняя граница нормированного отношения длительности буферизации к просмотру при передаче неадаптивных видеопотоков	86
3.6	Алгоритм планирования распределения ресурсов для нормированного отношения длительности буферизации к просмотру при передаче неадаптивных видеопотоков	90
3.7	Численный пример	100
3.8	Выводы по разделу	104
4	ОПТИМИЗАЦИЯ ПЕРЕДАЧИ АДАПТИВНЫХ ВИДЕОПОТОКОВ В БЕСПРОВОДНЫХ ЦЕНТРАЛИЗОВАННЫХ СЕТЯХ	106
4.1	Вводные замечания	106
4.2	Критерии качества восприятия адаптивного видеопотока	107
4.3	Постановка оптимизационной задачи	110
4.4	Решение вспомогательной оптимизационной задачи выпуклого программирования на основе теоремы Каруша-Куна-Таккера	114
4.5	Нижняя граница отношения длительности буферизации к просмотру при передаче адаптивных видеопотоков	118
4.6	Численный пример	122
4.7	Выводы по разделу	125
	ЗАКЛЮЧЕНИЕ	128
	СПИСОК ЛИТЕРАТУРЫ	129
	СПИСОК РИСУНКОВ	134
	СПИСОК ТАБЛИЦ	136
	ПРИЛОЖЕНИЕ А АППРОКСИМАЦИИ ФУНКЦИИ MOS ДЛЯ ПЕРЕДАЧИ ВИДЕОДАНЫХ	137

ПРИЛОЖЕНИЕ Б	ПАРАМЕТРЫ МОДЕЛИРОВАНИЯ	140
ПРИЛОЖЕНИЕ В	ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ЗАВИСИМОСТЕЙ РАЗДЕЛОВ	143

ВВЕДЕНИЕ

В настоящий момент времени огромной популярностью обладают сервисы хранения и передачи видео по протоколу прикладного уровня HyperText Transfer Protocol (HTTP). Данное явление вызвано множеством факторов, такими как бурное развитие мобильных устройств, увеличение аудитории социальных сетей и их плотная интеграция с сервисами хранения видеоконтента, рост популярности дистанционного обучения, видеокурсов, видеолекций и т. д. Подобная комбинация факторов приводит к доминированию передачи видеоданных в современных телекоммуникационных системах.

Важной особенностью передачи видео по протоколу HTTP является наличие двух технологий организации передачи видеоданных: неадаптивная (HTTP Progressive Download) и адаптивная (HTTP Adaptive Streaming), представленные в стандарте Dynamic Adaptive Streaming over HTTP, известным под сокращением DASH. Так же, в современных реалиях, пользователи обладают высокой мобильностью, что приводит к использованию беспроводных сетей связи в качестве носителя информации. Последним этапом развития беспроводных систем связи являются централизованные системы, позволяющие достичь высоких скоростей передачи информации и низких задержек в радиоканале. Однако, рост объемов видеотрафика приводит к экстремальным нагрузкам на беспроводную сеть, что проявляется в появлении эффектов деградации качества обслуживания абонентов: увеличение длительности ожидания начала воспроизведения и прерывание проигрывания видеоконтента при просмотре.

Общая производительность беспроводных централизованных сетей во многом определяется аспектами работы с беспроводным каналом связи, а именно методом распределения частотно-временных ресурсов между пользователями. В современных стандартах связи распределение ресурсов осуществляет алгоритм планирования (планировщик), установленный на уровне доступа ко среде, базовой станции. Однако, ввиду обширности требований, планировщики не регламентируются стандартами, и каждый производитель оборудования по собственному усмотрению выбирает принципы, в соответствии с которыми будет организовано распределение ресурсов радиоканала, что имеет непосредственное влияние на производительность системы в целом.

Таким образом, актуальной является задача анализа производительности и создания алгоритмов планирования для беспроводных централизованных сетей, которые обеспечивают высокую производительность и достаточный уровень качества восприятия при передаче видеоданных по протоколу HTTP.

При анализе производительности беспроводных централизованных сетей для передаче видео использовалась теория замкнутых систем массового обслуживания с конечным числом абонентов, проработанная А. Scherr, L. Kleinrock, В.М. Вишневым и А.И. Ляховым. Применение настоящей теории к анализу систем передачи видеоданных было представлено в ряде работ отечественных: Е.А. Бакин, Г.С. Евсеев, А.И. Парамонов, и зарубежных авторов: А. El Essaili, О. Oyman, V. Ramamurthi, исследующих производительность алгоритмов планирования для передачи видеоконтента. В большинстве подобных работ рассматривается неадаптивная технология передачи и предлагаются подходы для увеличения производительности в соответствии с рассматриваемым критерием качеством восприятия.

Целью настоящего диссертационного исследования является определение, вычисление и построение численных показателей максимально возможной производительности беспроводных централизованных телекоммуникационных систем при использовании адаптивной и неадаптивной технологии передачи видеоданных по протоколу HTTP, и предложение класса алгоритмов планирования распределения ресурсов беспроводного канала на базовой станции, производительность которых близка к максимально достижимой производительности беспроводных централизованных сетей.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Исследовать способы адаптивной и неадаптивной технологий передачи видеоданных по протоколу HTTP;
2. Исследовать методы оценки качества восприятия видеоряда для адаптивной и неадаптивной передачи видеоданных и выделить факторы, обладающие наибольшим влиянием на качество восприятия;
3. Ввести модель системы передачи видеоданных, включающую в себя модели компонентов системы передачи видео и беспроводной централизованной сети, и найти взаимосвязь между ее параметрами;

4. Предложить аналитические оценки максимально возможной производительности введенной модели телекоммуникационной системы для предложенных критериев качества восприятия адаптивной и неадаптивной передачи видеоданных по протоколу НТТР;
5. Разработать алгоритм планирования распределения ресурсов беспроводного канала связи на основе полученных аналитических результатов и продемонстрировать его производительность в сравнении с ними и существующими решениями.

Объект и предмет исследования. Объектом исследования является беспроводная централизованная телекоммуникационная система с доминированием передачи видеоданных по протоколу НТТР.

Предмет исследования составляет процесс распределения частотно-временных ресурсов беспроводного канала связи на базовой станции при передаче видеоданных по протоколу НТТР.

Методы исследования. При получении основных результатов работы использовались общие методы теории вероятностей и математической статистики, теории случайных процессов, методы математической оптимизации, в частности нелинейного и невыпуклого программирования, а также методы имитационного моделирования.

Научная новизна:

1. Построена трехкомпонентная модель системы передачи видеоданных по протоколу НТТР в беспроводных централизованных сетях связи, позволяющая провести аналитические исследования и сравнение производительности алгоритмов распределения ресурсов беспроводного канала;
2. Найдена взаимосвязь между характеристиками системы передачи информации и проигрывания видеоряда при передаче видео по НТТР протоколу;
3. Предложен способ вычисления нижней границы по всевозможным алгоритмам планирования, удовлетворяющим введенной модели, для нормированного отношения длительности буферизации к просмотру при неадаптивной передаче видеоданных;
4. Предложен и реализован алгоритм планирования, минимизирующий нормированное отношение длительности буферизации к просмотру при неадаптивной передаче видеоданных;

5. Найдена нижняя граница для отношения длительности буферизации к просмотру с учетом средней битовой скорости потока при адаптивной передаче видеоданных по всевозможным алгоритмам планирования и адаптации видеоряда, удовлетворяющим введенной модели.

Научная и практическая значимость диссертационной работы. Полученные в диссертационной работе результаты позволяют повысить производительность алгоритмов планирования распределения ресурсов беспроводного канала, и, как следствие, системы передачи информации в целом, для передачи видео по протоколу НТТР, что способствует увеличению емкости беспроводных систем. Так же полученные результаты могут быть использованы для формирования требований к разрабатываемым стандартам связи текущих и последующих поколений.

Степень достоверности. Результаты, полученные в диссертационной работе, согласуются с известными исследованиями передачи видеоданных по протоколу НТТР в беспроводных сетях. Основные результаты опубликованы в рецензируемых журналах и доложены на крупных международных конференциях.

Апробация работы. Основные результаты работы докладывались и обсуждались на следующих конференциях и симпозиумах в период с 2013 по 2017 гг.: на научных сессиях ГУАП; на конференции «СПИСОК-2014» на 15-й конференции «Conference of Open Innovations Association FRUCT»; на 16-м симпозиуме «Problems of Redundancy in Information and Control Systems».

Внедрение результатов. Результаты работы были использованы в рамках проекта «Разработка технических решений по построению внутриобъектовой сети мобильной радиосвязи» ПАО «Информационные телекоммуникационные технологии». Кроме того, результаты работы используются в учебном процессе кафедры инфокоммуникационных систем и кафедры безопасности информационных систем ГУАП.

Личный вклад. Все результаты, представленные в тексте диссертационной работы, получены автором лично.

Публикации. Материалы, отражающие основное содержание и результаты диссертационной работы, опубликованы в 14 печатных работах. Из них 2 работы опубликованы в рецензируемых научных журналах, утвержденных в перечне ВАК, и 4 работ опубликованы в журналах, индексируемых в Scopus.

Основные положения, выносимые на защиту:

1. Модель беспроводной централизованной системы связи при передаче видеоданных по протоколу НТТР, позволяющая производить аналитические исследования алгоритмов планирования распределения ресурсов беспроводного канала .
2. Взаимосвязь характеристик беспроводной централизованной сети и воспроизведения видеоряда при передачи видео по протоколу НТТР.
3. Нижняя граница для нормированного отношения длительности ожидания к просмотру по всевозможным алгоритмам планирования, удовлетворяющим введенной модели, при передаче неадаптивных видеопотоков.
4. Алгоритм планирования распределения ресурсов беспроводного канала для минимизации нормированного отношения длительности ожидания к просмотру при передаче неадаптивных видеопотоков.
5. Нижняя граница для отношения длительности буферизации к просмотру с учетом средней битовой скорости потока при адаптивной передаче видеоданных по всевозможным алгоритмам планирования и адаптации видеоряда, удовлетворяющим введенной модели.

Объем и структура работы. Диссертация состоит из введения, четырёх разделов, заключения и двух приложений. Полный объём диссертации составляет 143 страницы с 29 рисунками и 11 таблицами. Список литературы содержит 47 наименований.

1 ОСОБЕННОСТИ ПЕРЕДАЧИ ВИДЕОДАННЫХ В ТЕЛЕКОММУНИКАЦИОННЫХ СЕТЯХ

1.1 Вводные замечания

В течение последних лет наблюдается существенное развитие телекоммуникационных систем, особенно сетей мобильной связи. Это проявляется в росте пропускной способности каналов передачи информации и увеличении числа обслуживаемых устройств в сети. Так же увеличивается производительность мобильных устройств, что позволяет им обрабатывать больший объем информации и приводит к увеличению нагрузки на беспроводные сети. Наибольшей популярностью среди пользователей мобильных сетей пользуются сервисы просмотра видео контента по протоколу HyperText Transfer Protocol (YouTube, Vimeo и т. д.) плотно интегрированные с социальными сетями. Подобное явление приводит к колоссальному росту нагрузки на существующие мобильные сети. Следует отметить, что за 2016-й год объем передаваемой информации через телекоммуникационные мобильные сети увеличился на 63 процента и доминирующим видом передаваемой информации является видео контент, который занял 60 процентов объема трафика прошедшего через мобильные сети. По прогнозам ведущих производителей сетевого оборудования видеоданных, займет около 78-ми процентов всего мобильного трафика к 2021-му году [1].

Основной проблемой сетей мобильной связи является ограниченность доступной частотной полосы для организации беспроводного канала, что приводит к ограничениям на пропускную способность каналов связи и число абонентов, одновременно активных в сети. Увеличение объема трафика в сети негативно влияет на качество обслуживания абонентов, наиболее сильно выражающегося в уменьшении скорости передачи информации, увеличении среднего значения и дисперсии задержки при доставке данных абоненту. Наиболее чувствительной, к данным факторам обслуживания, является передача видеоданных. Как следствие, операторы мобильной связи и производители телекоммуникационного оборудования ищут пути оптимизации совместной передачи различных видов трафика в существующих и разрабатываемых системах, где наибольшим приоритетом обладает именно передача видеоданных.

Для организации исследования в области передачи видеоданных важным является установление ключевых показателей производительности сети передачи информации и критерия качества восприятия с позиции доставки видео контента. Основным критерием производительности сетей передачи видеоданных является именно критерий качества восприятия, который является субъективной реакцией человека на характеристики воспроизведения видеопоследовательности. Важной отличительной чертой передачи видеоданных является влияние на качество восприятия особенностей работы воспроизводящего устройства, называемого видеоплеером.

Данный раздел посвящен рассмотрению особенностей передачи видеоданных в современных телекоммуникационных сетях и организован следующим образом. В начале рассматривается система передачи видео в целом и описывается формат хранения видеопоследовательностей. Далее производится анализ существующих видеоплееров, и на его основе предлагается модель генерируемого трафика при просмотре видео. В заключении производится обзор ключевых показателей производительности телекоммуникационных сетей и методик оценки качества восприятия при передаче видео контента.

Ввиду большого числа англоязычных терминов, используемых в данном разделе, далее представлен список основных используемых сокращений, их расшифровок и кратких описаний.

- Key Performance Indicators (KPI) – ключевые показатели производительности;
- Quality of Experience (QoE) – критерий качества восприятия;
- Mean Opinion Score (MOS) – усредненное мнение по группе пользователей;
- Transmission Control Protocol (TCP) – протокол транспортного уровня для гарантированной доставки информации;
- Media Presentation Description (MPD) – формат файла, описывающий представление видеопоследовательности на сервере;
- HyperText Transfer Protocol (HTTP) – протокол прикладного уровня для обмена информацией в телекоммуникационных сетях;
- HTTP Progressive Download (HPD) – группа плееров не осуществляющих адаптацию видеопотока;

- HTTP Adaptive Streaming (HAS) – группа плееров осуществляющая адаптацию видеопотока;
- Moving Picture Experts Group - Dynamic Adaptive Streaming over HTTP (MPEG-DASH) – международный стандарт передачи видеоданных, поддерживающий адаптацию видеопотока.

1.2 Структура систем передачи видеоданных

Современные системы передачи видеоданных по протоколу HyperText Transfer Protocol состоят из трех основных компонентов (рисунок 1.1):

- Видео Контент Сервер – программно-аппаратный комплекс обеспечивающий хранение и доступ к видеоданным;
- Сеть передачи информации – среда передачи информации, организующая соединение между узлами сети;
- Пользовательское устройство – программно-аппаратный комплекс производящий загрузку и демонстрацию видео пользователю.

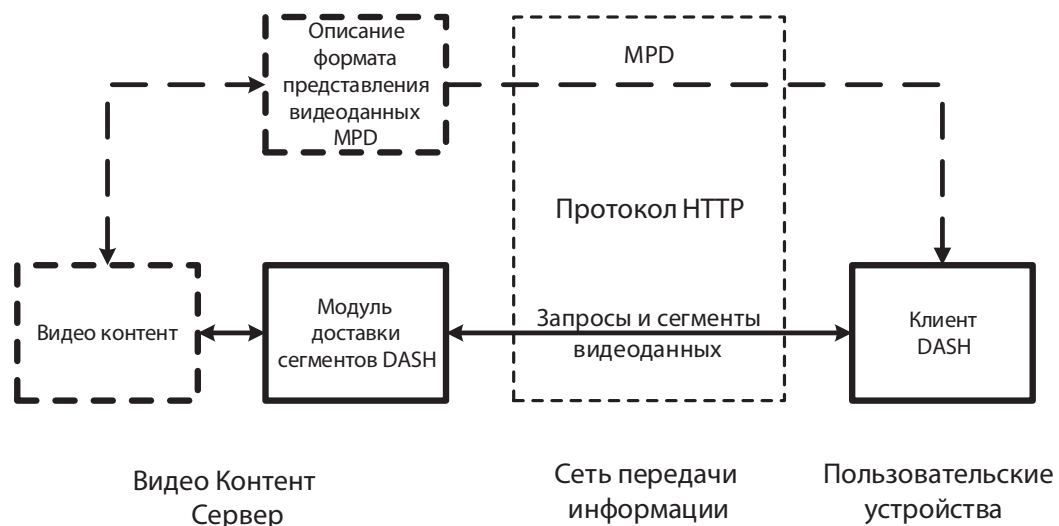


Рисунок 1.1 — Обобщенная структура систем передачи видеоданных

При передаче видеоданных, удаленно в сети Интернет установлена система серверов (Видео Контент Сервер), хранящая видео данные. К данной системе по средствам HyperText Transfer Protocol (HTTP), HyperText Transfer Protocol Secure (HTTPS) или Transport Layer Security (TLS) протоколам через сеть переда-

чи информации подключаются пользовательские устройства. При установке соединения с каждым пользователем производится формирование сессии по протоколу гарантированной доставки данных TCP (Transmission Control Protocol). Данный протокол обеспечивает надежную передачу информации между узлами сети за счет использования дополнительных сообщений, подтверждающих доставку данных (квитанций). Важно отметить, что загрузкой видеоданных управляет программный комплекс установленный на пользовательском устройстве – видеоплеер. Именно видеоплеер принимает решение о порядке загрузки данных путем формирования последовательности запросов на сервер хранения информации, в зависимости от воспроизведения и статистики получения информации.

Исходя из специфики работы сетевых протоколов HTTP и TCP, при прохождении через сеть передачи информации происходит только задержка данных при доставке на пользовательское устройство. Как следствие, невозможна ситуация, когда происходит потеря данных при передаче через сеть. Важной особенностью систем, основанных на протоколе HTTP, является их инвариантность к сети передачи информации: для обеспечения работы такой системы необходима лишь корректная работа протокола HTTP и нижележащих уровней. Однако, использование на транспортном уровне протокола TCP вносит небольшую избыточность, обусловленную квитированием и повторными передачами данных.

Задержка при передаче информации может привести к появлению негативных эффектов воспроизведения, таких как длительное время ожидания начала проигрывания, остановка воспроизведения ввиду недостатка загруженных данных и, как следствие, уменьшению удовлетворенности пользователя просмотром.

Целью данного раздела является именно установление взаимосвязи между характеристиками сети передачи информации и удовлетворенностью пользователя проигрыванием потока прошедшего через нее. Так как только на основе данной взаимосвязи возможно построить решения для увеличения производительности телекоммуникационных сетей для передачи видеоданных. Первым шагом для достижения поставленной цели является рассмотрения формата представления видео информации при передаче через телекоммуникационные сети.

Основные результаты данного раздела опубликованы в работе [2].

1.3 Представление видеоданных при передаче через телекоммуникационные сети

Для дальнейшего повествования крайне важно провести описание формата хранения информации на Видео Контент Сервере для передачи по средствам HTTP протокола. Каждая последовательность, хранящаяся в памяти сервера, разбита на периоды равной длительности называемыми сегментами (в англоязычной литературе Segments или Chunks). Каждый сегмент характеризуется уникальным идентификатором (порядковым номером и идентификатор видео) и репрезентацией [3].

Репрезентация является общепринятой характеристикой видеопотока. Она включает в себя три параметра:

- Битовая скорость потока – объем информации, необходимое для хранения в одну единицу времени. Общепринятая единица измерения битовой скорости потока Мбит/с (мегабит в секунду). В русскоязычной литературе для обозначения данного термина иногда используется термин битрейт, являющийся транскрипцией англоязычного аналога – Bitrate.
- Рекомендованное разрешение – количество точек экрана занимаемое при демонстрации сегмента. Данная характеристика измеряется в прогрессивной развертке (p) плотно ассоциирована с размерами экранов воспроизводящих устройств, например, 720p определяет экран формата высокого разрешения (High Definition) с 720 точками по вертикали. Исходя из стандартного соотношения сторон экранов 16:9, определяется экран размера 1280 точек по горизонтали и 720 по вертикали.
- Частота кадров – количество кадров воспроизводимое в одну единицу времени. Общепринятой единицей измерения частоты кадров является кадр в секунду.

Важно отметить, что при передачи сегмента через сеть информация, содержащаяся в сегменте представляется в виде последовательности из пакетов равного объема. В реальной системе размер пакета ограничен сверху значением максимального размера полезного блока информации одного пакета, передача которого возможна без фрагментации данных (в англоязычной литературе Maximum

Transmission Unit или MTU). В сети Интернет значение MTU варьируется в отрезке от 536 до 1440 байт.

Для демонстрации взаимосвязи перечисленных выше параметров приводится таблица 1.1, демонстрирующая рекомендации по настройкам репрезентации сегментов [4; 5]. В данной таблице частота кадров считается Стандартной, если она не превышает 30-ти кадров в секунду, и Высокой если она превосходит данное значение.

Таблица 1.1

Рекомендованные настройки репрезентации сервисом YouTube

Разрешение	Стационарные устройства		Мобильные устройства
	Стандартная частота кадров	Высокая частота кадров	
2160p (4K)	35–45 Мбит/с	53–68 Мбит/с	13.5 Мбит/с
1440p (2K)	16 Мбит/с	24 Мбит/с	6 Мбит/с
1080p (Full HD)	8 Мбит/с	12 Мбит/с	3 Мбит/с
720p (HD)	5 Мбит/с	7.5 Мбит/с	1.5 Мбит/с
480p	2.5 Мбит/с	4 Мбит/с	0.7 Мбит/с
360p	1 Мбит/с	1.5 Мбит/с	0.45 Мбит/с
240p	-	-	0.25 Мбит/с

Из анализа таблицы 1.1 возможно вывести следующие закономерности:

1. С увеличением разрешения экрана на одну позицию, битовая скорость возрастает примерно в два раза;
2. Битовая скорость видеопотока для мобильных устройств в несколько раз ниже, чем для стационарных. Данный факт обусловлен разницей между размерами экранов мобильных и стационарных устройств.

При преобразовании исходной видеопоследовательности в формат хранения данных на контент сервере возможны два режима кодирования: с неизменяемой и изменяемой битовой скоростью потока (в англоязычной литературе Constant и Variable Bitrate соответственно). Под неизменяемой битовой скоростью потока понимается следующее: все сегменты последовательности в заданном разрешении имеют равные битовые скорости. При изменяемой битовой скорости каждый сегмент даже в одном разрешении может иметь отличные друг от друга битовые скорости. Наличие двух режимов кодирования вызвано динамичностью

сцен исходной последовательности, например, статичные сцены могут быть более эффективно сжаты видеокодеком, чем динамичные, как следствие в зависимости от сложности сцен исходного видео имеется возможность уменьшить объем хранимой информации без потери качества.

На качественном уровне, наличие двух возможных режимов кодирования последовательностей затрудняет разработку решений для оптимизации передачи видеоданных. Так как появляется дополнительная зависимость требований к ресурсам телекоммуникационной системы от содержания видео. На данный момент сервис YouTube поддерживает только режим с неизменяемой битовой скоростью репрезентации [4].

Для описания всей имеющейся информации о видеопоследовательности был разработан специальный формат представления Media Presentation Description (MPD) файл, описывающий каждый сегмент последовательности: идентификатор, длительность и репрезентация. Файл данного формата загружается на видеоплеер перед началом загрузки видео, и на основе информации, записанной в нем, видеоплеер будет осуществлять загрузку видеопоследовательности.

После определения формата хранения и представления видеопотока необходимо ответить на вопрос: каким образом, используя описанный выше формат представления, организована передача видео с Контент Сервера на пользовательское устройство. Исходя из представленной информации в подразделе 1.2 следует, что ведущую роль при передаче видео исполняет видеоплеер. Как следствие, следующим шагом необходимо провести анализ существующих видеоплееров и на основе данного анализа сформировать модель трафика, генерируемого при просмотре видео контента.

1.4 Организация передачи видео через телекоммуникационные сети

Важным моментом при организации передачи видео через телекоммуникационные сети является работа воспроизводящего устройства и модель поведения пользователя. Реакция воспроизводящего устройства на действия пользова-

теля, информацию о характеристиках видеопоследовательности, статистику получения данных определяет модель генерируемого трафика.

Изначально введем модель поведения пользователя при просмотре видео контента (рисунок 1.2). В начальный момент времени пользователь не просматривает видео данные (не активен). Через некоторый случайный промежуток времени будет начат просмотр видеоряда: дана команда видеоплееру, установленном на пользовательском устройстве, организовать передачу и демонстрацию загруженного контента с сервера хранения видео. Важно отметить, что пользователь просматривает заказанное видео полностью, не смотря на любые возможные эффекты деградации воспроизведения, и считается активным в период просмотра. После окончания просмотра каждой видеопоследовательности, через случайный промежуток времени (паузу) пользователь снова начнет просматривать видео.

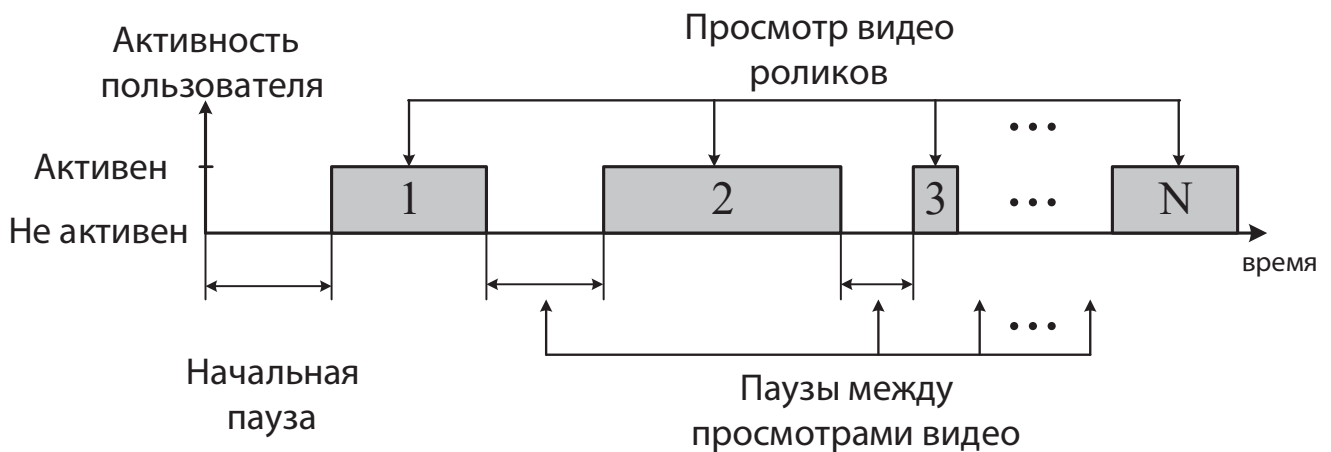


Рисунок 1.2 — Модель поведения пользователя при просмотре видео

Описанная выше модель поведения пользователя может быть определена следующим набором случайных величин:

- Длительность интервала времени перед заказом первого видео;
- Длительность заказанного видеоряда;
- Длительность пауз между просмотрами видео.

Таким образом, при наблюдении за трафиком, проходящим через сетевой (и нижележащие) уровни семиуровневой модели OSI, фиксируется чередование периодов наличия и отсутствия (пульсации) трафика у абонента. Однако, данное утверждение задает лишь обобщенную модель трафика со стороны пользователя системы передачи видеоданных. Для того чтобы окончательно определить мо-

дель генерируемого трафика необходимо рассмотреть аспекты работы видеоплеера в активный период загрузки видеопоследовательности.

В настоящие дни существует огромное множество сервисов хранения видеоданных и каждый из них представляет собственный программный комплекс, обеспечивающий загрузку видео с уникального сервиса. Исходя из данного факта провести обзор всех возможных видеоплееров не представляет возможности, поэтому в данной работе будет представлена обобщенная модель видеоплеера. Данная модель не теряя общности описывает все важные аспекты работы всевозможных плееров видеоданных.

На рисунке 1.3 изображена машина состояний видеоплеера. В начальный момент времени плеер находит в состоянии *Ожидания* начала загрузки, нахождение в данном состоянии обусловлено ожиданием указаний от пользователя. После получения команды пользователя о начале загрузки будет установлено соединение с сервером хранения видео информации и произведен переход в состояние *Буферизация*.

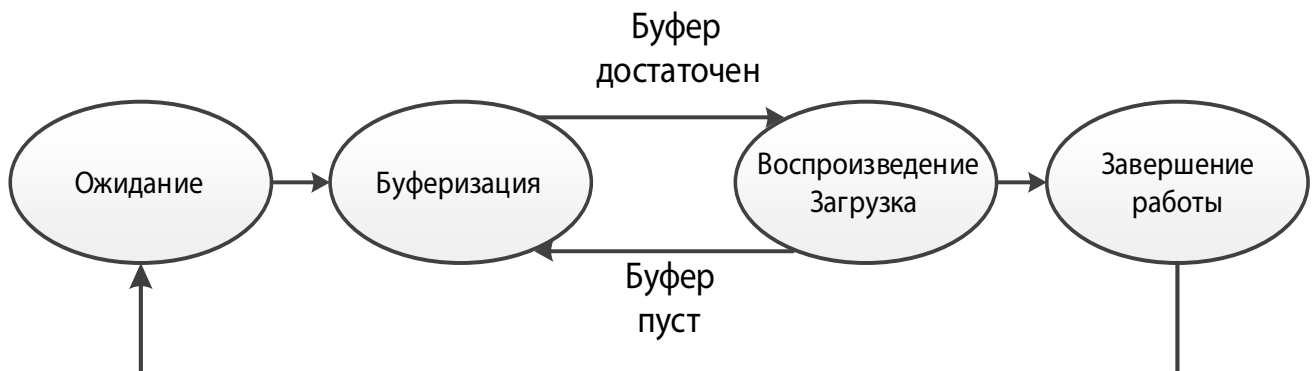


Рисунок 1.3 — Машина состояний видеоплеера

В состоянии *Буферизация* видеоплеер не осуществляет демонстрацию загружаемых видеоданных, а производит накопление информации до определенного уровня для последующей демонстрации. Осуществление буферизации на воспроизводящем устройстве призвано уменьшить негативные эффекты воспроизведения, вызванные изменением свойств канала передачи информации, что особенно актуально для беспроводных систем передачи информации. Прибывание в данном состоянии является негативным эффектом при просмотре видео. После накопления достаточной длительности видеопоследовательности в буфере видеоплеер переходит в состояние *Воспроизведение и Загрузка*.

В течении состояния *Воспроизведение и Загрузка* плеер решает две важные задачи: демонстрация контента и адаптация загружаемого потока для максимизации длительности возможного времени воспроизведения. Вторая задача не является тривиальной и решается, в рассмотренных далее реализациях, эвристическими методами. Важно отметить, что в данном состоянии происходит одновременное наполнение и опустошение буфера на пользовательском устройстве, и если скорость наполнения будет меньше скорости опустошения, то в некоторый момент времени в буфере не найдется необходимого объема данных для обеспечения демонстрации. Описанное событие называется опустошение буфера, и при его наступлении видеоплеер возвращается в состояние *Буферизация*, где будет находиться до накопления достаточного объема видеоданных.

В некоторый момент времени будет закончена загрузка видеоданных на пользовательское устройство. И видео плеер осуществит переход в состояние *Завершение Работы*. В данном состоянии будет осуществлено закрытие соединения с видео контент сервером и ожидается окончание просмотра пользователем загруженных данных, после чего видеоплеер переходит в состояние *Ожидания* начала загрузки.

Завершающим шагом для описания модели генерируемого трафика при просмотре видео является рассмотрение существующих методов адаптации загружаемого потока для максимизации длительности возможного времени воспроизведения в состоянии *Воспроизведения и Загрузки*. Всевозможные плееры видеоданных разделяются на два множества:

- HTTP Progressive Download (HPD) – множество плееров не осуществляющих адаптацию видеопотока;
- HTTP Adaptive Streaming (HAS) – множество плееров осуществляющая адаптацию видеопотока.

Для упрощения терминологии и уменьшения используемых англоязычных аббревиатур далее в работе при обозначении плееров HPD будет использоваться термин *неадаптивные*, а для обозначения плееров HAS – *адаптивные*

В начале тысячелетия в сети Интернет все видео были представлены лишь в одной репрезентации, и для их доставки использовались реализации неадаптивных плееров. Однако, даже в настоящее время неадаптивные плееры очень широко распространены. Они используются в браузере Safari для загрузки видеоданных с сервиса YouTube, для просмотра видео в социальной сети Facebook

и при передаче трансляций в режиме реального времени (спортивные соревнования, прессконференций и т. д.). Как правило, для мероприятий, транслирующихся в реальном времени, не организуется кодирование видеопотока для нескольких репрезентаций, а доступна лишь одна для ознакомления с содержанием мероприятия.

Далее приведено описание работы неадаптивного плеера в состоянии *Буферизация* и *Воспроизведение и Загрузка*. После получения команды от пользователя о начале загрузки некоторого видео на пользовательское устройство загружается MPD файл с описанием последовательности, и пользователю предлагается выбрать репрезентацию для просмотра. На основе выбранной репрезентации будет отправлен запрос на контент сервер, который организует последовательную отправку сегментов видео на воспроизводящее устройство. Далее происходит буферизация некоторой длительности последовательности и после окончания буферизации начнется воспроизведение. Если по каким-либо причинам пользователь желает изменить репрезентацию просматриваемого видео, то на сервер будет отправлен новый запрос с ее указанием и идентификатором сегмента, с которого необходимо начать передачу видеоданных (рисунок 1.4). Логика отправки запросов на контент сервер регламентируется правилом: «запрос на новый сегмент отправляется после получения текущего загружаемого сегмента».

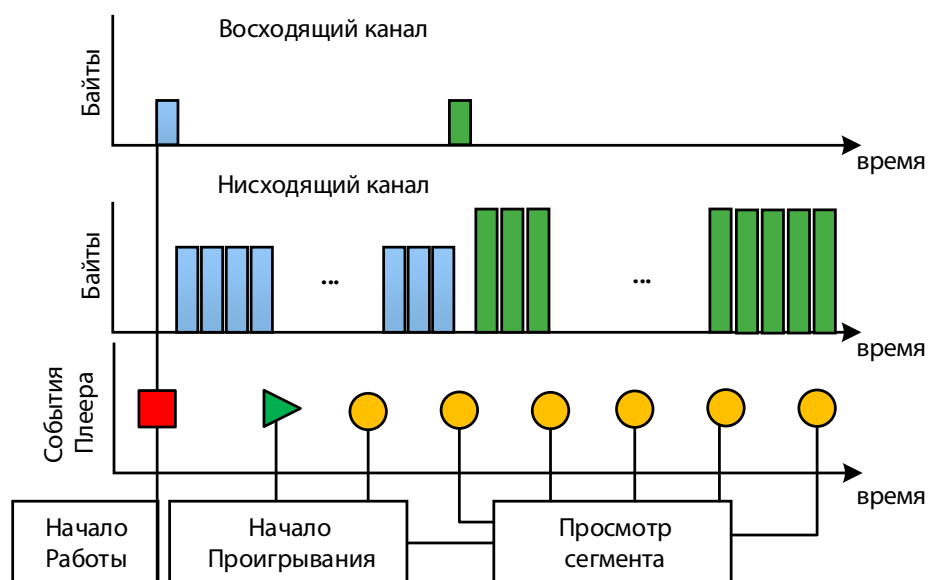


Рисунок 1.4 — Логика работы неадаптивного видеоплеера

Во время работы неадаптивного видеоплеера в состоянии *Воспроизведение и Загрузка* на сетевом уровне наблюдается постоянный поток видеоданных в нис-

ходящем канале (от видео контент сервера к пользователю), а в восходящем канале редкие запросы от воспроизводящего устройства.

Таким образом, модель генерируемого трафика при работе неадаптивного видеоплеера представляется последовательностью периодов высокой нагрузки на сеть передачи информации со случайными паузами между ними. Длительность периодов высокой нагрузки определяется только длительностью видео и выбранной репрезентации для просмотра, без учета текущей ситуации в сети. Подобная модель поведения значительно увеличивает нагрузку на систему передачи информации.

За последние десять лет количество обслуживаемых устройств в сетях передачи информации росло в геометрической прогрессии. И в начале первого десятилетия 21-го века стало очевидно, что существующие телекоммуникационные сети, особенно мобильные, не обладают достаточной емкостью для обеспечения достаточного качества обслуживания большого количества абонентов одновременно просматривающих видеоконтент. Под емкостью сети понимается количество пользователей, которое может одновременно находиться в сети, с необходимым уровнем качества обслуживания каждого активного пользователя. Ввиду невозможности кардинального изменения существующих телекоммуникационных систем для доставки видео был разработан новый стандарт для передачи видеоданных Moving Picture Experts Group Dynamic Adaptive Streaming over HTTP (MPEG-DASH), который позволяет решить проблему емкости сети [3]. В настоящее время представлена версия стандарта 3GP-DASH в спецификации консорциума, разрабатывающего решения для беспроводной связи 3GPP [6]. Стандарт 3GP-DASH включает в себя описание адаптивной и неадаптивной технологии передачи видео по протоколу HTTP.

Стандарт DASH описывает лишь формат представления видеоданных (подраздел 1.3) и общие идеи правил формирования запросов видеоплеера к серверу на сегменты видео, которые позволяют организовать адаптивную передачу видеопотока для специфичных требований каждого конкретного сервиса доставки контента. Обобщенная логика работы видеоплеера, описывающая основную идею представленную в стандарте DASH, демонстрирует рисунок 1.5.

После установки соединения с контент сервером и загрузки MPD файла видеоплеер самостоятельно генерирует и отправляет первый запрос на видеоданные. После загрузки каждого сегмента производится оценка скорости получения

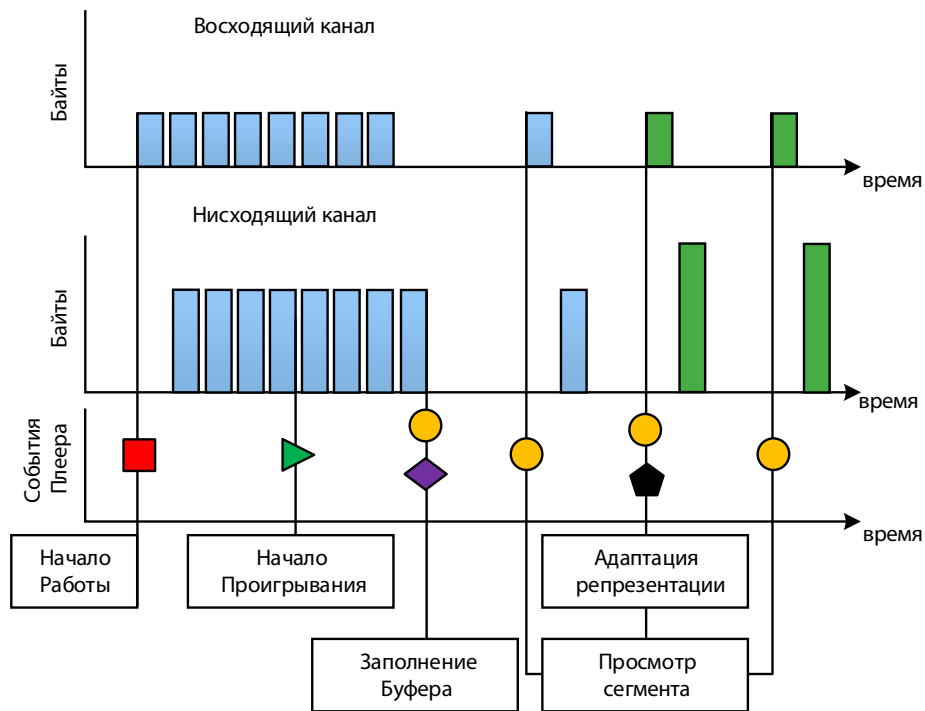


Рисунок 1.5 — Логика работы адаптивного видеоплеера

информации и длительности загруженного видео, на основе которых механизм адаптации видеоряда выберет репрезентацию для следующего сегмента. В состоянии *Воспроизведение и Загрузка* изменяется логика отправки запросов на сегменты данных с «запрос на новый сегмент отправляется после **получения текущего загружаемого сегмента**» на «запрос на новый сегмент отправляется после **просмотра текущего сегмента**», что приводит к появлению пауз между загрузками сегментов. Наличие алгоритма изменения репрезентации и пауз между загрузками сегментов приводит к уменьшению нагрузки на телекоммуникационные сети, и, как следствие, увеличению их емкости.

Таким образом, важными отличительными особенностями адаптивного видеоплеера от неадаптивного являются:

- Отправка плеером запросов на контент сервер для каждого уникального сегмента видеопоследовательности;
- Наличие пауз между заказами сегментов в состоянии *Воспроизведение и Загрузка*;
- Наличие алгоритма адаптации репрезентации под характеристики сети передачи информации.

Наибольший интерес из перечисленных выше особенностей представляет алгоритм адаптации репрезентации, так как он имеет непосредственное влияние на модель генерируемого трафика. Как было сказано ранее стандарт DASH не регламентирует данный алгоритм. Однако, существует реализация опорного видеоплеера, полностью соответствующего стандарту DASH с открытым исходным кодом, представленный консорциумом DASH Industry Forum, который занимается продвижением настоящего стандарта и предоставляет открытую реализацию видеоплеера и серверных приложений. Данный плеер был реализован в проекте Dash.js, исходный код которого доступен на платформе разработчиков программного обеспечения GitHub [7]. Плеер Dash.js полностью соответствует требованиям стандарта DASH, постоянно обновляется и дополняется при консультациях с группой экспертов, разрабатывающий данный стандарт. Далее в работе будет описана работа опорного плеера Dash.js версии 1.5. В ходе проведения исследования был проанализирован исходный код на языке JavaScript и проведен комплекс испытаний для подтверждения полученных в ходе анализа исходного кода.

Для наглядного представления результатов проведенного анализа был создан рисунок 1.6, описывающий логику адаптации битовой репрезентации видеоплеером в состояниях *Буферизация* и *Воспроизведение и Загрузка*. Все решения алгоритма адаптации репрезентации основаны на значении параметра «*Уровень Буфера*» – длительность загруженной видеопоследовательности на пользовательском устройстве в секундах, доступной для воспроизведения. Например, при значении «*Уровень Буфера*» равном 5-ти секундам на воспроизводящем устройстве доступно к воспроизведению 5 секунд видео в текущий момент времени, и если плеер находится в состоянии *Воспроизведение и Загрузка* и в течение 5-ти секунд не будет загружен ни один сегмент, то воспроизведение прекратится и плеер перейдет в состояние *Буферизация*.

Плеер Dash.js является многопоточным приложением и его работа организована в трех основных потоках, одновременно исполняющихся на пользовательском устройстве: ПРОИГРЫВАНИЕ, ЗАГРУЗКА и МОНИТОРИНГ.

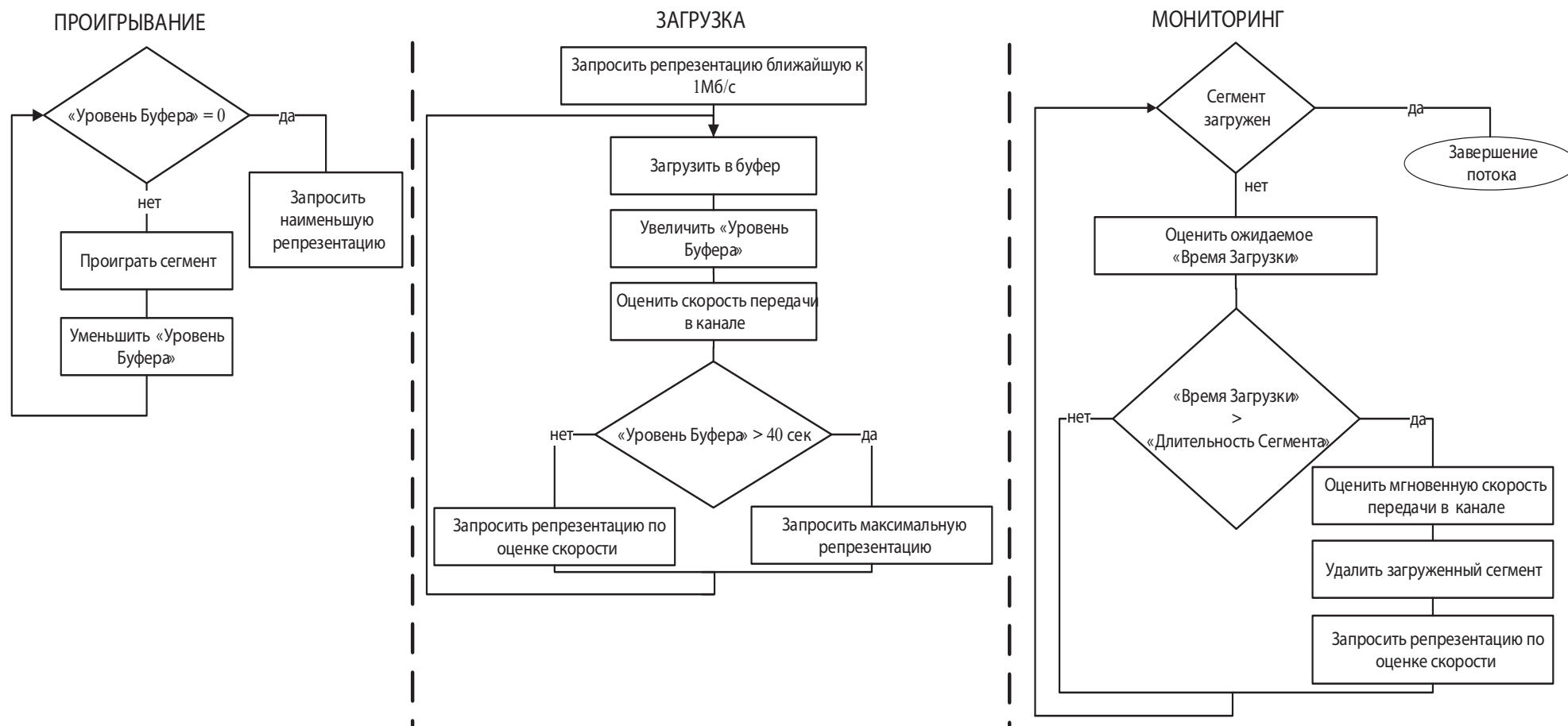


Рисунок 1.6 — Логика адаптации репрезентации опорного видеоплеера Dash.js

В начальный момент времени работы плеера доступен MPD файл с описанием каждого сегмента видеопоследовательности, и необходимо решить задачу выбора репрезентации для первого сегмента видеопоследовательности. Для решения данной задачи реализована функция подбора репрезентации по оценке скорости получения информации. Словесно данная функция может быть описана следующим образом: найти ближайшую снизу репрезентацию с битовой скоростью не превышающую оценку скорости. Она вызывается в блоках, где присутствует команда «Запросить». Для первого сегмента будет подобрана репрезентация ближайшая к скорости 1 Мбит/с, отправлен запрос на сервер и запущен поток приема информации – ЗАГРУЗКА.

В потоке ЗАГРУЗКА происходит прием информации с нижележащих уровней сети, формирование сегментов видеопоследовательности и помещение их в буфер на воспроизводящем устройстве. Рассмотрим цикл загрузки одного сегмента видео. После отправки запроса на сервер в нисходящем канале начинается передача данных, и необходимо организовать их прием, контроль корректности, порядка и полноты загрузки данных. Когда сегмент будет полностью доставлен на воспроизводящее устройство он будет загружен в буфер, и значение параметра «Уровень Буфера» увеличено на длительность загруженного сегмента. Далее для выбора репрезентации для следующего сегмента производится оценка скорости загрузки информации в два этапа:

1. Вычисление оценки скорости загрузки текущего сегмента, как отношение размера сегмента к длительности его загрузки;
2. Вычисление результирующей оценки скорости загрузки как среднего арифметического оценок скоростей нескольких последних сегментов и умножение, полученного значения, на понижающий коэффициент. Занижение результирующего значения необходимо для компенсации возможных колебаний скорости загрузки информации.

Далее значение параметра «Уровень Буфера» сравнивается с порогом достаточного уровня буфера. Если значение «Уровень Буфера» выше, то следующий сегмент будет заказан в максимальной доступной репрезентации, иначе выбор репрезентации будет основан на вычисленной ранее оценке.

Одновременно с потоком ЗАГРУЗКА работает поток МОНИТОРИНГ, задачей которого является контроль длительности загрузки сегментов. Для этого в каждый момент времени загрузки происходит оценка ожидаемого времени за-

грузки сегмента и если данная оценка превышает произведение длительности сегмента и коэффициента порогового ожидания загрузки, то загрузка текущего сегмента будет прекращена. При прекращении загрузки все загруженные данные удаляются и в этот же момент, сервер уведомляется о прекращении загрузки сегмента и на него отправляется новый запрос с указанием репрезентации, выбранной на основе оценки скорости загрузки только текущего сегмента.

После накопления размера достаточной для начала воспроизведения длительности видео будет запущен поток ПРОИГРЫВАНИЕ, который производит демонстрацию загруженного в буфер видео. При проигрывании сегмента он изымается из буфера, и значение «Уровень Буфера» уменьшается на длительность, изъятого сегмента. Данный поток влияет на выбор репрезентации только в том случае, когда уровень буфера равен нулю (в буфере нет данных для воспроизведения), тогда плеер прекращает демонстрацию видео, переходит в состояние *Буферизация*, а следующий сегмент будет заказан в минимальной репрезентации.

На представленный выше алгоритм адаптации видеоряда под условия канала передачи информации налагается ограничение на частоту переключения репрезентаций видео. Очевидным является факт, что частое переключение битовых репрезентаций негативно влияет на качество восприятия видеопотока: в некоторых случаях быстрое переключение репрезентаций при просмотре приводит к проявлению симптомов неврологических расстройств. Поэтому в современных реализациях видеоплееров количество переключений в единицу времени ограничено сверху [8].

Основные параметры видеоплеера Dash.js были агрегированы в таблице 1.2.

Для адаптивного видеоплеера модель генерируемого трафика зависит как от характеристик просматриваемого видеопотока, так и системы передачи информации. Она определяется настройками видеоплеера и алгоритма адаптации репрезентации.

На текущий момент был представлен анализ организации передачи видео в телекоммуникационных сетях. Рассмотрены основные виды видеоплееров и показано влияние плеера на модель генерируемого трафика. Исходя из проведенного анализа следует, что обслуживание передачи видеоданных является нетривиальной задачей. Наиболее важным для исследования становится вопрос о влиянии характеристик сети передачи информации и видеоплеера на качество обслуживания абонента. Именно данному вопросу посвящена следующая часть раздела.

Основные параметры адаптивного видеоплеера Dash.js

Название параметра	Описание	Значение
Общие параметры		
Длительность сегмента видеоданных	Длительность сегментов, на которые разделена видеопоследовательность	[2, 10] секунд
Размер начальной буферизации	Длительность видеопоследовательности, которая должна быть загружена перед началом воспроизведения	12 секунд
Максимальный размер буфера	Максимальная длительность видеопоследовательности, которая может быть помещена в буфер	60 секунд
Параметры адаптации репрезентации		
Начальная оценка канала	Оценка канала передачи данных, на основе которой будет выбрана репрезентация первого сегмента	1 Мбит/с
«Память» при оценке скорости канала	Число сегментов, на основе которых будет производиться оценка скорости канала передачи информации	3
Понижающий коэффициент оценки скорости	Коэффициент для понижения оценки скорости передачи информации, при оценке скорости загрузки	0.9
Достаточный уровень буфера	Длительность буферизированного видеопотока, при котором плеер будет заказывать только максимальную репрезентацию	40 секунд
Коэффициент порогового ожидания загрузки сегмента	Длительность числа сегментов, которых будет ожидать плеер, перед удалением текущего сегмента	1.5

1.5 Методики оценки качества передачи видеоконтента

Оценка качества передачи видеоконтента является сложной задачей ввиду субъективного восприятия пользователем характеристик воспроизведения. Существуют два основных подхода для оценки качества для всевозможных сервисов:

- Экспертная оценка [9];
- Оценка пользовательского опыта использования сервиса [10].

Изначально для определения качества телекоммуникационных сервисов использовались экспертные оценки. Получение подобной оценки состояло из нескольких этапов: подбор членов экспертной комиссии, формулировка строго поставленных вопросов и ожидание результата. Данная процедура занимает много времени, как следствие, использование экспертной оценки имеет смысл только уже к разработанным системам передачи информации, однако для оценки качества на этапе разработки системы использовать данный подход не представляется возможным. В настоящее время мнения экспертов используют для формирования требований к разрабатываемым стандартам связи следующего поколения.

В тоже время, наличие положительной экспертной оценки не является гарантией удовлетворенности пользователя использованием сервисом. Данный факт привел к появлению второго метода оценки качества: оценка пользовательского опыта использования сервиса, в англоязычной литературе этот термин известен как Quality of Experience (QoE). Термин QoE очень общий и включает в себя огромное число методик и критериев для различных сервисов, любой критерий с системой сравнения результатов, отражающий пользовательский опыт, считается QoE.

Для оценки качества телекоммуникационных сервисов была принята методология оценки, известная как Mean Opinion Score (MOS) [11]. Результат оценки MOS представляется числом в отрезке от 1-цы до 5-ти, где единица описывает наименьшую степень удовлетворенности пользователя, а пять наивысшую. Считается, что пользователь удовлетворен качеством сервиса, если значение MOS больше или равно трем. Для получения оценки MOS собирается статистически оправданная группа пользователей, с учетом расового, полового, возрастного и т.д. факторов, производится демонстрация сервиса, и на ее основе члены группы

выставляют оценки. Результирующее значение выражается как среднее арифметическое всех данных оценок. Описанный выше путь получения оценки обладает такими же недостатками, как и экспертная оценка, однако полученный результат является более показательным, и на его основе можно организовать сравнение систем связи.

Множество усилий было направлено на ускорение получения оценки восприятия качества обслуживания, и в настоящее время для каждого вида сервисов существуют разработанные рекомендации по вычислению значения MOS. В каждой подобной рекомендации присутствует описание методологии, которая на основе огромного массива объективных характеристик доступных на пользовательском устройстве вычисляет значение оценки. Полученное значение, по утверждению авторов, не отличается от оценки, поставленной опрашиваемыми пользователями. Важно отметить следующий факт: применимость различных методологий оценки качества восприятия ограничивается типом используемого видеоплеера, например, существуют методологии, которые могут вычислить достоверное значение оценки только если при демонстрации использовался неадаптивный видеоплеер.

Организацией International Telecommunication Union, известной в РФ как Международный Союз Электросвязи, была стандартизирована процедура вычисления MOS для видеопоследовательностей с неизменяемой репрезентацией длительностью от 30-ти до 60-ти секунд [12]. Ввиду большого объема методологии, описанной в [12], в данной работе будет представлено лишь ее описание на качественном уровне. Далее в работе для сокращения длины названий в качестве обозначения данной методологии будет называться термин ITU MOS.

Изначально на основе объективных характеристик рассчитываются два основных фактора при просмотре видео: фактор качества видео и фактор воспроизведения (рисунок 1.7). Фактор качества видео оценивает восприятие пользователем характеристик неискаженного видеопотока на конкретном воспроизводящем устройстве. В качестве входных параметров данный фактор учитывает разрешение экрана воспроизводящего устройства, битовую скорость, частоту кадров и настройки кодирования видео и аудио потоков. Наличие, в приведенном выше списке, настроек видеокодека обусловлено влиянием алгоритма сжатия на потерю качества видеоданных. На выходе возвращается значение, нормированное в отрезке от 1-цы до 5-ти.



Рисунок 1.7 — Схема вычисления оценки качества передачи видео в методологии ITU MOS

Фактор воспроизведения оценивает восприятие пользователем характеристик воспроизведения. Выделяются два основных негативных эффекта, имеющих определяющее влияние на оценку пользователем воспроизведение видео: начальная буферизация и опустошение буфера. Начальная буферизация характеризуется длительностью интервала времени от получения от пользователя команды о запросе видео до начала демонстрации загруженных данных (длительность нахождения в состоянии *Буферизация* в начале работы видеоплеера). Считается, если длительность начальной буферизации превышает некоторое значение, то пользователь откажется от просмотра видео не смотря на его качество.

Второй негативный эффект воспроизведения видео – опустошение буфера. Он характеризуется событием перехода видеоплеера из состояния *Воспроизведение* и *Загрузка* в состояние *Буферизация*, ввиду недостатка видеоданных загруженных в буфер воспроизводящего устройства. Наиболее важными в методологии ITU MOS считаются моменты времени переходов из состояний, длительности повторного нахождения в состоянии *Буферизация* после осуществления начальной буферизации. Из моментов времени переходов рассчитывается частота опустошений буфера при воспроизведении видео. В методологии ITU MOS наличие хотя бы одного опустошения критично снижает оценку качества воспроизведения. На выходе фактора воспроизведения так же представляется числом от 1-цы до 5-ти.

Важным моментом работы методологии ITU MOS является метод агрегации представленных выше факторов, определяемый следующим выражением:

$$ITU MOS = \max([VideoQualityFactor - 5 + VideoPlayingFactor], 1),$$

где *VideoQualityFactor* – фактор качества видео и *VideoPlayingFactor* – фактор воспроизведения.

Из вида данного выражения следует, что результирующее значение MOS не может превышать оценки фактора качества видео: при идеальном воспроизведении (оценка фактора воспроизведения равна 5-ти) результирующее значение равно оценке фактора качества видео, а любое уменьшение качество воспроизведения уменьшает результат.

Основным недостатком методологии ITU MOS является его ограниченная сфера применимости, так как она может быть использована для видеопоследовательностей с неизменной битовой скоростью (неадаптивный видеоплеер) и длительностью не превышающих 60-ти секунд. Также реализация данной методологии затруднено ввиду использования огромного объема информации для вычисления оценки. В работе методология ITU MOS приведена в качестве опорного метода оценки качества восприятия видеоряда, так как все остальные методологии основаны на данной.

В настоящее время предложено несколько методологий оценки качества восприятия просмотра видео. Ярким представителем таких методологий является решение компании Huawei по названию U-vMOS (User, Unified, Ubiquitous-Mean Opinion Score for Video), которая ставит перед собой задачу оценки качества видео без ограничений на их длительность и изменение битовых скоростей. Методология Huawei U-vMOS была представлена в апреле 2016-го года на выставке London TV Connect Exhibition [13].

Подробное описание данной методологии представлено в презентации на сайте компании [14]. В отличии от методологии ITU MOS, для расчета значения U-vMOS требуется намного меньшее число параметров.

Для расчета результирующей оценки качества восприятия видео по методологии U-vMOS должны быть оценены три фактора для одного сегмента видео:

- Фактор качества видеоданных (sQuality);
- Фактор ожидания начала проигрывания (sInteraction);
- Фактор качества воспроизведения (sView).

Далее, для обозначения представленных факторов, будут использоваться оригинальные англоязычные термины.

Фактор *sQuality* описывает восприятие пользователем просмотра видеоряда некоторого разрешения на типовом экране воспроизводящего устройства. Авторами методологии U-vMOS, в качестве типовых экранов, были рассмотрены экран телевизионного устройства (42 дюйма) и экран смартфона (5.5 дюйма). Значение *sQuality* задано таблицей 1.3.

Таблица 1.3

Фактор качества видеоданных (*sQuality*) в методологии U-vMOS

Разрешение репрезентации	Оценка <i>sQuality</i>	
	Телевизионный экран	Экран смартфона
2880p (5K)	4.78	5
2160p (4K)	4.65	4.78
1440p (2K)	4.2	4.58
1080p (Full HD)	4	4.45
720p (HD)	3.15	4
480p	2.44	3.64
360p	1.66	3

Фактор *sInteraction* описывает удовлетворенность пользователем длительностью ожидания в секундах начала воспроизведения потока и задается таблицей 1.4. Данная оценка является общей для всех сегментов видео и вычисляется только для первого сегмента.

Важным фактором воспроизведения является *sView*. В качестве оцениваемой характеристики используется процент времени ожидания возобновления просмотра после события опустошения буфера к длительности сегмента видео. Оценка качества воспроизведения так же является таблично заданной функцией (таблица 1.5).

После того как были рассчитаны все вышеперечисленные показатели качества результирующее значение для одного сегмента может быть получено следующим образом:

$$UvMOS = (sQuality - 1) \left[\frac{\alpha(sInteraction - 1) + \beta(sView - 1)}{4(\alpha + \beta)} \right] + 1,$$

Таблица 1.4

Фактор ожидания начала воспроизведения (sInteraction) в методологии U-vMOS

Оценка sInteraction	Длительность начальной буферизации	
	Телевизионный экран	Экран смартфона
5	< 100 мс	< 100 мс
4	1 сек	1 сек
3	2 сек	3 сек
2	5 сек	5 сек
1	≥ 8 сек	≥ 10 сек

Таблица 1.5

Фактор качества воспроизведения (sView) в методологии U-vMOS

Оценка sView	Процент времени прерывания воспроизведения	
	Телевизионный экран	Экран смартфона
5	0%	0%
4	0.1%	5%
3	1%	10%
2	5%	15%
1	$\geq 10\%$	$\geq 30\%$

где коэффициенты α и β заданы таблицей 1.6.

Таблица 1.6

Коэффициенты для расчета значение MOS в методологии U-vMOS

Коэффициент	Телевизионный экран	Экран смартфона
α	0.66	0.71
β	0.77	0.77

Из вида функции U-vMOS следует, что если зафиксировать некоторое значение показателя sQuality, то результирующее значение не может превышать значение данного показателя. Таким образом, чтобы получить значение MOS превышающее 3, то необходимо демонстрировать пользователю видеопоток с разрешением не менее 720p и 360p для телеэкранов и смартфонов соответственно.

Показатели качества sInteraction и sView, отличные от 5-ти только понижают результирующее значение MOS. Как следствие, для обеспечения значения U-vMOS больше трех следует обеспечивать более высокие репрезентации.

Так же компанией Huawei был представлен программный комплекс для разработчиков систем, который позволяет оценить качество воспроизведения на стороне пользовательского устройства для большинства платформ [15]. Реализованный программный комплекс представлен в виде библиотеки на языке C++. Он позволяет оценить качество для адаптивного видеопотока неограниченной длительности, на основе методики U-vMOS. На вход поступают «сырые» данные для каждого просмотренного сегмента: разрешение, битовая скорость, прерывание воспроизведения и длительность ожидания начала воспроизведения. На выходе представлены оценки воспроизведения потока (sQuality, sInteraction, sView) и результирующее значение оценки удовлетворенности пользователя от просмотра.

Таким образом, метод оценки Huawei U-vMOS применим для адаптивного и неадаптивного видео потока. Данное свойство позволяет использовать его для большего множества видеоконтента, по сравнению с стандартом ITU MOS. Из описания методик оценок качества воспроизведения видео ITU MOS и Huawei U-vMOS, следует что для их использования необходимо иметь доступ к информации на видеоплеере. Данное ограничение значительным образом сужает возможную сферу их применения.

Существует ряд работ, в которых ставится задача оптимизации передачи видео через существующие сотовые сети, по средствам аналитического исследования функции MOS. Авторы данных работ отмечают, что описанные выше методы невозможно использовать в качестве целевых функций, ввиду большого числа параметров и их сложной, нелинейной взаимосвязи. В подобных работах рассматриваются аппроксимации MOS, которые опираются на меньшее число параметров. Так же утверждается, что оптимизация (минимизация или максимизация) таких аппроксимаций приводит к максимизации функции MOS. В таблице А.1 (приложение А) представлены некоторые аппроксимации, отсортированные в порядке возрастания количества используемых критериев для оценки качества обслуживания.

Таблицу А.1 возможно использовать следующим образом: зафиксировать некоторый объем данных, доступный разработчику сети, и выбрать те критерии качества обслуживания, которые могут оценить удовлетворенность пользовате-

ля. Например, при знании битовой скорости просматриваемого потока можно использовать метод, описанный в работе [16], однако, если добавить информацию о длительности опустошения буфера и просмотра возможна более точная оценка на основе критерия [17].

После осуществления обзора различных методик оценки качества восприятия видео остается важный вопрос о взаимосвязи объективных объективных показателей производительности сети передачи видеоданных и оценки качества восприятия (подраздел 1.5). Данному вопросу посвящена заключительная часть раздела.

1.6 О взаимосвязи объективных показателей производительности сети и оценки качества восприятия видео

При передаче видеоданных к телекоммуникационной сети предъявляются специфичные требования, обусловленные характеристиками видеоконтента и настройками воспроизводящего устройства. Для разработчиков сетевого оборудования важно обеспечить максимально возможную емкость сети передачи информации, что ставит задачу на основе некоторого объема информации понять удовлетворен ли пользователь качеством обслуживания. При разработке или модернизации сетей передачи данных разработчикам доступен ограниченный набор показателей качества обслуживания, который может быть с достаточной точностью оценен на физическом, канальном, сетевом и транспортном уровнях. Однако, показатели работы воспроизводящего устройства, которые имеют наибольшее влияние на удовлетворенность пользователя при просмотре видео, недоступны ввиду шифрования трафика, что делает невозможным получение данной информации на нижележащих уровнях сети.

Вопрос о взаимосвязи характеристик сети и качества обслуживания абонентов прорабатывался в большом количестве работ. Авторы подобных работ представляют список объективных характеристик (в англоязычной литературе обозначается как Key Performance Indicators или сокращенно KPI) сети и указывают их значения, для того чтобы пользователь был удовлетворен качеством обслуживания. Так как в настоящее время современные телекоммуникационные сети и

воспроизводящие устройства бурно развиваются, приведенные в данных работах количественные значения показателей изменялся в зависимости от момента написания работ. Однако, список объективных показателей качества обслуживания видеотрафика был неизменным.

На основе анализа работ [5; 18; 19] был выделен следующий список объективных показателей качества обслуживания пользователя, имеющих наибольшее влияние на передачу видео:

- *Задержка* – интервал времени, который проходит от момента отправки пакета передающей стороной до момента получения подтверждения от приемной стороны о получении пакета. В англоязычной литературе для обозначения данного вида задержки используется термин End - to - End, в русскоязычной двухсторонняя.

В реальной системе задержка является случайной величиной. Как следствие, во всех работах в качестве количественного показателя приводится оценка математического ожидания данной случайной величины. Кроме математического ожидания задержку так же характеризуют параметром, который в англоязычной литературе называют Packet Delay Variation [20]. Достаточно часто как в русскоязычной, так и англоязычной литературе данный параметр называют джиттером (Jitter).

- *Джиттер* – возможные колебания задержки. В рекомендации [20] описывается несколько возможных видов данного параметра, далее в работе будет рассматриваться двухточечный джиттер. Двухточечный джиттер для одного пакета принимается равным значению разности между односторонней задержкой, измеренной для данного пакета, и некоторым эталонным значением задержки. Иногда эталонное значение задержки принимают равным минимальному значению, измененному за продолжительное время. В настоящее время влияние джиттера на производительность систем передачи информации нивелировано за счет введения на узлах сети буферов для сглаживания возможных колебаний задержки.
- *Время отклика* – интервал времени, необходимый для организации соединения между приемной и передающей стороной. В некоторых работах данный показатель называется начальной задержкой.
- *Скорость загрузки данных* – скорость передачи данных в сети.

После определении ключевых показателей производительности сети передачи информации и критериев качества восприятия видео пользователем возникает актуальный вопрос о их взаимосвязи. Данный вопрос является основополагающим для реализации методов увеличения производительности телекоммуникационных сетей для передачи видеоконтента.

Изначально нужно определить каким образом возможно реализовать увеличение производительности сети передачи информации для видеоконтента. Для разработчика сетей передачи данных может быть доступна информация с сетевого, транспортного, канального и физического уровней, в зависимости от вида и доступных механизмов анализа системы. Список доступной информации совпадает со списком, представленным ранее в настоящем подразделе, и он является очень ограниченным. Однако, исходя из описания критериев качества восприятия видео (подраздел 1.5), для точной оценки качества необходима информация, доступная только на уровне приложения. И подобный список ограничивается тремя позициями:

- Битовая скорость и разрешение (репрезентация) воспроизводимого потока в текущий момент времени;
- Длительность начальной буферизации для текущего воспроизводимого видео;
- Статистика опустошений буфера на пользовательском устройстве.

Обладание данной информацией для каждого активного пользователя в сети и сегмента просматриваемого им видео позволяет в перспективе сформировать оптимальное управление системой передачи информации для увеличения ее производительности.

На практике все характеристики сети, такие как задержка, джиттер и т.д., агрегируются в полезную скорость передачи информации для каждого пользователя. Полезная скорость – это скорость получения информации на уровне приложения без учета накладных расходов на передачу информации (заголовки, квитанции и повторные передачи протоколов транспортного уровня). При наличии ограниченного ресурса, грамотное перераспределение полезной скорости передачи данных между активными пользователями может увеличить производительность системы в целом. Полезная скорость может быть с достаточной точностью оценена на промежуточных узлах сети как отношение объема информации, прошедшего в нисходящем канале, к длительности его доставки. Данная оценка мо-

жет быть улучшена за счет анализа служебных сообщений протоколов транспортного уровня. Наиболее важным является существование эффективных механизмов управления полезной скоростью передачи данных, например, в проводных сетях возможно вносить задержку в доставку пакетов пользователю для ее ограничения сверху, а в беспроводных влиять на распределение частотно-временных ресурсов радиоканала.

Полезная скорость передачи информации имеет огромное влияние на работу всех видов видеоплееров (подраздел 1.4). Для неадаптивного видеоплеера это выражается в длительности начальной буферизации и прерываниях воспроизведения из-за опустошения буфера, а в адаптивном плеере дополнительно на основе оценки скорости получения информации принимается решение о выборе репрезентации для сегментов видео. Все эти факторы имеют непосредственное влияние на качество восприятия видео, и, как следствие, на производительность сети.

Таким образом, основным элементом управления телекоммуникационной системой при доставке видео является полезная скорость передачи информации и существуют механизмы реализации управления передачей информации.

В исследовании передачи видеоданных по протоколу DASH [5] для методологии оценки качества восприятия Huawei U-vMOS представлена взаимосвязь между объективными критериями качества обслуживания и субъективной оценкой качества восприятия (таблица 1.7). Таким образом для данной методологии существует взаимно-однозначное отображение объективных показателей обслуживания и субъективной оценки качества восприятия видеопотока. Более того, опираясь на таблицу 1.7, возможно осуществить оценку качества восприятия на промежуточных узлах сети для методологии Huawei U-vMOS.

Из представленной выше таблицы следуют следующие выводы:

- На оценку качества восприятия влияют два периода при просмотре видео: начальная буферизация и воспроизведение;
- Скорость загрузки информации при начальной буферизации должна быть значительно выше, чем при воспроизведении для обеспечения требуемого значения U-vMOS. Данный факт объясняется критическим влиянием длительности начальной буферизации при расчете значения U-vMOS (подраздел 1.5), и необходимостью ее минимизации.

Таблица 1.7

Влияние объективных показателей качества обслуживания при передаче видео на критерий качества восприятия U-vMOS

Значение U-vMOS	Объективные показатели				
	Уровень приложения		Промежуточные узлы		
	Разрешение видео	Длительность начальной буферизации	Задержка	Скорость загрузки при начальной буферизации	Скорость загрузки при воспроизведении
3.0	720p	3.5 с	80 мс	5 Мбит/с	2 Мбит/с
3.3	720p	1.5 с	80 мс	7 Мбит/с	2 Мбит/с
4.0	1080p	1 с	40 мс	13 Мбит/с	5 Мбит/с
	2K	1.6 с	40 мс	11.7 Мбит/с	8 Мбит/с
4.5	4K	1 с	10 мс	32 Мбит/с	18 Мбит/с

- Наибольшее влияние на качество восприятия имеет полезная скорость загрузки информации.

Продолжим рассмотрение качества восприятия Huawei U-vMOS, описанный в подразделе 1.5, при использовании адаптивного видеоплеера для мобильных устройств. Покажем, каким образом управление полезной скоростью получения информации влияет на качество восприятия видео и какие цели преследует оптимизация передачи видеоданных. Зафиксируем длительность начальной буферизации для видео равной 2-м секундам (sInteraction) и построим плоскость всевозможных значений критерия U-vMOS. Исходя из описания методологии, для получения оценки, требуется информация о разрешении видеопотока (sQuality) и процента времени прерывания воспроизведения (sView). Для получения оценки sQuality используем значения из таблицы 1.1 о взаимосвязи между битовой скоростью потока и его разрешением для мобильных устройств, а в качестве оценочного значения выберем среднее значение. Для получения промежуточных значений для битовой скорости и всех таблично заданных функций показателей качества U-vMOS был использован метод кусочной интерполяции кубическим полиномами на отрезках (Piecewise Cubic Hermite Interpolating Polynomial (PCHIP)). Полученный результат представлен на рисунке 1.8.

По оси абсцисс отложена средняя битовая скорость просматриваемого видео, по оси ординат средний процент времени прерывания воспроизведения ви-

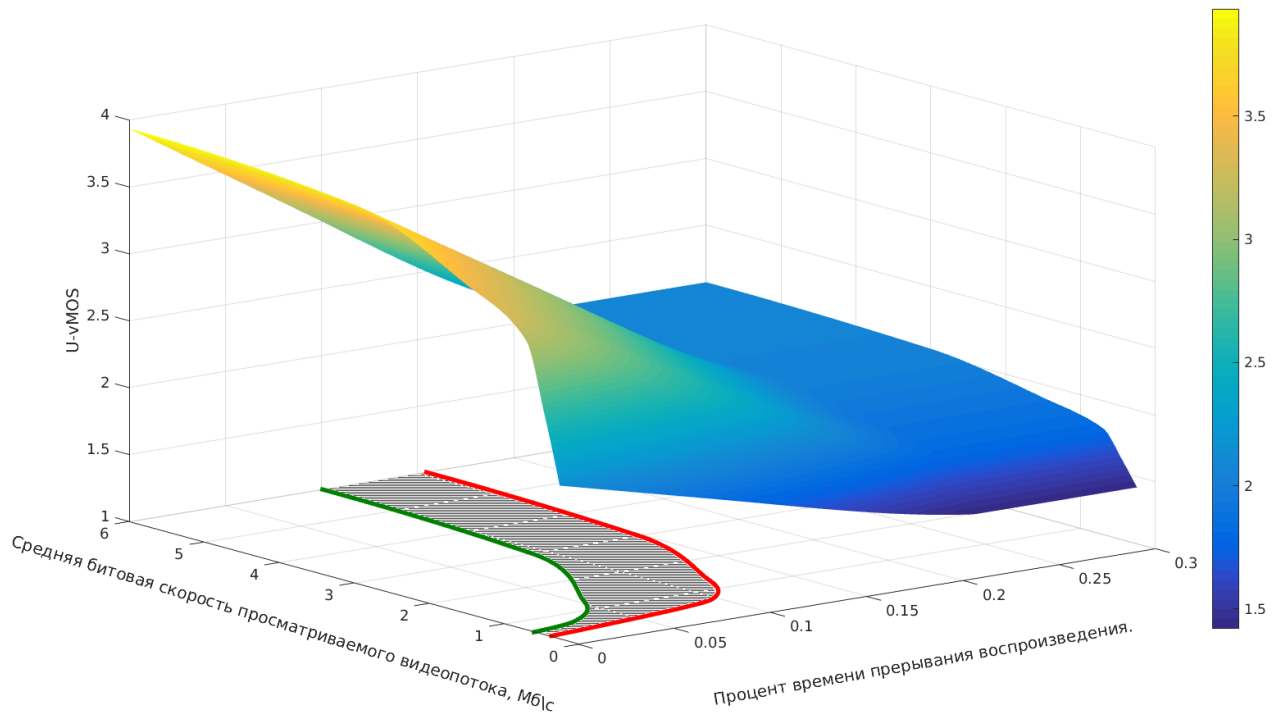


Рисунок 1.8 — Влияние объективных характеристик на критерий качества восприятия U-vMOS

део (повторные буферизации) и ось аппликат представляет результат вычисления критерия качества восприятия U-vMOS, при заданных параметрах. Процесс управлением полезной скоростью передачи информации, во время просмотра видео, может привести в одну из возможных точек полученной плоскости. Значения по оси аппликат находятся в диапазоне от 1-цы до 4-х, разделим плоскость на три области:

- Область высокого значения критерия восприятия (левее заштрихованной области) соответствует значениям в выше 3-х. Если пользователю была обеспечена полезная скорость загрузки информации для попадания в данную область, то он будет удовлетворен качеством обслуживания.
- Граничная область (заштрихованная область) соответствует значениям в отрезке от 2.5 до 3-х. Граничная область характеризует пограничное состояние удовлетворенности пользователя.
- Область низкого значения критерия восприятия (правее заштрихованной области) соответствует значениям в ниже 2.5. В данной области полезной скорости недостаточно для обеспечения достойного уровня обслуживания, таким образом пользователь считается неудовлетворенным.

Целью управления, увеличивающего производительность телекоммуникационной сети, является максимизация числа пользователей, находящихся в области высокого значения критерия восприятия. Это может быть достигнуто путем перераспределения ресурсов системы для пользователей в граничной области, за счет пользователей в области низкого значения критерия, так как перераспределение ресурсов для них не может привести к увеличению их удовлетворенности. Подобный анализ может быть проведен для любого критерия качества восприятия, представленного в подразделе 1.5.

1.7 Выводы по разделу

В качестве выводов по разделу можно отметить, что задача оптимизации передачи видеоданных в телекоммуникационных сетях актуальна, ввиду бурного развития мобильных устройств. Основной целью оптимизации передачи видео в телекоммуникационных сетях является максимизация числа видеопользователей, одновременно активных в сети. Данная цель может быть достигнута путем введения управления полезной скоростью передачи информации, которое основывается на знании о формате представления видеоданных при передаче (подраздел 1.3), особенностей организации передачи видео и типах воспроизводящих устройств (подраздел 1.4), объективных, субъективных показателей качества обслуживания и их взаимосвязи (подразделы 1.5 и 1.6 соответственно).

Увеличение производительности телекоммуникационных систем для передачи видеоконтента является особенно актуальным для мобильных сетей связи, ввиду ограниченности ресурсов и динамичности состояний беспроводного канала передачи информации.

В качестве основных результатов стоит отметить следующие:

- В настоящее время существуют две технологии доставки видеоданных по протоколу HTTP: неадаптивная и адаптивная (подраздел 1.4);
- Оценка качества восприятия видеоданных является сложной задачей ввиду субъективности пользователя, и более того зависит от технологии передачи видео (подраздел 1.5);

- Основные факторы, влияющие на качество восприятия видеопотока в зависимости от технологии передачи (подраздел 1.6):
 - **Неадаптивная технология** – длительность буферизации в период просмотра видео;
 - **Адаптивная технология** – длительность буферизации и битовая скорость просматриваемого потока.

Информация, представленная в данном разделе, во многом носит обзорный характер, с использованием методов обратной разработки (в англоязычной литературе Reverse Engineering) и анализа «черного ящика». Представленное исследование демонстрирует основную проблематику передачи видео в телекоммуникационных системах и является основополагающим для последующих разделов работы.

2 ПЕРЕДАЧА ВИДЕОДАНЫХ В БЕСПРОВОДНЫХ ЦЕНТРАЛИЗОВАННЫХ СЕТЯХ

2.1 Вводные замечания

Оптимизация передачи видеоданных наиболее актуальна для беспроводных систем связи по причине наличия ограничений на пропускную способность радиоканала, который считается «узким местом» (адаптированный перевод англоязычного термина «bottle neck») сети в целом. Как следствие, скорость передачи информации от видео контент сервера до пользовательского устройства определяется именно скоростью передачи данных по радиоканалу. Основная проблематика всех исследований в области передачи видеоданных состоит не в предложении методов управления для телекоммуникационных сетей, а в обосновании их эффективности. Любое возможное управление может увеличить производительность системы, используя дополнительную информацию о передаваемом контенте, однако, в настоящее время не существует опорных теоретических исследований, с которыми возможно сравнить производительность предлагаемых решений.

В настоящее время наиболее распространены централизованные беспроводные системы связи с коммутацией пакетов, так как подобные системы позволяют добиться больших скоростей передачи информации и меньших задержек в беспроводном канале связи. Наиболее ярким представителем данного класса систем является стандарт The Long-Term Evolution (LTE) [21].

Настоящий раздел посвящен задаче передачи видеоданных в современных централизованных беспроводных сетях. В разделе рассматривается специфика построения современных мобильных сетей связи, и решается задача формирования аналитической модели для проведения исследований подобных систем при передаче видеоданных. Предлагаемая модель состоит нескольких вложенных моделей, описание которых последовательно будет представлено в настоящем разделе. В заключении раздела предлагается основополагающий результат, описывающий взаимосвязь между характеристиками сети, влияющими на ее производительность.

Основные результаты данного раздела опубликованы в работах...

2.2 Структура современных беспроводных централизованных сетей передачи информации

В настоящее наибольшее распространение получили централизованные мобильные сети с коммутацией пакетов, так как они обеспечивают наиболее высокие скорости и малую задержку при передаче информации [1]. В централизованных беспроводных сетях вся территория обслуживания разделена на зоны, называемые сотами. В каждой соте установлено устройство, управляющее передачей информацией всех активных пользователей в зоне его ответственности, называемое Базовой станцией. Базовая станция является программно-аппаратным комплексом обеспечивающим обмен информацией между пользователем в радиосети и остальной сетью передачи информации.

Современные беспроводные централизованные системы связи состоят из трех основных компонентов (рисунок 2.1):

- *Фиксированная магистральная сеть передачи данных.* Сеть передачи информации, в которая обеспечивает соединение с удаленными устройствами (в системах передачи видеоданных таким устройством является Видео Контент Сервер). Она характеризуется высокой надежностью, большой скоростью передачи информации, низкими задержкой и джиттром.
- *Опорная сеть оператора.* Система устройств, обеспечивающая взаимодействие пользовательского устройства с удаленными узлами в фиксированной магистральной сети. Данный участок сети обладает схожими характеристиками с магистральной сетью.
- *Радиосеть.* Беспроводная сеть передачи информации, организующая обмен данными между пользовательскими устройствами и остальными компонентами сети. Характеризуется несравнимо меньшими скоростями передачи информации и пропускными способностями каналов, большими задержкой и джиттером. Важной отличительной особенностью радиосети является изменяемость характеристик беспроводного канала передачи информации во времени.

Исходя из представленной выше структуры следует, что аналитическая модель должна агрегировать в себе множество аспектов и параметров, присущих беспроводным централизованным системам, что делает ее сложной и многоуров-

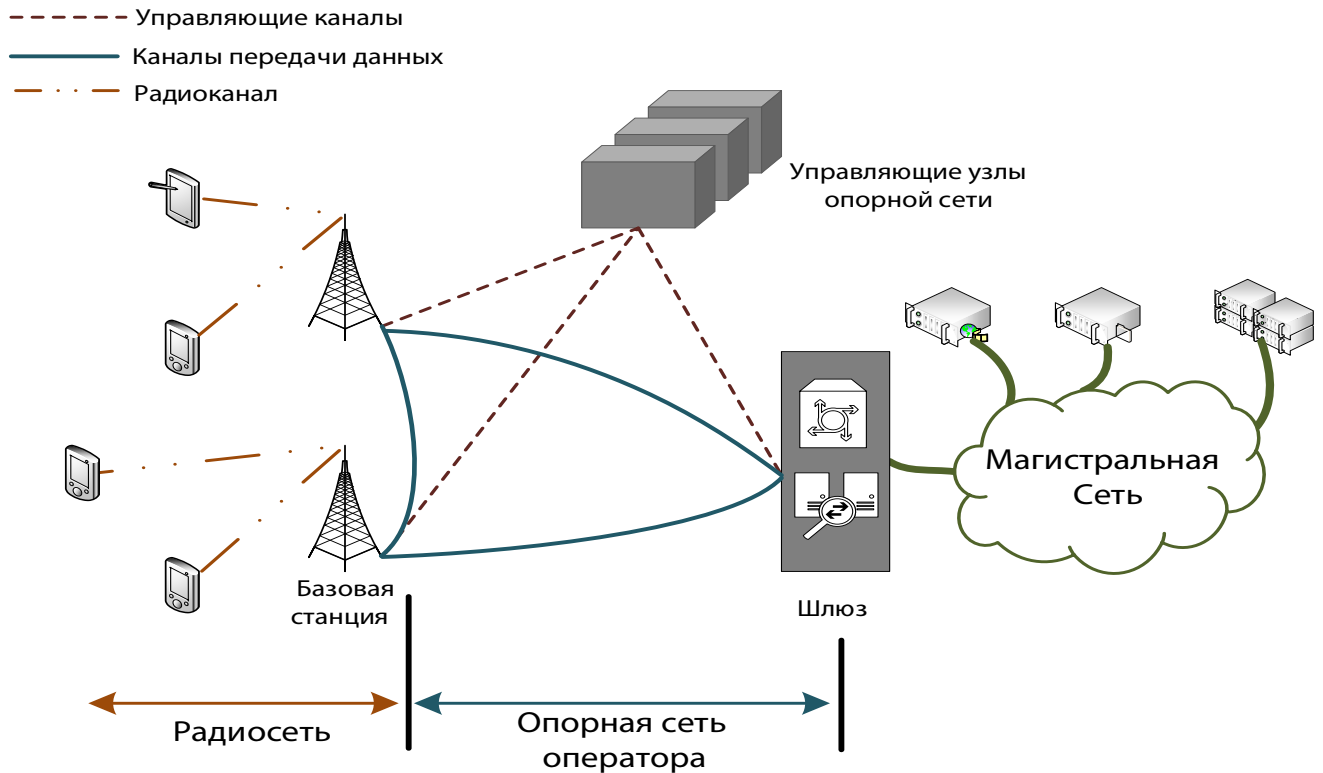


Рисунок 2.1 — Структура современных беспроводных централизованных систем

невой. Как следствие, дальнейшее представление будет разделено на несколько этапов, последовательно описывающих представленную структуру:

- Общие положения предлагаемой модели;
- Модель радиосети:
 - Модель беспроводного канала передачи информации;
 - Модель базовой станции.
- Модель воспроизводящего устройства.

В завершении описания будет представлена система используемых допущений.

2.3 Общие положения аналитической модели

В настоящей работе рассматривается система с конечным числом абонентов: в системе находятся N пользователей, которые используют сервис просмотра видеоконтента. Каждый пользователь обладает устройством с установленным видеоплеером и характеризуется уникальным номером – i , однозначно ассоции-

руемым с воспроизводящим устройством (на рисунках обозначается сокращенно «Устр. i »).

В современных централизованных беспроводных системах базовые станции в различных сотах работают независимо друг от друга: невозможна ситуация, когда одно и то же устройство обслуживают несколько станций одновременно, так же решение о выделении ресурсов беспроводного канала принимается независимо, от работы базовых станций в соседних сотах. Следовательно, каждую соту возможно рассматривать как отдельную независимую систему, в которой необходимо производить оптимизацию передачи видео. На основании вышеизложенного, далее в работе будет рассматриваться работа только одной соты.

Все N абонентов подключены по беспроводному каналу к одной базовой станции и имеют соединение с контент сервером, установленным в магистральной сети. Предполагается, что магистральные и опорные сети связи используют высокоскоростные каналы связи, что позволяет достигать колоссальных скоростей и минимально возможных задержек и джиттера при передаче информации. Таким образом, задержкой прохождения информации через данные участки сети можно пренебречь, и считать передачу данных через них мгновенной.

На удаленном видео контент сервере представлены видео различной длительности, каждое из которых представлено в некотором наборе битовых репрезентаций (подраздел 1.3). Каждое видео представлено в виде последовательности из сегментов равной длительности d секунд. Длительность видеопоследовательностей на контент сервере является некоторой случайной величиной m . Репрезентация каждой видеопоследовательности однозначно сопоставляется с битовой скоростью потока. Далее в работе, для обозначения битовой скорости потока сегмента j , просматриваемого пользователем i , будет использоваться конструкция $R_{i,j}$. Предполагается, что все видеопоследовательности на контент сервере доступны в непрерывном отрезке битовых скоростей от R_{min} до R_{max} :

$$R_{i,j} \in [R_{min}, R_{max}], i = \overline{1, N}. \quad (2.1)$$

В реальной системе непрерывность битовых скоростей может быть достигнута использованием на удаленном сервере траскодеров видеопотока. Траскодер – программный комплекс, позволяющий в режиме реального времени получать видеопоток с заданной битовой скоростью из потока с более высокой битовой скоростью. Использование траскодирования видеопотока позволяет видеоплееру

подбирать битовую скорость видео под конкретные условия системы передачи видеоданных. Общая структура и особенности работы траскодеров видеопотока описаны в работах [22; 23].

Сегмент видеоданных j , при передаче через сеть пользователю j , представляется в виде последовательности из $K_{i,j}$ пакетов равного размера. Общий объем переданной информации сегмента j на пользовательское устройство i составляет $R_{i,j} \cdot d$ бит.

Введем модель, описывающую поведение пользователя при просмотре видеоконтента. Каждый абонент просматривает серию видео (рисунок 2.2) путем отправки запросов на видео контент сервер и последовательной загрузки сегментов видео данных через восходящий и нисходящий каналы связи соответственно. При заказе каждого сегмента видеоданных воспроизводящее устройство генерирует и отправляет на контент сервер уникальный запрос с указанием характеристик сегмента. В начале просмотра видео производится начальная буферизация, после окончания которой начинается демонстрация видеопотока и продолжится загрузка оставшихся видеоданных. После окончания просмотра видео пользователь i через случайный интервал времени (паузу) τ_i закажет просмотр нового видео. Если скорости загрузки информации недостаточно для обеспечения просмотра видео, то могут возникнуть прерывания воспроизведения, вызванные необходимостью накопления достаточного объема видеоданных после опустошения буфера.

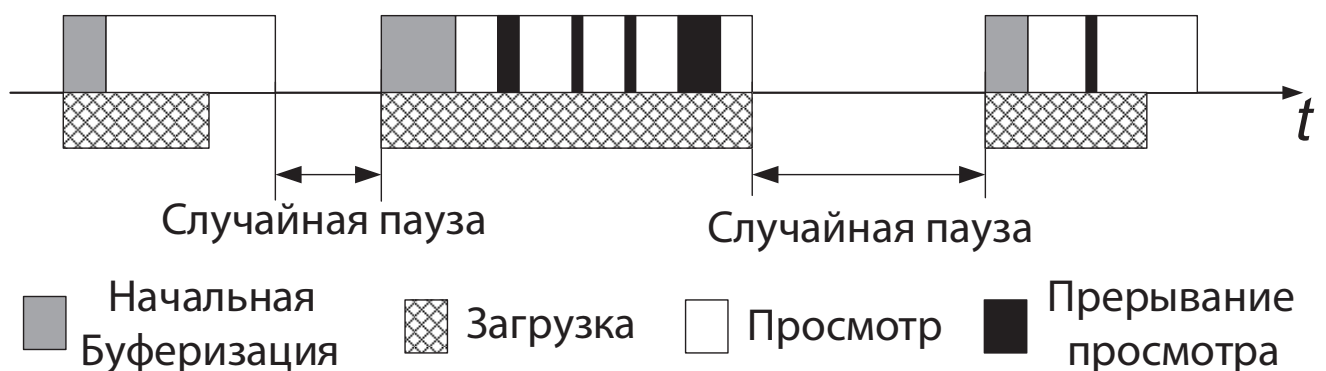


Рисунок 2.2 — Временная диаграмма поведения пользователя при просмотре видео

Из модели поведения пользователя следует, что все время для конкретного пользователя можно разделить на три периода:

- *Буферизация*: пользователь ожидает начала или возобновления проигрывания видео (включает в себя начальные буферизации и прерывания воспроизведения, вызванные опустошением буфера);
- *Просмотр*: штатное состояние, когда буфер на пользовательском устройстве достаточен и пользователь просматривает видеопоток;
- *Пауза*: пользователь выбирает следующее видео для просмотра (в данном периоде просмотр и загрузка видео не осуществляются).

В дальнейшем будут использоваться конструкции b_i^T , w_i^T и p_i^T для обозначения общей длительности буферизации, просмотра и пауз пользователя i за время в отрезке $[0, T]$ соответственно.

На уровне пользователя данную модель поведения возможно описать двумя основными характеристиками: коэффициентом разряженности видеопотока и отношением длительности буферизации к длительности просмотра для каждого конкретного пользователя.

Определение 2.1. Коэффициент разряженности видеопотока пользователя i – это отношение суммы длительностей просмотра и пауз к длительности просмотра пользователя i за интервал времени $T \rightarrow \infty$:

$$\gamma_i = \lim_{T \rightarrow \infty} \frac{w_i^T + p_i^T}{w_i^T}.$$

По определению коэффициент разряженности потока является величиной большей или равной единице: $\gamma_i \geq 1, i = \overline{1, N}$. Данное значение описывает активность пользователя при просмотре видео и показывает взаимосвязь между общим временем просмотра и пауз для конкретного абонента. Например, сравним двух пользователей l и r , которые просмотрели одинаковую длительность видео за время T , однако $\gamma_l > \gamma_r$, это означает что пользователь l делал большие паузы между просмотрами видео, чем пользователь r .

Определение 2.2. Отношение длительности буферизации к просмотру – это отношение общей длительности буферизации к длительности просмотра пользователя i за интервал времени $T \rightarrow \infty$:

$$q_i = \lim_{T \rightarrow \infty} \frac{b_i^T}{w_i^T}.$$

По определению значение отношения времени буферизации к просмотру для каждого пользователя является неотрицательной величиной: $q_i \geq 0, i = \overline{1, N}$. Исходя из подраздела 1.5, буферизация является негативным эффектом воспроизведения видео, который имеет критично большое влияние на качество восприятия абонентом сервиса передачи видеоданных. Как следствие, данный критерий обладает обратной корреляцией с удовлетворенности пользователя: чем больше значение q_i , тем меньше удовлетворенность пользователя.

2.4 Модель беспроводного канала передачи информации

Одно из важнейших положений в описании представляемой аналитической модели занимает модель радиосети. В современных централизованных беспроводных системах радиосеть состоит из трех основных компонентов:

- Беспроводной канал передачи информации;
- Базовая станция;
- Пользовательское устройство.

В начале будет описана структура и аспекты работы беспроводного канала передачи информации. В существующих беспроводных централизованных сетях на физическом уровне ресурсы задаются двумя измерениями: частотой и временем (рисунок 2.3). Во временной области все время работы разделено на периоды равной длительности, называемые слотами. Типовая длительность одного слота равняется 1-й миллисекунде (мс). В англоязычной литературе для обозначения слота используется Slot или Transmission Time Interval (TTI), обозначающие минимальное дробление времени для передачи информации. В частотной области вся полоса передачи информации разделена на области равной ширины, в современных стандартах передачи информации ширина данных областей равняется 180 кГц. Минимальной единицей ресурсов является Ресурсный Блок – одна частотная область длительностью в один слот.

Множество ресурсных блоков формируют общий канал передачи информации, ресурсы которого могут быть распределены между пользователями в радиосети. Помимо общего канала связи реализованы служебные каналы, решающие задачи поддержания синхронизации и управление передачей пользовательских

устройств, оценки качества канала, обеспечения надежной передачи информации по ненадежному каналу и т.д. Данные каналы занимают ресурсы в начале и внутри ресурсных блоков.

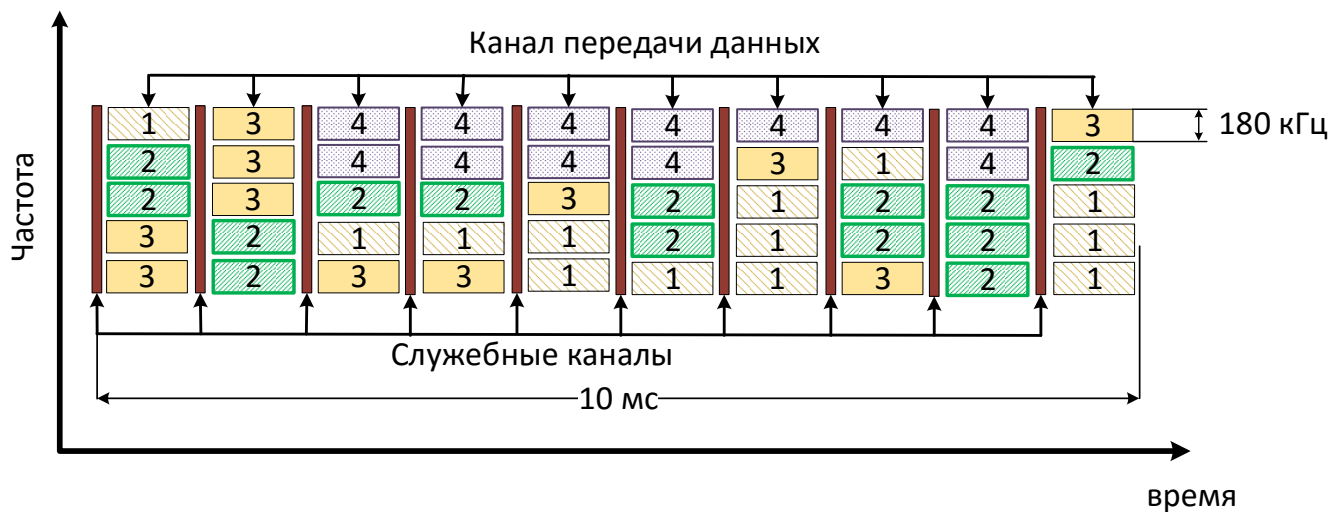


Рисунок 2.3 — Структура беспроводного канала

Важной отличительной особенностью беспроводных каналов связи является изменчивость его характеристик как во времени, так и в частотной области. Подобная изменчивость вызвана наличием помех от работы других систем связи, движением пользователя и окружения, и т.д. Данный факт приводит к тому, что для одного абонента в разных ресурсных блоках может быть передан разный объем информации. Для каждого ресурсного блока на базовой станции вычисляется затухание распространения сигнала на основе информации от служебных каналов связи. Для описания качества канала на вышележащих уровнях базовой станции, полученное значение затухания сигнала для конкретного ресурсного блока и абонента преобразуется в кодо-модуляционную схему. Кодо-модуляционная схема (Modulation Code Scheme или MCS) – целое число в отрезке от 0 до 28, описывающее используемую модуляцию (QPSK, QAM-16, QAM-64) и настройки помехоустойчивого кодирования. Значение кодо-модуляционной схемы имеет прямую корреляцию с уровнем качества канала: чем выше значение MCS, тем выше качество канала. Используемая кодо-модуляционная схема определяет количество байт, которое возможно передать в конкретном ресурсном блоке.

Рассматриваемая модель предполагает, что пользователи в соте могут находиться в различных условиях беспроводного канала, вызванные удаленностью от

базовой станции. Последовательно рассмотрим восходящий и нисходящий беспроводные каналы. В современных централизованных системах связи восходящий канал обладает сравнимыми характеристиками с нисходящим, однако, его загруженность несравнимо меньше по сравнению с нисходящей линией связи. Как следствие, в предлагаемой модели восходящий радиоканал считается абсолютно надежным и задержкой передачи запросов на сегменты видео можно пренебречь.

Аналитическая модель уделяет большое внимание нисходящему каналу связи, так как от его производительности зависит удовлетворенность пользователей в соте. В данной работе рассматривается модель канала, обладающая следующим свойством: затухание при распространении сигнала происходит одинаково по всей ширине полосы передачи информации для конкретного пользователя в одном моменте времени. В реальной системе данное допущение приводит к равенству значений выбранных кодо-модуляционных схем для всех ресурсных блоков в рамках одного слота конкретного пользователя. Следствием равенства кодо-модуляционных является равенство возможного передаваемого объема информации для всех ресурсных блоков в слоте.

Таким образом, состояние беспроводного канала пользователя i возможно охарактеризовать максимально достижимой скоростью канала.

Определение 2.3. Максимально достижимая скорость канала пользователя i – скорость передачи информации по беспроводному каналу, если все доступные ресурсы были выделены i -му пользователю.

Данная величина является аналогичной скорости передачи данных, при условии, что пользователь i находится один в соте. Далее в работе, для обозначения максимально достижимой скорости канала пользователя i будет использоваться конструкция C_i , а для сокращения текста слово «максимальная» будет опускаться.

Обобщая сказанное ранее, для каждого пользователя i существует случайный процесс изменения затухания распространения сигнала от времени $L_i(t)$. На основе значения $L_i(t)$ в момент времени t на базовой станции подбирается кодо-модуляционная схема, таким образом, что вероятность ошибки при передаче информации является пренебрежимо малой. Выбранная кодо-модуляционная схема определяет максимальную пропускную канала $C_i(t)$ в момент времени t .

Предполагается, что состояние беспроводного канала изменяется, таким образом, что в течении загрузки пакета k из сегмента j максимально достижимая скорость канала пользователя i постоянна:

$$C_i(t)|_{t_{i,j,k}}^{t_{i,j,k}+\Delta t_{i,j,k}} = C_{i,j,k}, \quad (2.2)$$

где $t_{i,j,k}$ – момент времени начала загрузки пользователем i пакета k из сегмента j , $\Delta t_{i,j,k}$ – длительность загрузки пользователем i пакета k из сегмента j , $C_{i,j,k}$ – максимально достижимая скорость канала пользователя i в течении загрузки пакета k из сегмента j .

Для современных централизованных беспроводных сетей в типовой соте для передачи информации в одном слоте доступны 48 ресурсных блоков (ширина полосы 10 МГц), и при использовании кодо-модуляционной схемы с модуляцией QPSK в одном слоте возможно передать 900 байт полезной информации. Таким образом, для пользователя в плохих радиоусловиях пакет может быть передан за 2 миллисекунды (подраздел 1.3). Данный факт демонстрирует, что представленное ограничение не сужает возможную сферу применения предлагаемой модели беспроводного канала.

Так же введем дополнительное обозначение $C_{i,j}^{-1}$ равное выборочному среднему для величин обратных к максимально достижимой скорости при передаче пакетов в течение одного сегмента видеоданных:

$$C_{i,j}^{-1} = \frac{1}{K_{i,j}} \sum_{k=1}^{K_{i,j}} \frac{1}{C_{i,j,k}},$$

где $K_{i,j}$ – количество пакетов в сегменте j пользователя i .

2.5 Распределение ресурсов беспроводного канала связи

После определения структуры беспроводного канала необходимо привести описание того как, используя данную структуру, происходит обмен данными между базовой станций и пользовательским устройством. Для этого рассмотрим функциональную структуру базовой станции.

Основной задачей базовой станции, является организация надежной передачи информации по ненадежному радиоканалу. Для решения данной задачи используется структура, состоящая из четырех уровней (рисунок 2.4):

- Уровень обработки пакетов;
- Уровень очередей для передачи информации по беспроводному каналу связи;
- Уровень доступа ко среде передачи информации;
- Уровень физической среды.

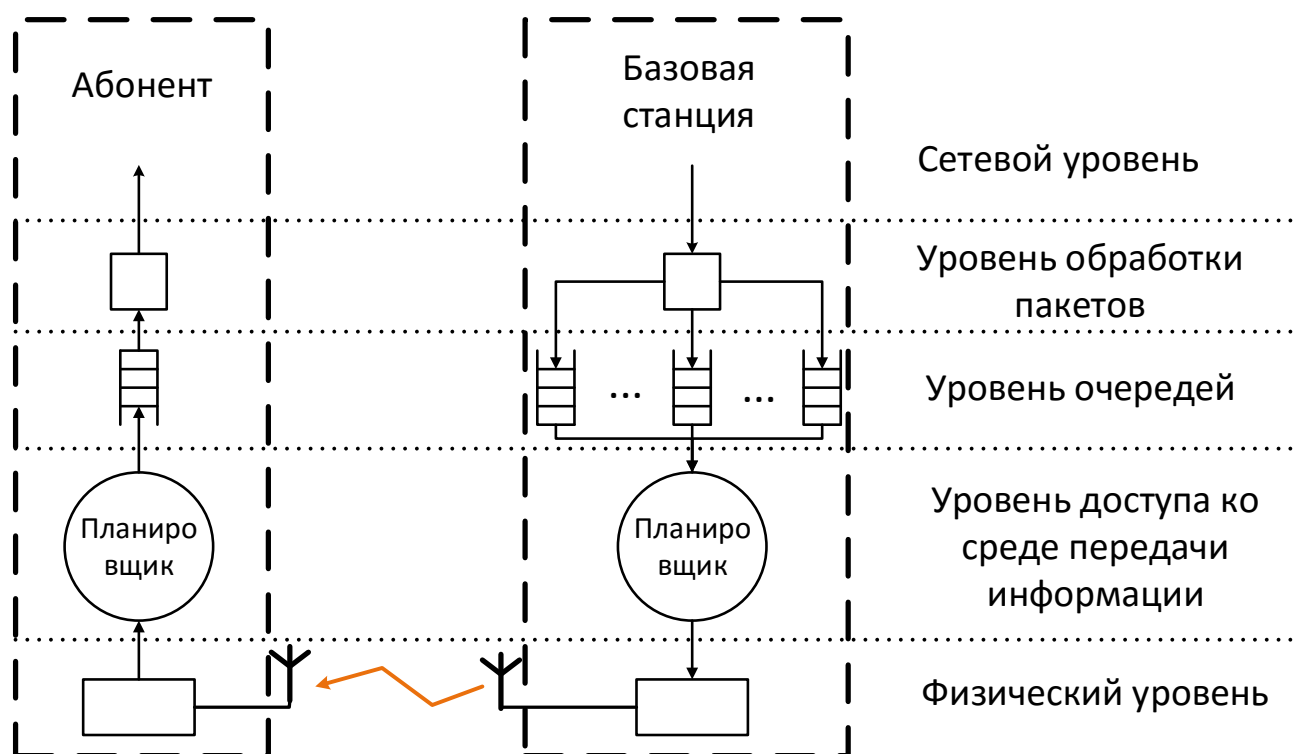


Рисунок 2.4 — Функциональная структура базовой станции и пользовательского устройства

Произведем описание доставки пакета с данными от момента его получения на базовой станции, до момента его появления на сетевом уровне пользовательского устройства. Из опорной сети оператора пакет попадает на уровень обработки пакетов, данный уровень выполняет две функции. Первой функцией является сжатие заголовков пакетов транспортного и сетевого уровня, для уменьшения объема передаваемой информации по беспроводному каналу. Реализация сжатия заголовков пакета позволяет нивелировать накладные издержки при передаче информации по беспроводному каналу и считать, что в пакете нет никакой избыточности, добавляемой сетевыми протоколами. Второй функцией является

определение абонента и передача информационной части пакета в его очередь на нижележащем уровне.

Ниже расположен уровень очередей для передачи информации по беспроводному каналу связи. На данном уровне для каждого пользователя, подключенного к базовой станции, расположена очередь (буфер) для информации. Уровень очередей является промежуточным между уровнями обработки пакетов и доступа к среде передачи информации, выполняющего роль временного хранилища информации при ее передаче по беспроводному каналу. Обслуживанием очередей занимается нижележащий уровень: уровень доступа ко среде передачи (Medium Access Control или MAC).

Уровень доступа ко среде передачи информации решает ключевую задачу для системы в целом: на основе информации от уровня физической среды и вышележащих уровней базовой станции произвести распределение ресурсов беспроводного канала и обеспечить надежность передачи информации. Решением данной задачи занимается планировщик ресурсов беспроводного канала, установленный на базовой станции (в англоязычной литературе используется обозначение Media Access Channel Scheduler или MAC Scheduler). Далее в работе для сокращения объема текста в качестве обозначения «планировщика ресурсов беспроводного канала» будет использоваться термин «планировщик».

Планировщик в каждом слоте производит распределение ресурсов радиоканала (ресурсных блоков) в соответствии с некоторым алгоритмом. Важно отметить, что планировщик не выделяет ресурсы абонентам, у которых нет данных в данном слоте. В каждом слоте формируется карта распределения ресурсных блоков, которая будет передана на физический уровень. На рисунке 2.3 представлен пример распределения ресурсов в течении десяти слотов для четырех активных пользователей. Таким образом, работу алгоритма планирования можно представить в виде распределения долей канала между пользователями. Например, в первом слоте пользователям 1, 2 и 3 были выделены доли беспроводного канала 0.2, 0.4 и 0.4 соответственно.

Сформированная карта распределение ресурсных блоков будет передана на уровень физической среды, который обеспечит передачу информации из уровня очередей в выделенных частотно-временных ресурсах. В дальнейшем, при корректной работе всех описанных уровней базовой станции пакет будет доставлен по беспроводному каналу на пользовательское устройство и, пройдя стек уров-

ней в обратном порядке, станет доступен на сетевом уровне пользовательского устройства.

Несмотря на важную роль планировщика, алгоритмы планирования не стандартизованы для существующих систем связи, поэтому каждый производитель базовых станций использует собственные реализации алгоритмов планирования. Создание подобных алгоритмов является открытой задачей, так как требования к ним очень обширны и не существует строгих теоретических исследований о их максимальной производительности. В настоящее время известны два эвристических алгоритма планирования ресурсов: Proportional Fair и Round Robin. Алгоритм планирования Proportional Fair ставит своей задачей обеспечения равных скоростей передачи информации для всех активных пользователей. Алгоритм планирования Round Robin обеспечивает равный доступ к ресурсам радиоканала для всех активных пользователей. Поэтому настоящая диссертационная работа ставит своей основной задачей исследование алгоритмов планирования, с целью увеличения их производительности и, как следствие, всей беспроводной централизованной сети для передачи видеоданных.

Центральное положение, в представляемой аналитической модели, занимает алгоритм планирования распределения ресурсов беспроводного канала, установленный на базовой станции. Введем определение алгоритма планирования:

Определение 2.4. Алгоритм планирования – это правило, в соответствии с которым базовая станция в момент времени t распределяет доли ресурсов беспроводного канала $\alpha_i(t)$ для пользователя i .

Таким образом, работу алгоритма планирования в момент времени t можно описать набором функций: $A(t) = \{\alpha_i(t), i = \overline{1, N}\}$. Очевидным ограничением на возможные значения функций $\alpha_i(t)$ является следующее неравенство:

$$\forall t : \sum_{i=1}^N \alpha_i(t) \leq 1. \quad (2.3)$$

Неравенство (2.3) может быть интерпретировано следующим образом: в любой момент времени работы алгоритма планирования общий объем выделенных ресурсов не превышает доступного объема ресурсов для канала передачи информации. Следовательно, для любого пользователя i , мгновенная скорость передачи информации в нисходящем канале связи $S_i(t)$ может быть вычислена следующим

образом:

$$S_i(t) = \alpha_i(t)C_i(t). \quad (2.4)$$

Алгоритм планирования в каждый момент времени решает задачу распределения ресурсов беспроводного канала. Для решения данной задачи ему доступна информация о предыстории, а именно доли выделенных ресурсов канала, значения максимально достижимых скоростей канала и объем переданной информации для каждого пользователя:

$$A(t) = \mathcal{A} \left(S_i(\tau), C_i(\tau); \tau < t, i = \overline{1, N} \right), \quad (2.5)$$

где $\mathcal{A}(\cdot)$ является алгоритмом планирования. В формуле (2.5) информация о предыстории выделенных долей беспроводного канала и объемах переданной информации для пользователя i агрегированы в значении $S_i(\tau)$, так как данные параметры задаются соотношением (2.4).

Важно отметить факт, что для планировщика базовой станции пользователь в каждый момент времени может находиться в одном из двух состояний: активном и неактивном. Пользователь считается активным, если у данного пользователя есть информация для передачи в нисходящем канале на базовой станции, иначе пользователь считается неактивным. В реальной системе, активность абонента так же определяется на основе заполненности очередей. Основываясь на модели поведения пользователя при просмотре видео, пользователь может является активным только в период осуществления загрузки данных (рисунок 2.2). Важно отметить, что загрузка информации состоит из последовательной передачи пакетов с видеоданными, и пользователь считается неактивным во время пауз между их загрузками.

В данной диссертационной работе рассматриваются алгоритмы планирования, которые удовлетворяют представленному ниже набору свойств:

- В каждый момент времени активному пользователю гарантировано выделение минимальной доли ресурсов канала α_i^{min} :

$$\begin{cases} \alpha_i(t) = 0, & \text{пользователь } i \text{ неактивен в момент времени } t \\ \alpha_i(t) \geq \alpha_i^{min}, & \text{пользователь } i \text{ активен в момент времени } t \end{cases}. \quad (2.6)$$

Минимальная доля ресурсов канала является величиной, отличной от нуля: $\alpha_i^{min} > 0$, и сумма минимальных долей ресурсов канала для всех пользователей не превышает общего объема ресурсов, доступных для планирования: $\sum_{i=1}^N \alpha_i^{min} \leq 1$.

- Ресурсы беспроводного канала не могут быть выделены неактивному абоненту.
- В каждый момент времени планировщик распределяет все доступные ресурсы между активными абонентами:

$$\sum_{i=1}^N \alpha_i(t) = \begin{cases} 0, & \text{в момент времени } t \text{ все пользователи неактивны} \\ 1, & \text{иначе} \end{cases}. \quad (2.7)$$

2.6 Модель воспроизводящего устройства

В завершении представления аналитической модели будет определена модель видеотрафика и видеоплеера. В данной работе предполагается, что видеоплеер и видео контент сервер поддерживают технологию адаптивной передачи видеоконтента DASH (подраздел 1.4) (рисунок 2.5). Videоплеер, установленный на пользовательском устройстве, осуществляет последовательную загрузку сегментов видеопоследовательности: следующий сегмент видео может быть заказан после окончания загрузки предыдущего и некоторой задержки. Пауза между заказами обусловлены стратегией наполнения буфера и моделью поведения пользователя. Данная задержка может быть короткой, если загружается сегмент в середине видео и достаточной длинной, если загружается первый сегмент, в случае если данная задержка включает в себя случайную паузы между просмотрами видеопоследовательностей, представленные на рисунке 2.2.

При заказе нового сегмента видеопоследовательности видеоплеер решает задачу выбора репрезентации для заказываемого сегмента. В предлагаемой модели репрезентация сегмента однозначно сопоставлена с битовой скоростью, таким образом задача выбора репрезентации сводится к задаче адаптации битовой скорости потока. Для решения данной задачи видеоплеер может использовать информацию о наполненности буфера и других статистиках: оценку скорости по-

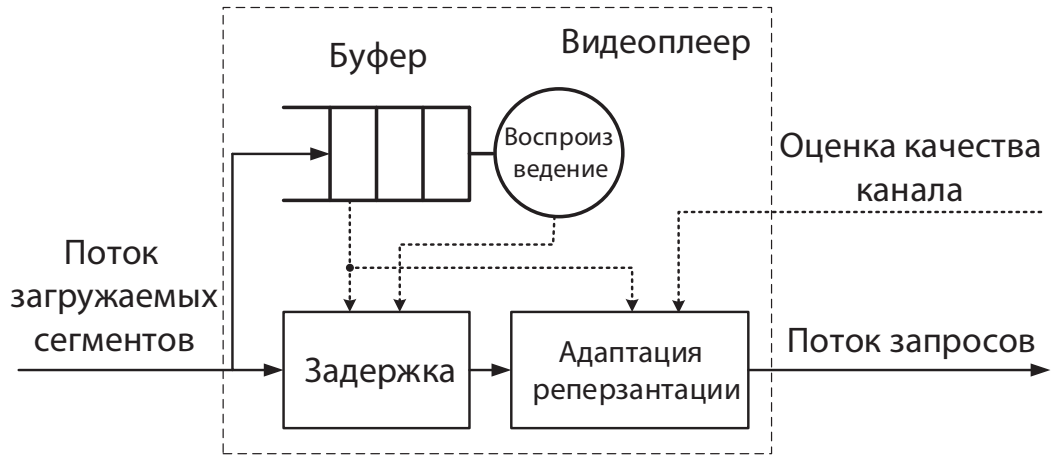


Рисунок 2.5 — Модель воспроизводящего устройства

лучения информации в течении предыдущих сегментов, информацию о качестве беспроводного канала и т.д. Задача адаптации битовой скорости потока может быть представлена следующим выражением:

$$R_{i,j} = \mathcal{B}(R_{i,k}, C_i(\tau), S_i(\tau); k < j, \tau < t_{i,j}),$$

где $\mathcal{B}(\cdot)$ является функцией вычисления битовой скорости j -го сегмента пользователя i , $R_{i,j}$ – битовая скорость потока j -го сегмента пользователя i .

Из описания, представленного в подразделе 1.4, следует, что часто на функцию $\mathcal{B}(\cdot)$ наложено ограничение на количество переключений в единицу времени. Обычно адаптация видеопотока под конкретные условия беспроводного канала происходит в короткий промежуток времени в начале загрузки нового видео. Таким образом, если существуют математическое ожидание: $E[R_i] = \lim_{j \rightarrow \infty} E[R_{i,j}]$ и среднее квадратичное отклонение: $\sigma[R_i] = \lim_{j \rightarrow \infty} \sigma[R_{i,j}]$ битовой скорости просматриваемого потока, то коэффициент вариации битовой скорости видеопотока ограничен сверху некоторой постоянной, значение которой зависит от типа и настроек видеоплеера:

$$\forall i : \frac{\sigma[R_i]}{E[R_i]} \leq \nu_i^R. \quad (2.8)$$

Введение модели видеоплеера завершает описание аналитической модели. Обобщая описанное выше, предлагается следующая система допущений для аналитической модели передачи видеоданных в централизованных беспроводных сетях.

2.7 Система допущений для модели передачи видеоданных

Общая структура предложенной аналитической модели передачи видеоданных представлена на рисунке 2.6. В представленной структуре магистральная и опорная сети были объединены в магистральную сеть передачи информации, ввиду схожести их характеристик. Ниже представлен список используемых допущений в настоящей работе.

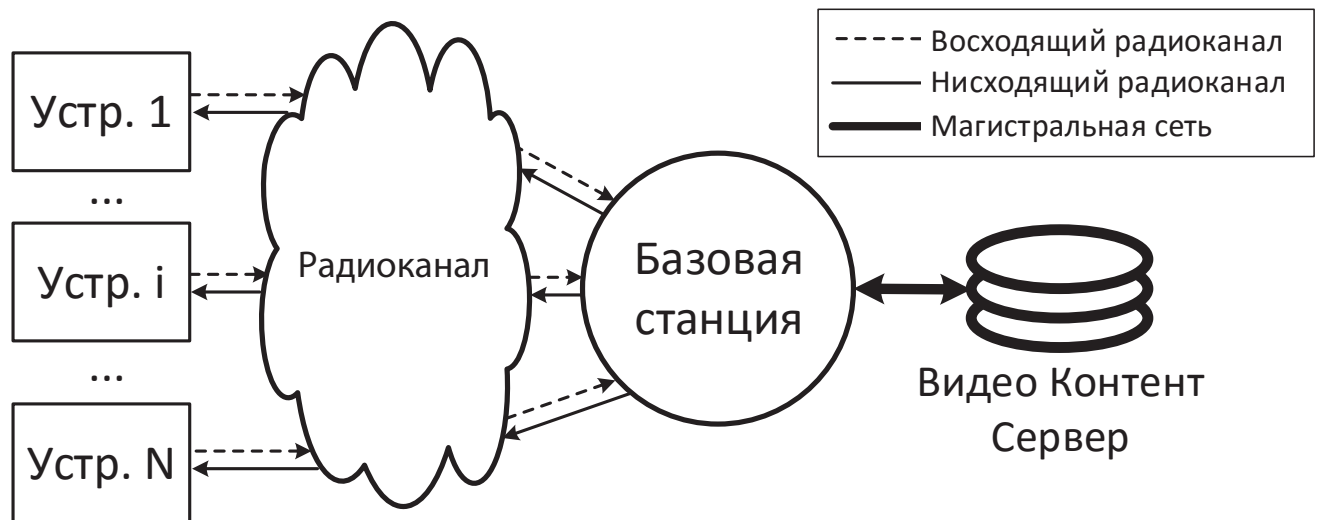


Рисунок 2.6 — Структура модели передачи видеоданных в централизованных беспроводных сетях

– *Формат представления видеоданных:*

1. Все видеопоследовательности разбиты на сегменты одинаковой длительности d секунд;
2. Каждый сегмент видеоданных j пользователя i представляется в виде последовательности из $K_{i,j}$ пакетов равного объема;
3. Все видеопоследовательности на видео контент сервер представлены в непрерывном отрезке битовых скоростей (2.1);
4. Длительности видеопоследовательностей m , представленных на видео сервере, являются независимыми случайными величинами с конечными математическим ожиданием и дисперсией.

– *Общие характеристики сети передачи информации:*

1. Задержка в восходящем канале считается пренебрежимо малой;

2. Загрузка сегментов видеоданных начинается незамедлительно в момент отправки запроса.
- *Модель поведения пользователя при просмотре видео:*
1. Паузы между просмотрами τ_i является независимыми случайными величинами с конечными математическим ожиданием ($\bar{\tau}_i$) и дисперсией.
- *Модель беспроводного канала:*
1. Затухание сигнала при распространении происходит одинаково по всей ширине полосы передачи информации для конкретного пользователя в одном моменте времени;
 2. Кодо-модуляционная схема подбирается таким образом, что вероятность ошибки при передаче информации по беспроводному каналу пренебрежимо мала;
 3. В течении передачи одного пакета k из сегмента j пользователем i в нисходящей линии связи максимально достижимая скорость канала постоянна (2.2);
 4. Последовательность случайных величин $C_{i,1}^{-1}, C_{i,2}^{-1}, \dots, i = \overline{1, N}$ формирует эргодический случайный процесс с конечными математическим ожиданием $E[C_i^{-1}]$ и коэффициентом вариации ν_i^C .
- *Модель планировщика ресурсов беспроводного канала:*
1. Алгоритм планирования выделяет частотно-временные ресурсы только активным абонентам;
 2. Алгоритм планирования обеспечивает каждому активному абоненту минимальную долю ресурсов беспроводного канала в каждом интервале планирования (2.6);
 3. Алгоритм планирования распределяет все доступные частотно-временные ресурсы (2.7).
- *Модель видеоплеера:*
1. Последовательность $R_{i,1}, R_{i,2}, \dots, i = \overline{1, N}$ формируют эргодический случайный процесс с конечным математическим ожиданием $E[R_i]$ и коэффициентом вариации ν_i^R (2.8).

Данная система допущений завершает представление аналитической модели.

2.8 Система передачи видеоданных как система массового обслуживания

Представленная выше аналитическая модель передачи видеоданных описывает достаточно современные тенденции развития телекоммуникационных сетей связи: передача видео и беспроводные сети. Несмотря на данный факт аналогичные системы рассматривались еще в 70-х годах 20-го века, они известны как замкнутые системы массового обслуживания (СМО) с конечным числом абонентов. Далее предлагается краткое описание таких СМО и анализ отличий данных систем от рассматриваемой модели в настоящей работе.

Изначально, СМО с конечным числом абонентов были рассмотрены Шерром в работе [24], в дальнейшем замкнутые СМО с конечным числом абонентов были широко исследованы в работах [25–27]. Общая идеология таких систем состоит в следующем. Абонент отправляет на обслуживающий прибор заявку, которая становится в очередь бесконечного размера. Заявки всех абонентов считаются одинаковыми. Обслуживающий прибор, исходя из зафиксированной дисциплины обслуживания, обрабатывает заявки в очереди. Когда заявка для некоторого абонента была обслужена, то по петле обратной связи для этого абонента будет доставлено уведомление об окончании выполнения заявки. После получения уведомления абонент отправляет следующую заявку в систему через некоторую паузу. В классических СМО с конечным числом абонентов время между получением уведомления об обработке заявки до посылки нового запроса в систему и длительность обслуживания заявки в очереди распределено по экспоненциальному закону.

Важной особенностью подобных систем является то, что в каждый момент времени в очереди на обслуживающем приборе количество заявок не превосходит количество абонентов в сети. Данный факт приводит к тому, что среднее время обслуживания абонентов до точки насыщения системы растет незначительно, а после нее растет линейно, с углом наклона зависящего от интенсивности обслуживания и временем между получением уведомления и отправкой новой заявки. Такое поведение среднего времени обслуживания отличается от классических СМО без обратной связи, в которых оно растет экспоненциально после достижения системой точки насыщения, в виду бесконечного роста количества заявок в очереди на обслуживающем приборе.

Опишем систему передачи видеоданных в терминах замкнутых СМО с конечным числом абонентов. Далее будет описана общая логика работы систем передачи видеоданных, в скобках будет представлено название в терминах СМО (рисунок 2.7).

Воспроизводящие устройства (*абоненты*) по восходящему каналу связи отправляют запрос на сегмент видеопоследовательности к контент серверу. Сервер, в свою очередь, отправляет воспроизводящему устройству сегмент видео (*заявку*). Важно отметить, что одновременно одно устройство (*абонент*) может загружать только один сегмент (*заявку*). Все сегменты видеоданных попадают в очереди на базовой станции (*обслуживающем устройстве*). Далее, в соответствии с алгоритмом планирования распределением ресурсов (*дисциплиной обслуживания*), сегменты (*заявки*) из очередей передаются на пользовательское устройства. После получения сегмента воспроизводящее устройство отправит через некоторую паузу новый запрос на контент сервер.

В данной системе запрос на сегмент видеоданных выполняет роль петли обратной связи. В предложенной модели планировщик на базовой станции в каждый момент времени распределяет доли ресурсов канала между активными абонентами. Распределение долей ресурсов канала возможно интерпретировать как разделение времени обслуживания активных абонентов. Следовательно, базовая станция может быть представлена в виде процессора с разделением времени.

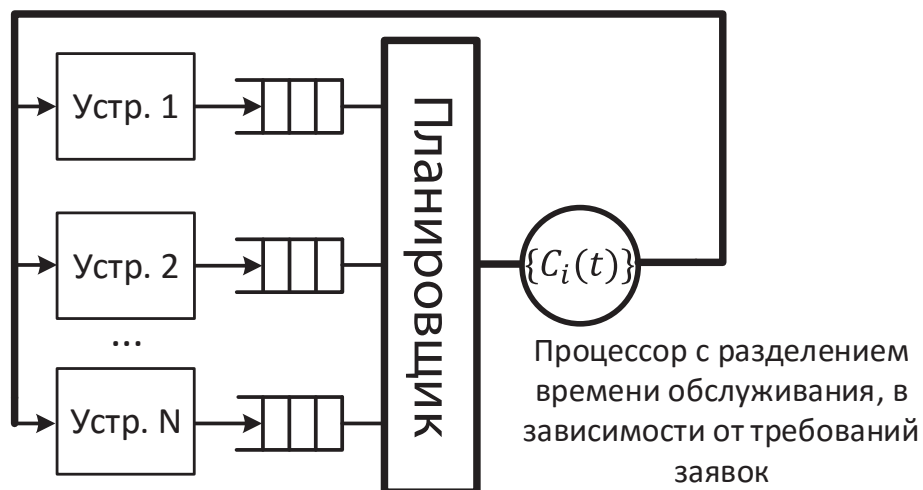


Рисунок 2.7 — Система передачи видеоданных как замкнутая система массового обслуживания

Однако, существует несколько важных отличий систем передачи видеоданных от замкнутых СМО с конечным числом абонентов:

- Размер загружаемых сегментов видеоданных изменяется во времени и зависит от скорости загрузки информации, и как следствие, алгоритма планирования;
- Длительность интервала времени между получением сегмента и отправкой нового запроса не распределено по экспоненциальному закону;
- Время передачи сегмента видео через базовую станцию (*время обслуживания заявки*) зависит от его размера и характеристик беспроводного канала всех активных абонентов в текущий момент времени (изменяемая интенсивность обслуживания во времени с дополнительной зависимостью от предыстории).

Данные отличия приводят к невозможности анализа системы передачи видеоданных методами, используемых при анализе классических СМО.

2.9 О взаимосвязи параметров системы передачи видеоданных

Представленная ранее аналитическая модель описывает важнейшие компоненты и параметры реальной системы передачи видеоданных. Основываясь на данной модели вводится утверждение 2.1, описывающая основные взаимосвязи между параметрами системы.

Утверждение 2.1. *Для всевозможных алгоритмов планирования, удовлетворяющих допущениям, представленных в подразделе 2.5, следующее неравенство истинно:*

$$\sum_{i=1}^N (1 - \nu_i^R \nu_i^C) \frac{E[R_i] E[C_i^{-1}]}{q_i + \gamma_i} \leq 1. \quad (2.9)$$

Доказательство. Рассмотрим процессы в системе передачи видеоданных, когда некоторый пользователь i загружает пакет k из сегмента j . В течении загрузки пакета через систему был передан следующий объем информации $\frac{R_{i,j}d}{K_{i,j}}$, который может быть вычислен исходя из наблюдения за скоростью передачи информации

в беспроводном канале:

$$\int_{t_{i,j,k}}^{t_{i,j,k}+\Delta t_{i,j,k}} S_i(t)dt = C_{i,j,k} \int_{t_{i,j,k}}^{t_{i,j,k}+\Delta t_{i,j,k}} \alpha_i(t)dt = \frac{R_{i,j}d}{K_{i,j}}.$$

Перенесем $C_{i,j,k}$ в правую часть последнего равенства и получим долю времени использования канала пользователем i при загрузке k -го пакета:

$$\int_{t_{i,j,k}}^{t_{i,j,k}+\Delta t_{i,j,k}} \alpha_i(t)dt = \frac{R_{i,j}d}{K_{i,j}} \frac{1}{C_{i,j,k}}.$$

Просуммируем левую и правую часть по всем пакетам в сегменте, получим долю времени использования канала пользователем i при загрузке j -го сегмента:

$$\int_{t_{i,j}}^{t_{i,j}+\Delta t_{i,j}} \alpha_i(t)dt = R_{i,j}d \left(\frac{1}{K_{i,j}} \sum_{k=1}^{K_{i,j}} \frac{1}{C_{i,j,k}} \right),$$

где $t_{i,j}$ – время начала загрузки пользователем i сегмента j , $\Delta t_{i,j}$ – длительность загрузки пользователем i сегмента j . Согласно введенным ранее обозначениям (подраздел 2.4) $C_{i,j}^{-1} = K_{i,j}^{-1} \sum_{k=1}^{K_{i,j}} C_{i,j,k}^{-1}$, тогда выражение примет вид:

$$\int_{t_{i,j}}^{t_{i,j}+\Delta t_{i,j}} \alpha_i(t)dt = R_{i,j}C_{i,j}^{-1}d.$$

Тогда доля времени использования канала пользователем i за время T :

$$\int_0^T \alpha_i(t)dt = d \sum_{j=1}^{H_i^T} R_{i,j}C_{i,j}^{-1} + \Delta s_i, \quad (2.10)$$

где H_i^T – число полностью загруженных сегментов пользователем i в течении времени T , Δs_i – длительность видео, загруженного пользователем i , но не воспроизведенный плеером. Просуммировав уравнение (2.10) для всех пользователей в системе ($i = \overline{1, N}$) и внося сумму под знак интеграла было получено следующее

выражении:

$$\int_0^T \sum_{i=1}^N \alpha_i(t) dt = \sum_{i=1}^N d \sum_{j=1}^{H_i^T} R_{i,j} C_{i,j}^{-1} + \Delta s,$$

где $\Delta s = \sum_{i=1}^N \Delta s_i$.

Так как в каждый момент времени t сумма выделенных ресурсов не может превышать единицы, согласно условию (2.7), следовательно $\int_0^T \sum_{i=1}^N \alpha_i(t) dt \leq T$:

$$d \sum_{i=1}^N \sum_{j=1}^{H_i^T} R_{i,j} C_{i,j}^{-1} + \Delta s \leq T.$$

Разделим левую и правую часть неравенства на T и внесем его под первый знак суммы, так как $T > 0$ знак неравенства не изменится.

$$\sum_{i=1}^N \frac{d \sum_{j=1}^{H_i^T} R_{i,j} C_{i,j}^{-1} + \Delta s_i}{T} \leq 1.$$

Для каждого пользователя время пребывания в системе состоит из буферизаций, просмотров и пауз между просмотрами видео: $T = b_i^T + w_i^T + p_i^T$, следовательно последнее неравенство представляется в следующем виде:

$$\sum_{i=1}^N \frac{d \sum_{j=1}^{H_i^T} R_{i,j} C_{i,j}^{-1} + \Delta s_i}{b_i^T + w_i^T + p_i^T} \leq 1.$$

Так как длительность просмотра равняется длительности всех загруженных сегментов, $w_i^T = H_i^T d, i = \overline{1, N}$:

$$\sum_{i=1}^N \frac{w_i^T \frac{1}{H_i^T} \sum_{j=1}^{H_i^T} R_{i,j} C_{i,j}^{-1} + \Delta s_i}{b_i^T + w_i^T + p_i^T} \leq 1.$$

Разделив числитель и знаменатель левой части полученного выражения на w_i^T и устремим время в бесконечность ($T \rightarrow \infty$). Для каждого абонента $i = \overline{1, N}$

воспользуемся определением q_i и получим следующее выражение:

$$\sum_{i=1}^N \frac{E[R_i C_i^{-1}]}{q_i + \gamma_i} \leq 1.$$

Данный переход возможен так как последовательности $R_{i,1}, R_{i,2}, \dots$ и $C_{i,1}, C_{i,2}, \dots, i = \overline{1, N}$ формируют эргодические случайные процессы с конечными первыми двумя моментами (подраздел 2.7).

Рассмотрим более подробно значение, получившее в числителе: $E[R_i C_i^{-1}]$. Важно отметить, что битовая скорость потока R_i для каждого пользователя выбирается в зависимости от оценки качества канала, которая определяется максимально достижимой скоростью канала C_i . Следовательно, значения R_i и C_i являются зависимыми величинами, и математическое ожидание их произведения:

$$E[R_i C_i^{-1}] = E[R_i] E[C_i^{-1}] + r(R_i, C_i^{-1}) \sigma[R_i] \sigma[C_i^{-1}],$$

где $r(R_i, C_i^{-1})$ является коэффициентом корреляции между величинами R_i и C_i^{-1} . По определению значение $r(R_i, C_i^{-1})$ принимает значения в отрезке $[-1, 1]$, следовательно:

$$E[R_i] E[C_i^{-1}] - \sigma[R_i] \sigma[C_i^{-1}] \leq E[R_i C_i^{-1}].$$

Исходя из данного факта:

$$\sum_{i=1}^N (1 - \nu_i^R \nu_i^C) \frac{E[R_i] E[C_i^{-1}]}{q_i + \gamma_i} \leq 1.$$

Таким образом, утверждение 2.1 доказано. □

Результат, представленный в утверждении 2.1, описывает взаимосвязь между всеми характеристиками сети передачи видеоданных от особенностей модели поведения пользователя (γ_i, q_i) и характеристик видеопотока $(E[R_i])$, до аспектов работы радиоканала $(E[C_i^{-1}])$ для всех пользователей в сети. Полученное выражение возможно использовать в качестве ограничения на достижимый объем ресурсов радиоканала при оптимизации передачи видеоданных в беспроводных централизованных сетях.

В дальнейшем, для упрощения математических выкладок в качестве обозначений $E[R_i]$ и $E[C_i^{-1}]$ будут использоваться конструкции \tilde{R}_i и \tilde{C}_i^{-1} соответственно.

2.10 Выводы по разделу

В данном разделе была предложена модель передачи видеоданных в современных централизованных беспроводных сетях. Предложенная модель описывает все важнейшие компоненты существующих сетей: магистральная и опорная сети передачи информации, базовая станция (структура беспроводного канала и алгоритм распределения его ресурсов) модель видеотрафика (формат представления видео и аспекты работы воспроизводящего устройства). Так же модель системы передачи видеоданных в беспроводных централизованных сетях была рассмотрена как замкнутая система массового обслуживания с конечным числом абонентов. В завершении раздела был представлен результат, описывающий основополагающие взаимосвязи между всеми характеристиками сети передачи видеоданных.

Основные результаты раздела возможно сформулировать следующим образом:

- Предложена модель передачи видеоданных в беспроводных централизованных сетях, включающая в себя:
 - Модель опорной и магистральной сети передачи информации (подраздел 2.3);
 - Модель беспроводного канала связи (подраздел 2.4);
 - Модель работы базовой станции и планировщика ресурсов беспроводного канала (подраздел 2.5);
 - Модель воспроизводящего устройства и видеотрафика (подраздел 2.6);
 - Систему допущений (подраздел 2.7).
- Проведен анализ беспроводных систем перечеи видео как замкнутых систем массового обслуживания с конечным числом абонентов (подраздел 2.8);

- Представлена взаимосвязь между всеми всеми характеристиками сети передачи видеоданных (подраздел 2.9).

3 ОПТИМИЗАЦИЯ ПЕРЕДАЧИ НЕАДАПТИВНЫХ ВИДЕОПОТОКОВ В БЕСПРОВОДНЫХ ЦЕНТРАЛИЗОВАННЫХ СЕТЯХ

3.1 Вводные замечания

В настоящее время в беспроводных сетях, несмотря на бурное развитие технологий адаптивной передачи видеоконтента, достаточно распространена передача видео с использованием неадаптивных видеоплееров. Наиболее ярким примером является социальная сеть Facebook, весь видеоконтент которой представлен в двух репрезентациях: стандартном и высоком разрешениях, в англоязычной терминологии Standard Definition и High Definition соответственно. Пользователь, при просмотре видео, может выбрать один из двух представленных выше вариантов качества видеопотока и оно не будет изменяться во времени. Подобная модель поведения абсолютно соответствует описанию неадаптивного видеоплеера, представленного в подразделе 1.4.

В настоящем разделе производится аналитическое исследование производительности беспроводных централизованных систем передачи информации с доминированием передачи видеоданных по протоколу прикладного уровня НТТР. В начале раздела предлагается критерий качества обслуживания для неадаптивных видеопотоков. Далее вводятся дополнительные допущения, специфичные для неадаптивных видеопоследовательностей. Затем формулируется оптимизационная задача и предлагается нижняя граница для введенного критерия качества обслуживания. В завершении предлагается численный пример, демонстрирующий производительность стандартных алгоритмов планирования распределения ресурсов беспроводного канала по сравнению с полученной нижней границей.

Основные результаты данного раздела опубликованы в работах [28; 29].

3.2 Критерий качества восприятия неадаптивного видеопотока

Обобщая информацию, представленную ранее в настоящей работе, можно выделить три основных фактора, влияющих на восприятие видеопоследовательности:

- Репрезентация (рекомендуемое разрешение экрана и частота кадров видео), определяемая битовой скоростью потока (подраздел 1.5).
- Прерывание воспроизведения, вызванное буферизацией видеопотока, включающее начальную буферизацию видео (подраздел 1.5);
- Гладкость воспроизведения видеопоследовательности – частота изменения битовой скорости во времени (подраздел 2.6).

Произведем рассмотрение каждого из перечисленных выше факторов с точки зрения неадаптивных видеопоследовательностей.

При работе неадаптивного видеоплеера решение о выборе репрезентации видеоряда принимает пользователь некоторым образом, отвечающим индивидуальным требованиям к качеству просматриваемого видео и не изменяет его во времени. На основе факта о неизменности репрезентации видео во времени, следует что гладкость воспроизведения, которая характеризует частоту изменения репрезентации во времени, не влияет на качество восприятия видеоряда.

Таким образом, на удовлетворенность пользователя от просмотра неадаптивных видеопоследовательностей влияют прерывания воспроизведения, вызванные опустошением буфера и начальной буферизацией на пользовательском устройстве.

Очевидно, что если в соте находится большое число абонентов, то могут проявиться эффекты деградации воспроизведения видео у пользователей, ввиду конечности количества частотно-временных ресурсов беспроводного канала и ограниченности снизу минимальной битовой скорости репрезентации видео. Например, если некоторому пользователю базовая станция обеспечивает скорость загрузки информации равную 0.25 Мбит/с, а битовая скорость минимальной репрезентации равняется 0.5 Мбит/с, то у данного пользователя при просмотре видео с высокой вероятностью будет длительная начальная буферизация и неизбежно возникнут прерывания воспроизведения видео. Основываясь на данном факте введем определение перегрузки:

Определение 3.1. Перегрузка – состояние системы передачи информации, характеризующее недостаток ресурсов для обеспечения достаточного уровня удовлетворенности всех активных пользователей.

Как было представлено в подразделе 2.3, все время T нахождения пользователя в системе разделено на три: буферизация (b_i^T), просмотр (w_i^T) и паузы между просмотрами видео (p_i^T). В настоящем разделе предлагается оценивать качество восприятия видеопотока для пользователя как нормированное отношение длительности буферизации к просмотру за время T :

Определение 3.2. Нормированное отношение длительности буферизации к просмотру для пользователя i – это отношение общей длительности ожидания к длительности времени нахождения пользователя в системе, исключая паузы между просмотрами видео:

$$g_i = \lim_{T \rightarrow \infty} \frac{b_i^T}{w_i^T + b_i^T}. \quad (3.1)$$

Критерий восприятия (3.1) описывает процентное соотношение общих длительностей буферизации и просмотра для пользователя i за время T . На основе данного критерия возможно оценить удовлетворенность пользователя следующим образом:

$$g_i = \begin{cases} 0, & \text{пользователь } i \text{ удовлетворен просмотром} \\ (0,1) & \text{пользователь } i \text{ наблюдает негативные эффекты при просмотре} \\ 1, & \text{пользователь } i \text{ полностью неудовлетворен просмотром} \end{cases}$$

Если пользователь не ожидал в течение времени ($b_i^T = 0$), тогда значение $g_i = 0$. Если пользователь, за время T , не просмотрел ни одной секунды видео ($w_i^T = 0$), то значение g_i будет равно 1. Изначально критерий 3.2 был предложен в работе [30], в англоязычной литературе данный критерий известен под названием «*rebuffering percentage*». Термин «*rebuffering*» в англоязычной литературе используется для обозначения времени ожидания воспроизведения в течении просмотра видео при передаче с использованием протокола HTTP.

В качестве критерия качества работы системы в целом предлагается использовать среднее значение отношение длительности буферизации к просмотру по

всем пользователям, присутствующих в системе: $\bar{g}(\mathcal{A})$. Важно отметить, что значение критерия качества обслуживания имеет прямую зависимость от алгоритма планирования распределения ресурсов беспроводного канала $\mathcal{A}(\cdot)$.

После определения критерия качества обслуживания в дальнейшем будет произведена постановка оптимизационной задачи.

3.3 Постановка оптимизационной задачи

В настоящем подразделе будет продолжено рассмотрение критерия качества обслуживания для неадаптивных видеопотоков, введенного в подразделе 3.2. На основе проведенного рассмотрения будет предложена постановка оптимизационной задачи, решение которой характеризует максимальную возможную производительность алгоритма планирования при передаче неадаптивных видеопотоков.

Рассмотрим критерий качества для неадаптивного видео: нормированного отношения длительности буферизации к просмотру g_i (3.2). По определению, удовлетворенность пользователя имеет обратную корреляцию с данным критерием: чем меньше значение g_i , тем выше удовлетворенность пользователя i . Как следствие, основной целью увеличения производительности системы является поиск минимального значения \bar{g} по всем возможным алгоритмам планирования $\mathcal{A}(\cdot)$, удовлетворяющих введенным допущениям:

$$\bar{g}(\mathcal{A}) = \frac{1}{N} \left(\sum_{i=1}^N g_i(\mathcal{A}) \right).$$

Однако, множество возможных алгоритмов планирования, удовлетворяющих допущениям, является бесконечным и несчетным, поэтому вместо поиска минимального значения в данном разделе будет найдена нижняя граница (инфинум) по всем возможным алгоритмам $\mathcal{A}(\cdot)$:

$$G = \inf_{\mathcal{A}} \bar{g}(\mathcal{A}). \quad (3.2)$$

Ранее в подразделе 2.3 было введено определение для q_i : отношения времени буферизации к просмотру для пользователя i (определение 2.2). Покажем, каким образом взаимосвязаны значения g_i и q_i . По определению данных величин, если продолжительность буферизации пользователя i за время T (b_i^T) равнялась нулю, то $q_i = g_i = 0$. Пусть значение (b_i^T) отлично от нуля, тогда значения q_i и g_i взаимосвязаны следующим образом:

$$q_i = \frac{g_i}{1 - g_i}.$$

На основе представленной выше взаимосвязи значений критериев качества обслуживания выражение (2.9), полученное в результате доказательства Утверждения 2.1 (подраздел 2.9) примет следующий вид:

$$\sum_{i=1}^N (1 - \nu_i^R \nu_i^C) E[R_i] E[C_i^{-1}] \frac{1 - g_i}{g_i + \gamma_i(1 - g_i)} \leq 1. \quad (3.3)$$

При просмотре неадаптивных видеопоследовательностей пользователь i в начальный момент времени выбирает репрезентацию с битовой скоростью R_i , качество которой его полностью удовлетворяет, и не изменяет ее в течении всего времени работы. Данный факт приводит к изменению допущения *Модели видеоплеера* номер 1, представленному в подразделе 2.7, следующим образом:

1. Значение битовой скорости видеопотока для каждого конкретного абонента неизменно во времени: $R_{i,1} = R_{i,2} = \dots = R_i, i = \overline{1, N}$.

Следствием из данного допущения является равенство коэффициента вариации $\nu_i^R, i = \overline{1, N}$ нулю, а математическое ожидания битовой скорости $E[R_i]$ равно R_i для каждого пользователя i . Таким образом, неравенство (3.3) принимает следующий вид:

$$\sum_{i=1}^N R_i E[C_i^{-1}] \frac{1 - g_i}{g_i + \gamma_i(1 - g_i)} \leq 1.$$

Для уменьшения объема последующих математических преобразований введем обозначение K_i как $R_i E[C_i^{-1}]$. Важно отметить, что значение K_i не является переменной, так как его компоненты задаются независимыми от оптимизации факторами: пожеланиями пользователя (R_i) и условиями беспроводного канала ($E[C_i^{-1}]$).

В дополнении к системе допущений *Модели поведения пользователя при просмотре видео*, введенному в подразделе 2.7, в рамках настоящего раздела вводится дополнительное допущение:

2. Коэффициент разряженности видеопотока γ_i (определение 2.1) для всех пользователей имеет равное значение: $\gamma_i = \gamma, i = \overline{1, N}$.

Настоящее допущение обусловлено зависимостью в реальной системе между длительностью видеоролика и паузы после его просмотра: после просмотра длительного ролика пользователь «осмысливает» его более длительный промежуток времени, чем при просмотре короткого ролика. Данный факт приводит приблизительно к равенству коэффициентов разряженности потока у всех пользователей.

Обобщая сказанное выше, в настоящем разделе предлагается оптимизационная задача следующего вида:

$$\begin{aligned} &\text{Минимизировать: } G = \sum_{i=1}^N g_i \\ &\text{При условии:} \\ &\begin{cases} \sum_{i=1}^N \frac{K_i(1-g_i)}{g_i + \gamma(1-g_i)} \leq 1 \\ g_i \geq 0, i = \overline{1, N} \\ g_i \leq 1, i = \overline{1, N} \end{cases} \end{aligned} \quad (3.4)$$

Оптимизационная задача (3.4) является задачей нелинейного программирования, ввиду нелинейности первого ограничения. Дальнейшая часть данного раздела посвящена поиску оптимального решения задачи (3.4), значение которого задает нижнюю границу по всем возможным алгоритмам планирования, удовлетворяющих допущениям введенных в подразделе 2.7.

3.4 Решение обобщенной задачи о непрерывном рюкзаке

Отыскание решения оптимизационной задачи (3.4) является сложной задачей, и для ее решения необходимо проделать несколько шагов. Изначально будет рассмотрено обобщение известной задачи о непрерывном рюкзаке и найдено его

решение. Далее будет показано, каким образом найденное решение может быть использовано при решении оптимизационной задачи (3.4). В данном подразделе рассматривается задача о непрерывном рюкзаке, ее обобщение и находится находится решение обобщенной задачи о непрерывном рюкзаке.

Представленное ниже решение опубликовано в работе [29].

Задача о непрерывном рюкзаке в англоязычной литературе известна под термином «*fractional knapsack problem*», и подробно описана в [31]. Задача формулируется следующим образом. Имеется набор из N предметов, каждый предмет i характеризуется двумя величинами: стоимостью (c_i) и весом (w_i). Также имеется рюкзак неограниченного объема, характеризующийся с весом предметов, помещенных в него. Считается, что человек, который будет нести этот рюкзак, не может поднять вес превышающий W . Все предметы считаются делимыми на бесконечное число частей, что позволяет выбрать долю каждого предмета x_i , которую он положит в рюкзак. В качестве задачи, предлагается найти такой вектор (x_1, x_2, \dots, x_N) , что вес взятых долей предметов не превышает W и при этом стоимость взятых предметов максимальна. Математическое представление задачи о непрерывном рюкзаке представлено системой (3.5):

$$\begin{aligned} &\text{Максимизировать: } \sum_{i=1}^N c_i x_i \\ &\text{При условии:} \\ &\left\{ \begin{array}{l} \sum_{i=1}^N w_i x_i \leq W \\ W \in (0, \infty) \\ x_i \in [0, 1], \quad i = \overline{1, N} \\ w_i \in (0, \infty), \quad i = \overline{1, N} \end{array} \right. . \end{aligned} \quad (3.5)$$

Таким образом задача о непрерывном рюкзаке относится к классу непрерывных задач линейного программирования от нескольких переменных.

Важным отличием задачи о непрерывном рюкзаке (3.5), от классической задачи о рюкзаке («*0-1 knapsack problem*») заключается в возможности деления предметов, когда в классической задаче предметы считаются неделимыми (величина x_i может принимать только два значения: 0 и 1). Известно, что задача о рюкзаке не имеет замкнутого решения и может быть решена методами динамическо-

го программирования. Однако, для задачи о непрерывном рюкзаке оптимальное решение может быть получено «жадным алгоритмом»:

Алгоритм 3.1 : Решение задачи о непрерывном рюкзаке

- 1: Сортировка предметов по возрастанию отношений $\frac{w_i}{c_i}$, и их нумерация, в соответствие с сортировкой;

2: Нахождение предмета с индексом j :
$$\begin{cases} \sum_{i=1}^{j-1} w_i \leq W \\ \sum_{i=1}^j w_i > W \end{cases}$$

Если $j \in \emptyset, j = N + 1$;

- 3: Нахождение оптимальных значений x_i :

$$x_i = \begin{cases} 1, & i < j \\ \frac{1}{w_j} \left(W - \sum_{i=1}^{j-1} w_i \right), & i = j. \\ 0, & i > j \end{cases}$$

Дальнейшее повествование в настоящем подразделе будет построено следующим образом. Первым шагом, будет представлено обобщение задачи о непрерывном рюкзаке. Вторым шагом будет показано каким образом алгоритм 3.1 может быть применен для решения обобщенной задачи о непрерывном рюкзаке. В завершении подраздела будет представлено сравнение сложности полученного решения со стандартными методами решения задач выпуклого программирования от нескольких переменных.

Обобщение задачи о непрерывном рюкзаке:

Максимизировать: $\sum_{i=1}^N f(x_i)$

При условии:

$$\begin{cases} \sum_{i=1}^N w_i x_i \leq W \\ W \in (0, \infty) \\ x_i \in [0, 1], \quad i = \overline{1, N} \\ w_i \in (0, \infty), \quad i = \overline{1, N} \end{cases}, \quad (3.6)$$

где $f(x_i)$ — непрерывная, дважды дифференцируемая, строго монотонно возрастающая, выпуклая вниз функция, ограниченная сверху и снизу в области определения x_i .

Оптимизационная задача (3.6) относится к классу выпуклого программирования от нескольких переменных с ограничениями типа неравенства. Важно отметить факт, что задачи выпуклого программирования являются подклассом задач нелинейного программирования.

Решение оптимизационной задачи (3.6) описывается в утверждении 3.1.

Утверждение 3.1. Если в алгоритме 3.1 принять $\forall i : c_i = 1$, то вектор $X = (x_1, x_2, \dots, x_N)$, полученный в результате выполнения алгоритма 3.1, является оптимальным решением задачи нелинейного программирования (3.6).

Доказательство. В основе предложенного далее доказательства был использован метод от противного. Предположим, что существует оптимальное решение $Y = (y_1, y_2, \dots, y_N)$, удовлетворяющее всем ограничениям системы (3.6), отличное от решения $X = (x_1, x_2, \dots, x_N)$, полученного на основе алгоритма 3.1 (рисунок 3.1). Важно отметить, что нумерация всех предметов происходит в порядке сортировки на шаге 1 алгоритма 3.1. Тогда в решении Y найдется пара индексов k и l такие что:

- k — наименьший индекс предмета, взятая в рюкзак доля которого меньше единицы: $y_k < 1$;
- l — наименьший индекс предмета, больший k , взятая в рюкзак доля которого отлична от нуля: $y_l > 0, l > k$.

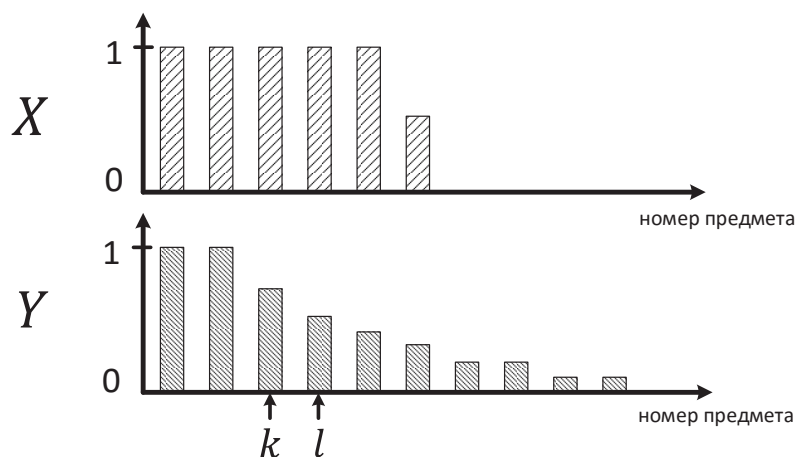


Рисунок 3.1 — Вид решений X и Y

Покажем, каким образом возможно перераспределить доли взятых предметов y_k и y_l под индексами k и l соответственно, для максимизации целевой функции системы (3.6).

Для этого введем следующие обозначения:

- w_s – суммарный вес взятых предметов k и l :

$$w_s = y_k w_k + y_l w_l.$$

Так как значение $y_k < 1$ и $y_l \in (0, 1]$, то $w_s < w_k + w_l$.

- Обновленное решение Y^* :

$$Y^* = (y_1, \dots, y_{k-1}, y_k^*, y_{k+1}, \dots, y_{l-1}, y_l^*, y_{l+1}, \dots),$$

где y_k^* и y_l^* – обновленные доли предметов под индексами k и l соответственно.

Решения Y и Y^* удовлетворяют ограничениям оптимизационной задачи (3.6) и отличаются только в долях предметов под индексами k и l . Далее в настоящем доказательстве, будет показано, что за счет преобразования долей y_k и y_l (решение Y) в y_k^* и y_l^* (решение Y^*) соответственно, и без изменения остальных значений долей предметов возможно увеличить значение целевой функции задачи (3.6). То есть будет показано, что для любого решения Y найдется решение Y^* , отличное в позициях k и l , такое что целевая функция при решении Y^* больше, чем при Y .

Для этого поставим промежуточную оптимизационную задачу:

Максимизировать: $f(y_k^*) + f(y_l^*)$

При условии:

$$\begin{cases} w_s = y_k^* w_k + y_l^* w_l \\ w_s < w_k + w_l \\ w_k < w_l \\ y_k^*, y_l^* \in [0, 1] \end{cases} \quad (3.7)$$

В промежуточной оптимизационной задаче (3.7) целевая функция состоит из суммы функций $f(\cdot)$ для долей предметов k и l , так как решение Y^* отлично от Y только в данных позициях. Как следствие, значения остальных элементов

суммы являются константами и не влияют на значение целевой функции. В ограничениях задачи (3.7) указывается, что суммарный вес предметов k и l не может быть изменен, а значит не будет нарушено ограничение на суммарный вес всех предметов взятых в рюкзак в системе (3.6).

Таким образом, целью промежуточной оптимизационной задачи (3.7) является отыскание таких y_k^* и y_l^* , при которых целевая функция задачи (3.7) принимает максимальное значение. Важно отметить, что максимизация целевой функции промежуточной задачи (3.7) приводит и к максимизации целевой функции в обобщенной задаче о непрерывном рюкзаке (3.6). Следствием этого является факт, что если решение Y может быть улучшено в промежуточной оптимизационной задаче, то оно не является оптимальным решением задачи (3.6).

Продолжим рассмотрение промежуточной оптимизационной задачи. Из системы (3.7) выразим y_k^* как функцию от y_l^* :

$$y_k^* = \frac{w_s - y_l^* w_l}{w_k}.$$

Подставим полученный результат в систему (3.7):

$$\text{Максимизировать: } f\left(\frac{w_s - y_l^* w_l}{w_k}\right) + f(y_l^*)$$

При условии:

$$\begin{cases} y_k^* = \frac{w_s - y_l^* w_l}{w_k} \\ w_s < w_k + w_l \\ w_k < w_l \\ y_k^*, y_l^* \in [0, 1] \end{cases} \quad (3.8)$$

Найдем вторую производную целевой функции задачи (3.8):

$$\left[f\left(\frac{w_s - y_l^* w_l}{w_k}\right) + f(y_l^*) \right]'' = f''\left(\frac{w_s - y_l^* w_l}{w_k}\right) \left(\frac{w_l}{w_k}\right)^2 + f''(y_l^*).$$

Так как функция $f(x)$ является выпуклой вниз, то $f''(x) > 0$. Следовательно выражение в правой части является положительным. Обобщая сказанное выше, экстремумом оптимизационной задачи (3.8), если он существует, является минимум, как следствие, максимальное значение целевая функция принимает на граничных значениях y_l^* (рисунок 3.2).

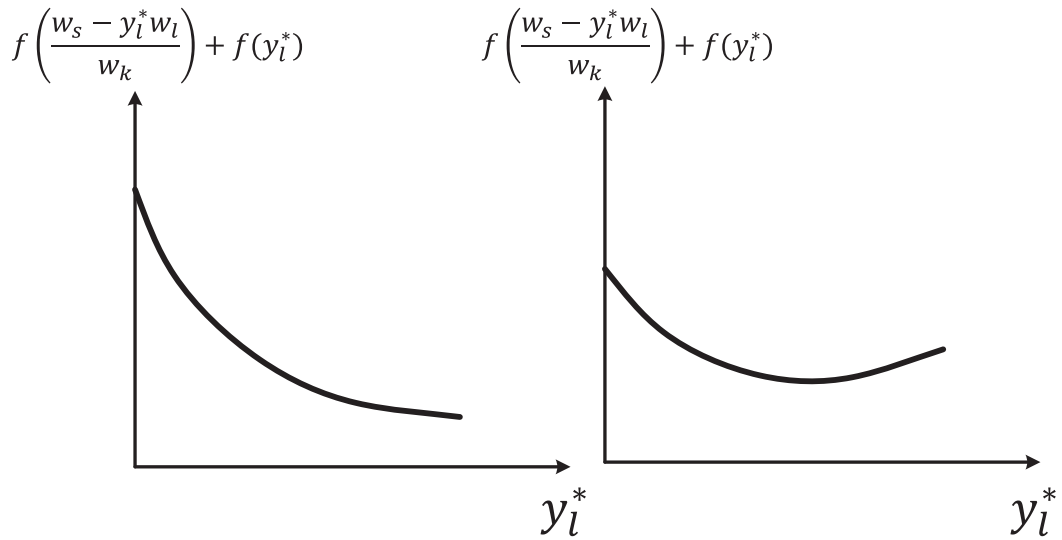


Рисунок 3.2 — Вид целевой функции промежуточной оптимизационной задачи

Будем рассматривать различные ситуации, в зависимости от значения w_s . Разобьем все возможные значения w_s на три интервала (рисунок 3.3):

- $I: w_s \in (0, w_k]$;
- $II: w_s \in (w_k, w_l]$;
- $III: w_s \in (w_l, w_k + w_l)$.

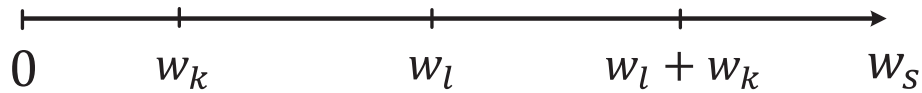


Рисунок 3.3 — Возможные значения w_s

Для каждого перечисленного выше интервала построим граничные значения целевой функции (3.8).

Интервал I:

Построим граничные значения для целевой функции оптимизационной задачи (3.8), при $w_s \in (0, w_k]$. Для этого необходимо соблюсти все ограничения оптимизационной задачи. Важным ограничением является $y_k^*, y_l^* \in [0, 1]$. Для обеспечения $y_k^* \geq 0$ необходимо, чтобы значение y_l^* не превышало $\frac{w_s}{w_l}$. Ограниченность значения w_s величиной w_k приводит невозможности превышения значениями y_l^* и y_k^* единицы.

Основываясь на данных рассуждениях, y_l^* может принимать значения на отрезке $\left[0, \frac{w_s}{w_l}\right]$, и граничные значения могут быть представлены следующим образом:

– Граничное значение $I.1$:

$$\begin{cases} y_k^* = \frac{w_s}{w_k} \\ y_l^* = 0 \end{cases}.$$

Значение целевой функции: $f\left(\frac{w_s}{w_k}\right) + f(0)$;

– Граничное значение $I.2$:

$$\begin{cases} y_k^* = 0 \\ y_l^* = \frac{w_s}{w_l} \end{cases}.$$

Значение целевой функции: $f(0) + f\left(\frac{w_s}{w_l}\right)$.

На основании фактов что функция $f(x)$ является монотонно возрастающей и $w_k < w_l$ следует, что $f\left(\frac{w_s}{w_k}\right) > f\left(\frac{w_s}{w_l}\right)$, так как $\frac{w_s}{w_k} > \frac{w_s}{w_l}$. Следовательно, граничное значение $I.1$ максимизирует целевую функцию на интервале I .

Интервал II :

На интервале, когда $w_s \in (w_k, w_l]$ важно ограничить сверху значение y_k^* единицей, для этого величина y_l^* должна превышать $\frac{w_s - w_k}{w_l}$. В тоже время неотрицательность y_k^* обеспечивается ограниченностью сверху y_l^* величиной $\frac{w_s}{w_l}$. Так же из $w_s \leq w_l$ следует, что $y_l^* \leq 1$.

Основываясь на данных рассуждениях, y_l^* может принимать значения на отрезке $\left[\frac{w_s - w_k}{w_l}, \frac{w_s}{w_l}\right]$, и граничные значения могут быть представлены следующим образом:

– Граничное значение $II.1$:

$$\begin{cases} y_k^* = 1 \\ y_l^* = \frac{w_s - w_k}{w_l} \end{cases}.$$

Значение целевой функции: $f(1) + f\left(\frac{w_s - w_k}{w_l}\right)$;

– Граничное значение $II.2$:

$$\begin{cases} y_k^* = 0 \\ y_l^* = \frac{w_s}{w_l} \leq 1 \end{cases}.$$

Значение целевой функции: $f(0) + f\left(\frac{w_s}{w_l}\right) \leq f(0) + f(1)$.

Очевидно, что граничное значение *III.1* максимизирует целевую функцию на интервале *II*.

Интервал *III*:

На интервале $w_s \in (w_l, w_k + w_l)$ проводя рассуждения, аналогичные представленным ранее, y_l^* может принимать значения на отрезке $\left[\frac{w_s - w_k}{w_l}, 1\right]$ для выполнения всех ограничений оптимизационной задачи (3.8). Граничные значения могут быть представлены следующим образом:

– Граничное значение *III.1*:

$$\begin{cases} y_k^* = 1 \\ y_l^* = \frac{w_s - w_k}{w_l} \end{cases}.$$

Значение целевой функции: $f(1) + f\left(\frac{w_s - w_k}{w_l}\right)$;

– Граничное значение *III.2*:

$$\begin{cases} y_k^* = \frac{w_s - w_l}{w_k} \\ y_l^* = 1 \end{cases}.$$

Значение целевой функции: $f\left(\frac{w_s - w_l}{w_l}\right) + f(1)$.

Не составляет трудностей показать, что $\frac{w_s - w_k}{w_l} > \frac{w_s - w_l}{w_k}$, основываясь на следующих неравенствах: $w_s < w_k + w_l$ и $w_k < w_l$. Следовательно, граничное значение *III.1* максимизирует целевую функцию на интервале *III*.

Исходя из описанного выше, решение оптимизационной задачи (3.7) задается таблицей 3.1. Таблица 3.1 организована следующим образом: по заданному значению w_s определяются оптимальные значения y_k^* и y_l^* .

Обобщая изложенное, при всех возможных значениях w_s перераспределение долей взятых предметов в пользу предмета k за счет всей взятой доли предмета l приводит к максимизации целевой функции оптимизационной задачи (3.7) (таблица 3.1). Следствием данного факта является, что решение Y , возможно улучшить за счет перераспределения долей предметов под индексами k и l . Вид решений X , Y и Y^* представлен на рисунке 3.4.

Таким образом, любое решение Y , в котором найдутся предметы с индексами k и l может быть улучшено до решения Y^* , значение целевой функции которой больше, чем у Y . Большее значение целевой функции при решении Y^* , по

Решение промежуточной оптимизационной задачи

Значение w_s	Оптимальное значение y_k^*	Оптимальное значение y_l^*
$(0, w_k]$	$\frac{w_s}{w_k}$	0
$(w_k, w_k + w_l]$	1	$\frac{w_s - w_k}{w_l}$

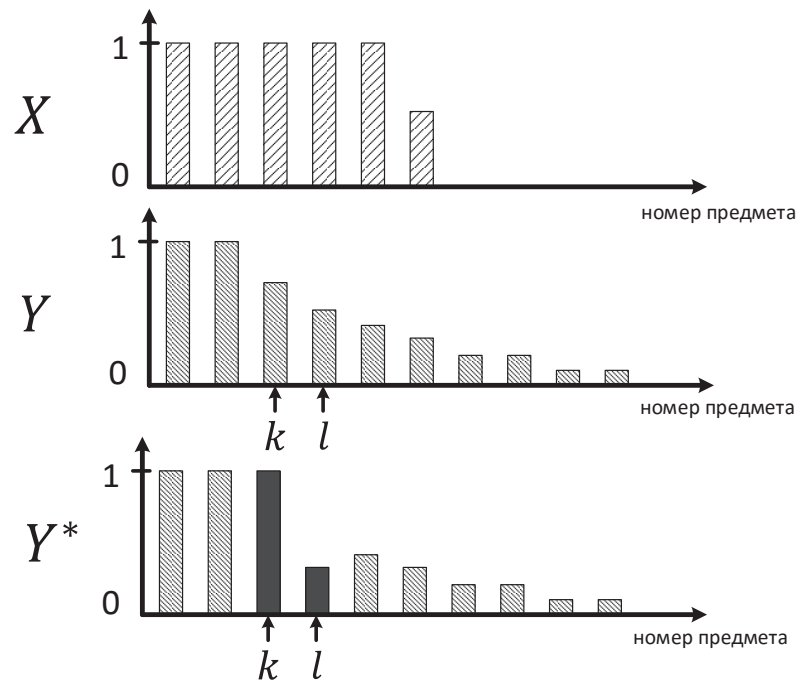
сравнению с Y , обеспечивается тем, что Y^* является решением промежуточной оптимизационной задачи (3.7).

Следовательно, всевозможные решения Y не являются оптимальными для обобщенной задачи о непрерывном рюкзаке (3.6), что является противоречием. Одновременно с этим важно отметить, что для решения X , полученного по средствам алгоритма 3.1, не найдется пары предметов с индексами k и l , следовательно оно не может быть улучшено и является оптимальным для оптимизационной задачи (3.6). \square

Проведем анализ сложности предложенного решения задачи о непрерывном рюкзаке. Для этого рассмотрим каждый шаг алгоритма 3.1:

- Шаг 1. Сортировка пользователей в порядке возрастания отношений $\frac{w_i}{c_i}$. Сложность данного шага составляет $O(N \log_2 N)$, $N \rightarrow \infty$ при использовании быстрой сортировки;
- Шаг 2. Поиск предмета с индексом j может быть осуществлен с линейной сложностью: $O(N)$, $N \rightarrow \infty$;
- Шаг 3. Вычисление оптимальных значений $x_i, i = \overline{1, N}$ может быть реализовано с линейной сложностью: $O(N)$, $N \rightarrow \infty$.

В терминах « O » большого сложность алгоритма в целом определяется самым сложным шагом. Следовательно, сложность алгоритма 3.1 определяется Шагом 1 и равняется $O(N \log_2 N)$, $N \rightarrow \infty$.

Рисунок 3.4 — Вид решений X , Y и Y^*

Для непрерывных задач выпуклого программирования от нескольких переменных с ограничениями типа равенств и неравенств существуют несколько способов отыскания экстремумов [32; 33]:

- Получение замкнутого выражения на основе условий Каруша-Куна-Таккера (ККТ);
- Получение численного решения на основе условий ККТ;
- Численные методы нелинейного программирования.

Стандартным методом решения задач выпуклого программирования является поиск решения удовлетворяющих условиям Каруша-Куна-Таккера [32; 33]. Данный метод сводится к нахождению решения системы нелинейных уравнений и неравенств. В результате решения данной системы может быть получено замкнутое выражение для условного экстремума. Тогда сложность полученного решения заключается в вычислении значений по заданным входным данным. Сложность вычисления решения растет линейно от числа переменных N : $O(N)$, $N \rightarrow \infty$.

Однако, ввиду большого числа переменных в системе, полученной на основе условий ККТ, получение замкнутого решения может быть затруднительно. Для решения подобных систем может быть применен метод, известный в англоязычной литературе как «*water-filling*», подробно описанный в [32]. Данный метод позволяет найти решение системы, которое может быть получено с исполь-

зованием численной процедуры. В настоящей работе, метод «*water-filling*» был применен для решения оптимизационной задачи, представленной в подразделе 4.4. Сложность вычисления решения определяется сложностью процедуры получения численного решения.

Для решения задач выпуклого программирования от нескольких переменных с ограничениями типа равенств и неравенств известны численные методы решения: методы спуска и штрафных функций [33]. Важно отметить, что для задач выпуклого программирования данные методы сходятся к оптимальному решению, так как в подобных задачах существует только один экстремум. Оценить сложность получения решения оптимизационных задач с использованием подобных методов не представляется возможным, так как данные методы являются итеративными и зависят от начальных значений поиска экстремума (невозможно предсказать количество шагов, необходимых для нахождения оптимальных значений оптимизационной задачи). Сравнение сложности методов решения оптимизационных задач выпуклого программирования от нескольких переменных с ограничениями типа равенств и неравенств представлены в таблице 3.2.

Таким образом, при сравнении сложности полученного решения на основе алгоритма 3.1 со стандартными методами решения непрерывных задач выпуклого программирования от нескольких переменных с ограничениями типа равенств и неравенств справедливо следующее утверждение. Сложность решения на основе алгоритма 3.1 является приемлемой в сравнении со стандартными методами решения подобных задач.

В дополнении сказанного выше, решение оптимизационной задачи (3.6) алгоритмом 3.1 может быть вычислено в режиме реального времени. Под режимом реального времени понимается возможность вычисления оптимальных значений в короткий промежуток времени. В настоящей работе за единицу времени принят интервал планирования равный 1 мс (подраздел 2.4). Элементарно показывается, что на современных процессорах алгоритм 3.1 может быть выполнен за промежуток времени не превышающий интервал планирования.

Таблица 3.2

Сравнение сложности методов решения обобщенной задачи о непрерывном рюкзаке

Название метода	Характеристика сложности метода при наличии N переменных	Возможность реализации в реальном масштабе времени
Замкнутое выражение (ККТ)	$O(N), N \rightarrow \infty$	+
Решение на основе алгоритма 3.1	$O(N \log_2 N), N \rightarrow \infty$	+
Численное решение системы условий ККТ	Зависит от процедуры получения численного решения	+/-
Численные методы нелинейного программирования	Высокая сложность	—

3.5 Нижняя граница нормированного отношения длительности буферизации к просмотру при передаче неадаптивных видеопотоков

В подразделе 3.3 была поставлена оптимизационная задача нелинейного программирования (3.4), на основе решения которой возможно найти нижнюю границу для средне-пользовательского значения нормированного отношения длительности ожидания к просмотру: G (определение 3.2) по всевозможным алгоритмам планирования $\mathcal{A}(\cdot)$, удовлетворяющих введенным ограничениям в подразделе 2.7. В подразделе 3.4 была представлена обобщенная задача о непрерывном рюкзаке, когда в целевой функции фигурирует сумма выпуклых

функций, и предложен алгоритм нахождения решения данной задачи, с низкой вычислительной сложностью. Настоящий подраздел ставит своей задачей основываясь на результатах, полученных в подразделе 3.4, найти нижнюю границу для критерия качества G .

Оптимизационная задача (3.4) имеет следующий вид:

$$\text{Минимизировать: } G = \sum_{i=1}^N g_i$$

При условии:

$$\begin{cases} \sum_{i=1}^N \frac{K_i(1-g_i)}{g_i+\gamma(1-g_i)} \leq 1 \\ g_i \geq 0, i = \overline{1, N} \\ g_i \leq 1, i = \overline{1, N} \end{cases},$$

где $K_i = R_i E[C_i^{-1}]$, γ – коэффициент разряженности видеопотока (определение 2.1).

Теорема 3.1. *Решение задачи нелинейного программирования (3.4), характеризующее нижнюю границу нормированного отношения длительности буферизации к просмотру при передаче неадаптивных видеопотоков, может быть получено по средствам алгоритма 3.2.*

Алгоритм 3.2 : Решение задачи (3.4)

- 1: Сортировка пользователей по возрастанию отношений $\frac{K_i}{\gamma}$, и их нумерация, в соответствие с сортировкой;

- 2: Нахождение пользователя с индексом j :
$$\begin{cases} \sum_{i=1}^{j-1} \frac{K_i}{\gamma} \leq 1 \\ \sum_{i=1}^j \frac{K_i}{\gamma} > 1 \end{cases}$$

если $j \in \emptyset, j = N + 1$;

- 3: Нахождение оптимальных значений g_i :

$$g_i = \begin{cases} 0, & i < j \\ \frac{\gamma(1-\xi)}{\xi+\gamma(1-\xi)}, & i = j, \\ 1, & i > j \end{cases}$$

$$\text{где } \xi = \frac{\gamma}{K_j} \left(1 - \sum_{i=1}^{j-1} \frac{K_i}{\gamma} \right).$$

Доказательство. Для доказательства теоремы 3.1 произведем несколько преобразований системы (3.4).

Изначально вынесем из знаменателя нелинейного ограничения γ :

$$\sum_{i=1}^N \frac{K_i(1 - g_i)}{g_i + \gamma(1 - g_i)} = \sum_{i=1}^N \left(\frac{K_i}{\gamma} \right) \frac{(1 - g_i)}{\frac{g_i}{\gamma} + (1 - g_i)}.$$

Далее произведем преобразование задачи (3.4) из задачи минимизации в задачу максимизации:

$$\textbf{Максимизировать: } \sum_{i=1}^N (1 - g_i)$$

При условии:

$$\begin{cases} \sum_{i=1}^N \left(\frac{K_i}{\gamma} \right) \frac{(1 - g_i)}{\frac{g_i}{\gamma} + (1 - g_i)} \leq 1 \\ g_i \geq 0, i = \overline{1, N} \\ g_i \leq 1, i = \overline{1, N} \end{cases}.$$

Введем обозначение:

$$x_i = \frac{(1 - g_i)}{\frac{g_i}{\gamma} + (1 - g_i)}.$$

Важно отметить, что значение x_i монотонно зависит от величины g_i и может принимать значения в отрезке $[0, 1]$. Выразим значения g_i и $(1 - g_i)$ из x_i :

$$g_i = \frac{\gamma(1 - x_i)}{x_i + \gamma(1 - x_i)}. \quad (3.9)$$

$$1 - g_i = \frac{x_i}{x_i + \gamma(1 - x_i)}.$$

Основываясь на введенных обозначениях, оптимизационная задача (3.4) принимает следующий вид:

$$\begin{aligned} &\text{Максимизировать: } \sum_{i=1}^N \frac{x_i}{x_i + \gamma(1-x_i)} \\ &\text{При условии:} \\ &\begin{cases} \sum_{i=1}^N \left(\frac{K_i}{\gamma} \right) x_i \leq 1 \\ x_i \geq 0, i = \overline{1, N} \\ x_i \leq 1, i = \overline{1, N} \end{cases} \end{aligned} \quad (3.10)$$

Проведем анализ целевой функции оптимизационной задачи (3.10) и покажем, что данная целевая функция соответствует требованиям, налагаемым на целевую функцию оптимизационной задачи (3.6). Из вида функции очевидно, что она непрерывна, строго монотонно возрастает и ограничена сверху и снизу на области определения x_i . Проведем анализ второй производной:

$$\left[\frac{x_i}{x_i + \gamma(1-x_i)} \right]'' = \frac{2\gamma(\gamma-1)}{(x_i + \gamma(1-x_i))^3}.$$

По определению, значение $\gamma \geq 1$, так же величина x_i , по условию оптимизационной задачи, может принимать значения в отрезке $[0,1]$, следовательно вторая производная неотрицательна в области определения x_i .

Основываясь на представленном анализе, $\frac{x_i}{x_i + \gamma(1-x_i)}$ – непрерывная, выпуклая вниз функция, дважды дифференцируемая, строго монотонно возрастающая и ограниченная на области определения x_i функция. Исходя и описанного выше и результатов, представленных в подразделе 3.4, справедливо следующее утверждение: решение задачи (3.10) может быть получено при использовании алгоритма 3.1, где $w_i = \left(\frac{K_i}{\gamma} \right)$, $c_i = 1, i = \overline{1, N}$ и $W = 1$.

Алгоритм 3.2 является адаптацией алгоритма 3.1 в обозначениях оптимизационной задачи (3.4) и взаимосвязи между значениями g_i и x_i (3.9). Данный факт завершает доказательство. \square

Результатом решения задачи (3.4) является вектор отношений длительности буферизации к просмотру для каждого пользователя: (g_1, g_2, \dots, g_N) , обладающий минимально возможной суммой элементов, удовлетворяющих всем ограничениям оптимизационной задачи. Среднее арифметическое элементов данного

вектора является нижней границей по всем возможным алгоритмам планирования распределения ресурсам, удовлетворяющим допущениям, введенным в подразделе 2.7. Обобщая сказанное ранее, в настоящем разделе была найдена граница, характеризующая максимально возможную производительность алгоритмов распределения ресурсов радиоканала в централизованных беспроводных телекоммуникационных сетях при передачи неадаптивных видеопотоков.

Настоящий результат имеет широкий спектр применения:

- **Теоретическая область.** Нижняя граница может быть использована для аналитической оценки производительности централизованных беспроводных для существующих и последующих стандартов беспроводной связи.
- **Практическая область.** Благодаря невысокой вычислительной сложности вычисления нижней границы (алгоритм 3.2), настоящая нижняя граница может быть непосредственно встроена в алгоритм распределения ресурсов беспроводного канала на базовой станции. Такая возможность позволяет создать семейство эвристических алгоритмов планирования для повышения производительности сети передачи информации в целом.

Завершающая часть настоящего раздела посвящена демонстрации области практической применимости полученного результата. Далее в разделе будет предложен эвристический алгоритм планирования ресурсов беспроводного канала на базовой станции и произведено его сравнение с известными планировщиками и нижней границей.

3.6 Алгоритм планирования распределения ресурсов для нормированного отношения длительности буферизации к просмотру при передаче неадаптивных видеопотоков

Настоящий подраздел описывает концепцию алгоритма планирования распределения ресурсов на базовой станции, в основу которого положен результат, представленный в подразделе 3.5. В настоящем подразделе будет предложен планировщик, опирающийся на концепцию совместного планирования ресурсов радиоканала между всеми активными пользователями, в отличии от классических

алгоритмов планирования (подраздел 2.5). Под совместным планированием подразумевается учет при распределении частотно-временных ресурсов конкретному пользователю не только его статистики, но и статистик всех активных пользователей, присутствующих в системе передачи информации в момент планирования.

Настоящий подраздел последовательно описывает этапы построения алгоритма планирования:

- Определение объема доступной информации для принятия решения;
- Статистики и методы их вычисления;
- Методы управления распределением частотно-временных ресурсов;
- Алгоритм распределения ресурсов беспроводного канала.

В начале будет представлено общее описание рассматриваемой системы. На рисунке 3.5 представлена концептуальная схема беспроводной централизованной сети при передаче видеоданных, основанная на описании представленному в подразделах 2.3, 2.4 и 2.5. Важным функциональным компонентом опорной сети оператора в современных стандартах связи является система анализа трафика, которая может определить тип используемого трафика и его характеристики (в англоязычной литературе данные системы известны под названием «*Deep Packet Inspection*» или DPI). Центральным элементом рассматриваемой схемы является планировщик базовой станции. Важным моментом в настоящем описании является объемы и моменты времени доступа к информации на основе которой алгоритм планировщика будет принимать решения распределении ресурсов беспроводного канала.

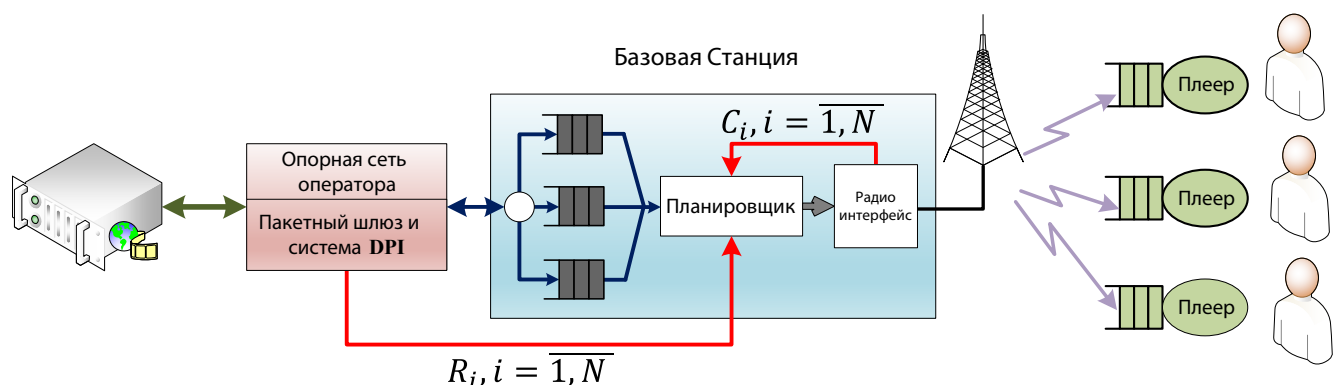


Рисунок 3.5 — Концептуальная схема предлагаемого алгоритма планирования

На первом шаге описания алгоритма планирования будет представлена информация, доступная алгоритму планирования, для принятия решения. В насто-

ящей работе, для планировщика доступна следующая информация для каждого пользователя i в каждый момент времени t :

- Битовая скорость потока $R_i(t)$, полученная от системы анализа трафика DPI;
- Значение максимально достижимой скорости канала $C_i(t)$, значение которой доступно от радио интерфейса;
- Объем данных доступный в очереди для передачи на пользовательское устройство $B_i(t)$;
- Объем переданных данных $P_i(t)$.

Важно отметить, что в существующих беспроводных централизованных системах связи время является дискретным и разбито на интервалы равной длительности Δt секунд.

На втором шаге описания алгоритма предлагаются методы агрегирования доступной информации. Представленный ранее объем информации должен быть агрегирован в некоторые значения, называемые статистиками, которые отражают текущую ситуацию в зоне ответственности базовой станции. Наиболее критичные статистики, необходимые для функционирования планировщика, представлены следующим списком:

- Активность пользователя $A_i(t)$;
- Оценка скорости передачи информации $S_i^{avg}(t)$;
- Оценка максимально достижимой скорости канала $C_i^{avg}(t)$.

Сложность вычисления данных статистик является важным ограничением, налагаемым на вычислительную процедуру, ввиду жесткого временного ограничения на длительность получения результата (подраздел 2.5).

В настоящей работе рассматривается модель поведения пользователя, просматривающего последовательность роликов (подраздел 2.3). Следовательно, пользователь может находиться в двух состояниях: активном (загружает видеоданные) или неактивном (выбирает следующий ролик для просмотра). Таким образом, на уровне планировщика необходимо произвести оценку $A_i(t)$: в каком из двух состояний находится пользователь i в момент времени t . Активность пользователя определяется логическим значением (истинно или ложно). Будем считать пользователя активным, если разность между текущим моментом времени и последним моментом времени, когда у абонента i были данные для передачи не

превышает t^{act} секунд:

$$A_i(t) = \mathbb{1}\{t - t_i^l \leq t^{act}\}, \quad (3.11)$$

где $t_i^l = \max_t \mathbb{1}\{B_i(t) > 0\}$.

Так как в настоящей работе рассматривается передача информации по беспроводному каналу связи, важной отличительной чертой которого является изменчивость характеристик во времени (подраздел 2.4). В качестве статистики для оценок $S_i^{avg}(t)$ и $C_i^{avg}(t)$ в настоящей работе предлагается использовать среднее значение за продолжительный промежуток времени (длительность усреднения) \bar{w}_S и \bar{w}_C соответственно. Ограничение интервала усреднения обусловлены изменчивостью канала связи и необходимостью поддержания актуальности статистик.

Простейшим подходом для вычисления средних значений является среднее арифметическое значение, однако, сложность его вычисления по операциям и затрачиваемым объемом памяти достаточно высока: хранение всего массива исходных данных и его обработка. Ввиду существующих ограничений на сложность в качестве метода оценки предлагается использовать метод экспоненциального забывания (в литературе встречается термин «экспоненциальное сглаживание»):

$$X_i^{avg}(t) = \left(1 - \frac{1}{\bar{w}_X}\right) X_i^{avg}(t-1) + \left(\frac{1}{\bar{w}_X}\right) X_i(t),$$

где $X_i(t)$ – элемент дискретного случайного процесса, \bar{w}_X – длительность усреднения для процесса X и $X_i^{avg}(0) = X_i(1)$.

Важной характеристикой методов оценки среднего значения за некоторый промежуток времени является скорость сходимости полученной оценки к истинному значению. На скорость сходимости среднего значения, вычисленного методом экспоненциального забывания, к истинному значению большое влияние оказывает длительности интервалов усреднения. Во многих случаях оценка среднего значения методом среднего арифметического и экспоненциального забывания показывает близкие значения. Однако, существует сценарий, на котором оценки, полученные по настоящим методикам, показывают разную скорость сходимости к истинному среднему значению. Подобная ситуация возможна, когда первое значение случайного процесса имеет сильное отличие от истинного среднего. Продемонстрируем это на примере: пусть существует дискретный случайный процесс

$X(t) \sim N(10, 1.5)$, при этом значение $X(1) = 0$, и необходимо произвести оценку его среднего значения за последние $\bar{w}_X = 10$ отсчетов (рисунок 3.6).

Из рисунка 3.6 следует, что метод экспоненциального ожидания при оценке среднего значения показывает медленную скорость сходимости к истинному среднему значению по сравнению с методом среднего арифметического. Поэтому в настоящем подразделе предлагается модифицировать метод экспоненциального забывания изменяемым во времени интервалом усреднения по следующему правилу:

$$w_X(t) = \min(t, \bar{w}_X).$$

В настоящей работе, метод оценки среднего значения, использующий изменяемый во времени интервал усреднения называется «методом экспоненциального забывания с быстрой адаптацией». Производительность данного метода оценки сравнима с методом среднего арифметического и демонстрирует достаточную скорость сходимости к истинному значению (рисунок 3.6).

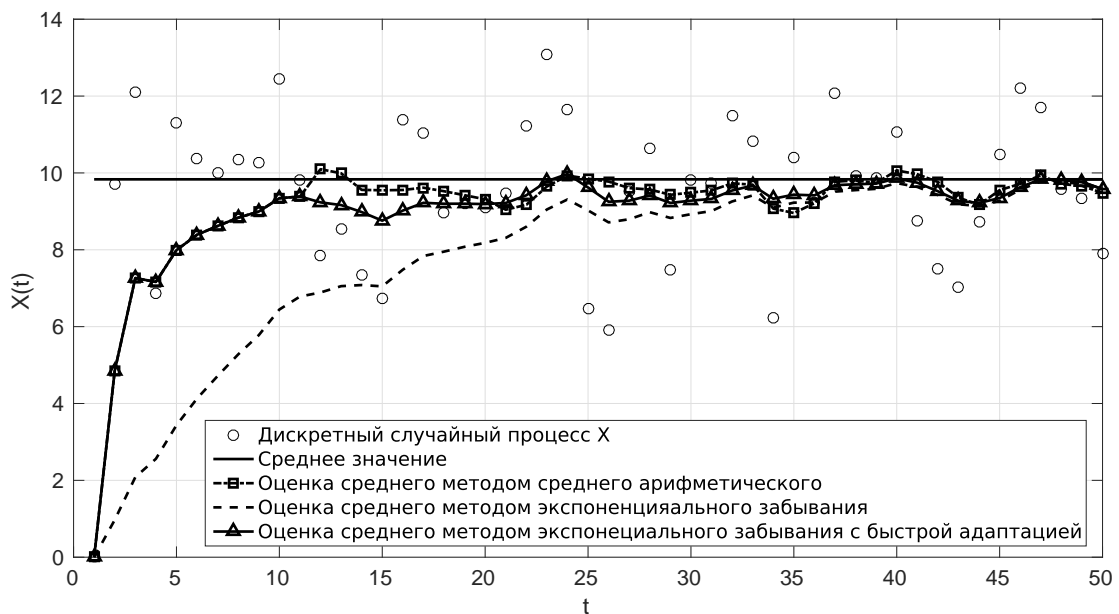


Рисунок 3.6 — Сравнение скорости сходимости методов оценки среднего значения

Обобщая сказанное выше, для оценки средних значений $C_i^{avg}(t)$ и $S_i^{avg}(t)$ предлагаются следующие методики, вычисление которых производится

только в активном состоянии пользователя:

$$C_i^{avg}(t) = \left(1 - \frac{1}{w_C(t)}\right) C_i^{avg}(t-1) + \left(\frac{1}{w_C(t)}\right) C_i(t) \quad (3.12)$$

$$S_i^{avg}(t) = \left(1 - \frac{1}{w_S(t)}\right) S_i^{avg}(t-1) + \left(\frac{1}{w_S(t)}\right) \frac{P_i(t)}{\Delta t}, \quad (3.13)$$

где $w_C(t) = \min(t - t_i^l + 1, \bar{w}_C)$ и $w_S(t) = \min(t - t_i^l + 1, \bar{w}_S)$ соответственно и $C_i^{avg}(0) = S_i^{avg}(0) = 0$.

Неотрицательность значений $w_C(t)$ и $w_S(t)$ обеспечиваются вычислением оценок только в активном состоянии пользователя ($t - t_i^l \geq 0$). Подобный метод оценки поддерживает актуальность статистик, так как канал может динамично меняться во времени, а так же обладает низкой вычислительной сложностью.

Следующим шагом в представлении алгоритма планирования является метод управления распределением частотно-временных ресурсов, имеющего непосредственное влияние на скорость передачи информации. Для представления данного метода необходимо представить описание каким образом существующие алгоритмы планирования распределяют ресурсы беспроводного канала связи. Как было описано в подразделе 2.4, беспроводной канал связи разделен на ресурсные блоки, являющиеся минимальной единицей ресурсов, которая может быть выделена пользователю. Количество ресурсных (N_{rb}) блоков в одном интервале планирования определяется шириной полосы.

В каждый момент времени t для конкретного ресурсного блока r вычисляется приоритет выделения для каждого пользователя i : $p_{r,i}(t)$, по некоторому правилу. Далее, для каждого конкретного ресурсного блока находится пользователь с максимальным приоритетом и выделяется найденному пользователю. Важно отметить, что все алгоритмы планирования не распределяют ресурсы пользователям, у которых отсутствуют данные для передачи на Уровне очередей. Именно правило вычисления приоритета пользователя на ресурсный блок определяет алгоритм планирования. Наиболее известными эвристическими алгоритмами планирования являются *Proportional Fair* и *Round Robin*:

- *Proportional Fair* ставит своей целевой функцией обеспечение равных скоростей передачи информации:

$$p_{r,i}^{PF}(t) = \frac{C_i^{avg}(t)}{S_i^{avg}(t)};$$

- *Round Robin* обеспечивает равный доступ к ресурсам беспроводного канала за счет циклического распределения ресурсных блоков между активными пользователями.

Важной отличительной чертой классических планировщиков является независимость вычисления приоритета пользователя для конкретного блока ресурсов от других пользователей. В настоящей работе предлагается иная концепция планирования: совместное планирование распределения ресурсов, в котором важную роль выполняет зависимость вычисления приоритета от характеристик всего множества пользователей: их числа, вида трафика, требований к обслуживанию и т.д.

Аналитические исследования производительности систем передачи информации, аналогичные представленным в подразделе 3.5, позволяют оценить скорость обслуживания для каждого конкретного абонента в зависимости от ситуации в соте. Таким образом, в каждый момент времени может быть вычислена «рекомендуемая» скорость передачи информации для максимизации производительности сети в соответствии с некоторым критерием. В свою очередь, рекомендуемая скорость обеспечивает максимальную производительность системы, при условии ее обеспечения. Обобщая сказанное выше, для реализации концепции совместного планирования распределения ресурсов необходимо решить две задачи:

- Предложить алгоритм расчета рекомендованной скорости передачи данных, на основе ограниченного объема информации доступного планировщику базовой станции;
- Предложить механизм обеспечения рекомендованной скорости передачи данных.

Изначально будет предложен механизм обеспечения рекомендуемой скорости передачи информации $S_i^{rec}(t)$ пользователю i . Основной функцией алгоритма планирования является выделение абонентам частотно-временных ресурсов: в каждый момент времени планировщик принимает решение о доле выделенных ресурсов для всех пользователей (подраздел 2.5). На основе данных возможностей, управление скоростью передачи информации на уровне доступа ко среде может быть осуществлено за счет контроля частоты выделения ресурсов. Пусть некоторому абоненту i необходимо обеспечить скорость получения информации $S_i^{rec}(t)$. Алгоритм планирования может оценить скорость передачи информации

в каждый момент времени t на основе выражения (3.13). Для того чтобы скорость получения информации не превышала рекомендованной скорости возможно применить следующий подход: пользователь i участвует в распределении ресурсов в момент t , если $\mathbb{1}\{S_i^{avg}(t) \leq S_i^{rec}(t)\}$.

Выберем в качестве основы алгоритм *Proportional Fair* и произведем его объединение с механизмом ограничения скорости передачи информации:

$$p_{r,i}(t) = \mathbb{1}\{S_i^{avg}(t) \leq S_i^{rec}(t)\} \frac{C_i^{avg}(t)}{S_i^{avg}(t)}. \quad (3.14)$$

Расчет приоритета пользователя на основе выражения (3.14) позволяет обеспечить скорость передачи информации не превышающую рекомендованной. Важно отметить, что если в некоторый момент времени у всех активных пользователей оценка скорости превышает рекомендованные, то у них всех будут равные приоритеты на ресурсные блоки, и они могут быть выделены случайным абонентам для полного использования ресурсов беспроводного канала.

После выбора механизма управления скоростью передачи информации необходимо детерминировать алгоритм вычисления рекомендованных скоростей передачи информации, чтобы на их основе организовать управление передачей данных. Предполагается, что на алгоритме планирования для всех пользователей, просматривающих видео, известна оценка битовой скорости видеопотока $(\hat{R}_i(t))$, просматриваемого в текущий момент времени, от системы анализа трафика DPI и оценка максимально достижимой скорости канала $(C_i^{avg}(t))$, на основе выражения (3.12). Данный объем информации позволяет произвести вычисление оценки нижней границы нормированного отношения длительности буферизации к просмотру при передаче неадаптивных видеопотоков $\hat{G}(t) = (\hat{g}_1(t), \hat{g}_2(t), \dots, \hat{g}_N(t))$ для всего множества активных пользователей в момент времени t ($i : A_i(t) > 0$), в предположении что $\forall i : \gamma_i = 1$, так как планирование происходит на непродолжительный интервал времени.

На основе оценки $\hat{G}(t)$ рекомендованные скорости предлагается вычислять следующим образом:

$$S_i^{rec}(t) = \begin{cases} \hat{R}_i(t), & \hat{g}_i(t) = 0 \\ \hat{R}_i(t)(1 - \hat{g}_i(t)), & \hat{g}_i(t) \in (0,1) , \\ S^{min}, & \hat{g}_i(t) = 1 \end{cases} \quad (3.15)$$

где S^{min} – минимальная скорость передачи информации, обеспечиваемая алгоритмом планирования. Значение минимальной скорости обслуживания абонентов определяется стандартами связи.

Используя описанные ранее результаты, предлагается эвристический алгоритм планирования, выполняемый в каждый интервал планирования, состоящий из четырех шагов (рисунок 3.7):

- **Определение множества активных пользователей.** На данном шаге все пользователи разделяются на два подмножества: активные и неактивные, в соответствии с выражением (3.11).
- **Вычисление рекомендуемых скоростей передачи данных.** На основе статистик $C_i^{avg}(t), S_i^{avg}(t), i : A_i(t) > 0$ и производится расчет оценки нижней границы для критерия G и рекомендуемых скоростей передачи данных для всех активных пользователей (3.15).
- **Распределение ресурсов беспроводного канала.** Выделение ресурсов беспроводного канала в соответствии с предложенным механизмом обеспечения рекомендованной скорости (3.14). На данном шаге также производится контроль за объемом выделенных ресурсов с целью предотвращения выделения ресурсных блоков, которые не могут быть использованы пользователем ввиду отсутствия достаточного объема информации в очереди на базовой станции.
- **Обновление статистик.** Для множества активных абонентов обновляются значения статистик в соответствии с выражениями (3.12) и (3.13).

В результате настоящего подраздела предложен алгоритм планирования распределения частотно-временных ресурсов беспроводного канала связи для централизованных систем передачи информации. Производительность предложенного алгоритма планирования будет продемонстрирована в подразделе 3.7 в сравнении с известными решениями и найденной в подразделе 3.5 нижней границей.

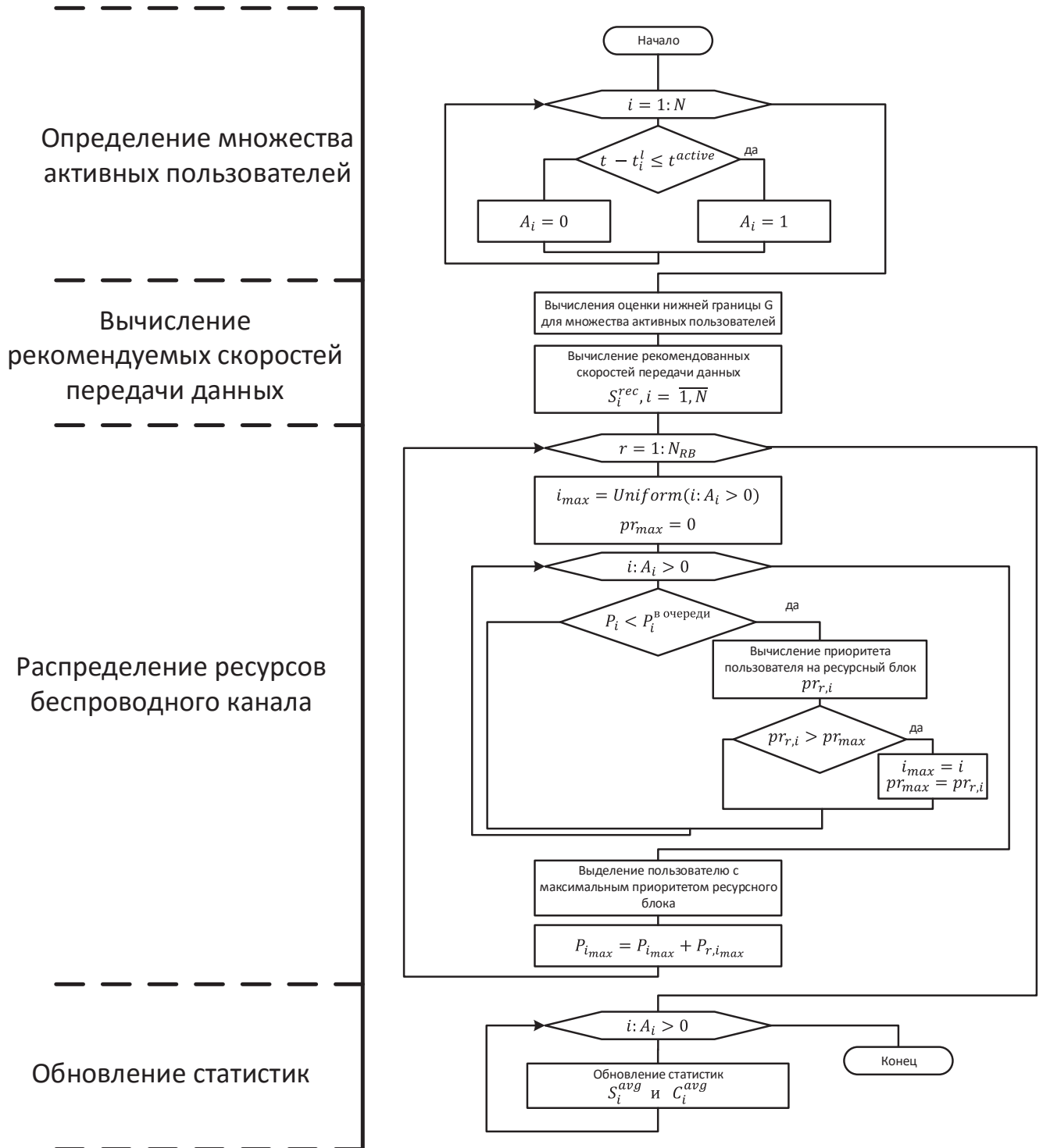


Рисунок 3.7 — Логика алгоритма совместного планирования распределения ресурсов для неадаптивных видеопотоков

3.7 Численный пример

Демонстрация результатов настоящей работы является нетривиальной задачей ввиду сложности анализируемой системы и невозможности постановки натурных экспериментов, так как оборудование централизованных сетей связи (базовые станции, опорная сеть оператора и т.д.) недоступно для рядового исследователя, что приводит к необходимости поиска платформы для моделирования. Таким образом, настоящий подраздел решает две задачи:

- Выбор стандарта беспроводной централизованной системы передачи информации, на котором будут демонстрироваться полученные результаты;
- Выбор платформы и сценария моделирования, обладающих достаточной достоверностью получаемых результатов моделирования.

Изначально проведем анализ современных стандартов связи, которые соответствуют рассматриваемой в настоящей работе модели системы (раздел 2). В настоящий момент времени наиболее известными технологиями, соответствующими введенной модели системы, являются два стандарта связи: IEEE 802.16 или *Worldwide Interoperability for Microwave Access* (WiMAX) [34] и *Long-Term Evolution Advanced* (LTE) [35]. В соответствии с анализом распространенности технологий связи, проведенным компанией Cisco, в настоящее время сети связи, построенные на основе стандарта WiMAX, не имеют широкого распространения, в свою очередь системы стандарта LTE получили широчайшее распространение в мире [1]. Более того, стандарт связи LTE постоянно обновляется и дополняется. Основываясь на сказанном выше, результаты настоящей работы будут демонстрироваться на основе стандарта связи LTE. Важно отметить, что полученные в настоящей работе результаты применимы ко всем беспроводным централизованным системам связи, а в данном разделе лишь демонстрируются на примере самого распространенного стандарта, удовлетворяющего модели системы.

Следующим шагом необходимо выбрать платформу моделирования систем стандарта LTE. В настоящий момент времени существует несколько возможных платформ:

- *Matlab*. Компанией MathWorks предлагается модуль анализа сетей стандарта LTE: LTE System Toolbox в рамках пакета Simulink, распространяемый с платной подпиской [36]. Данный модуль позволяет оценивать

характеристики беспроводного канала связи для данного стандарта путем имитационного моделирования. Однако, в настоящем пакете не представлены реализации вышележащих уровней стека протоколов беспроводного соединения стандарта LTE и опорной сети оператора. Данный факт приводит к невозможности использования данной платформы для демонстрации результатов настоящей работы.

- *Opnet LTE Simulation*. Компания Opnet предлагает платформу для моделирования сетей стандарта LTE с платной подпиской OPNET LTE SIMULATION [37]. Данная платформа включает в себя полный стек протоколов беспроводной связи и опорной сети оператора, однако, исходный код данного проекта является закрытым и его невозможно дополнить клиент-серверными приложениями, осуществляющие загрузку и демонстрацию видео. Следовательно, использовать данную платформу невозможно для демонстрации полученных результатов.
- Система моделирования NS-3. В 2012 году была создана низкоуровневая система моделирования сетевых протоколов с открытым исходным кодом NS-3 [38]. Данная система моделирования реализована на языке программирования C++ и ее компоненты соответствует стандартам связи. В NS-3 также реализован модуль стека протоколов LTE для беспроводной сети и опорной сети оператора, соответствующие стандарту [35]. Открытый исходный код позволяет производить дополнение и гибкую настройку системы для широкого спектра задач.

Таким образом, результаты настоящей работы будут демонстрироваться с использованием системы моделирования NS-3 для стандарта связи LTE. Достоверность полученных результатов обеспечивается соответствием системы моделирования стандарту связи LTE.

Для адаптации системы моделирования NS-3 был решен ряд технических задач:

- Реализация приложений видео контент сервера и клиентского приложения, соответствующие стандарту DASH (подраздел 1.4);
- Реализация в модуле базовой станции LTE системы анализа трафика DPI, позволяющая в каждый момент времени моделирования получить информацию о виде трафика и его характеристиках;

- Реализация механизмов оценки характеристик передачи информации для пользователей, подключенных к базовой станции, в соответствии с решениями, представленными в подразделе 3.6;
- Реализация алгоритма планирования, предложенного в подразделе 3.6.

Следующим важным шагом для демонстрации результатов является выбор сценария моделирования. В настоящей работе автор следовал указаниями, представленным в [39]. Рассматривается работа одной соты, в которой находится N пользователей на прямой линии с равным расстоянием друг от друга (рисунок 3.8). Расстояние между соседними пользователями вычисляется следующим образом:

$$\Delta x = \frac{X_{max}}{N + 1},$$

где X_{max} – максимально возможная удаленность пользователей от базовой станции.

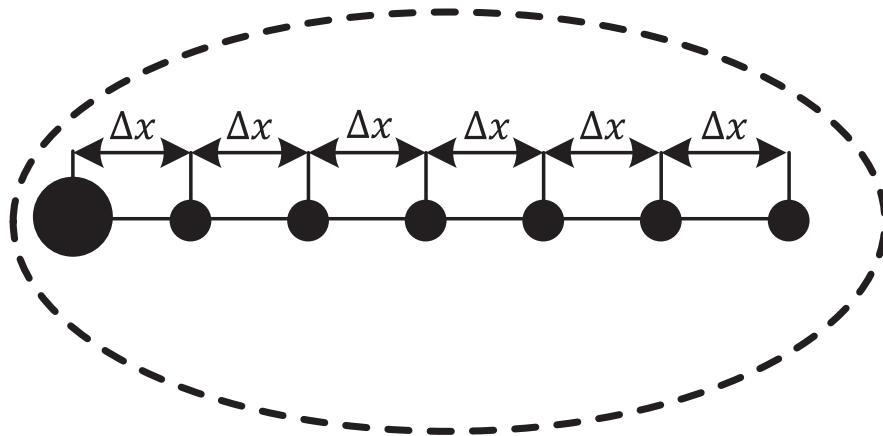


Рисунок 3.8 — Сценарий моделирования

В данном сценарии моделирования рассматривается канал связи с плоскими замираниями (подраздел 2.4). Для этого был сгенерирован случайный процесс, который изменяет затухание в канале от времени по всем доступным частотам для передачи. Таким образом затухание сигнала в канале от времени зависит от двух факторов: удаленности от базовой станции и плоскими замираниями в канале связи.

Полный список настроек сценария моделирования для каждого компонента модели системы представлен в таблице Б.1.

В сценарии моделирования наращивалось число абонентов до перехода системы в состоянии перегрузки (определение 3.1). И производится сравнение на

основе критерия G (3.2) производительности известных алгоритмов планирования (*Proportional Fair* и *Round Robin*) с предложенным алгоритмом и найденной нижней границей. Результат проведенного сравнения представлен на рисунке 3.9.

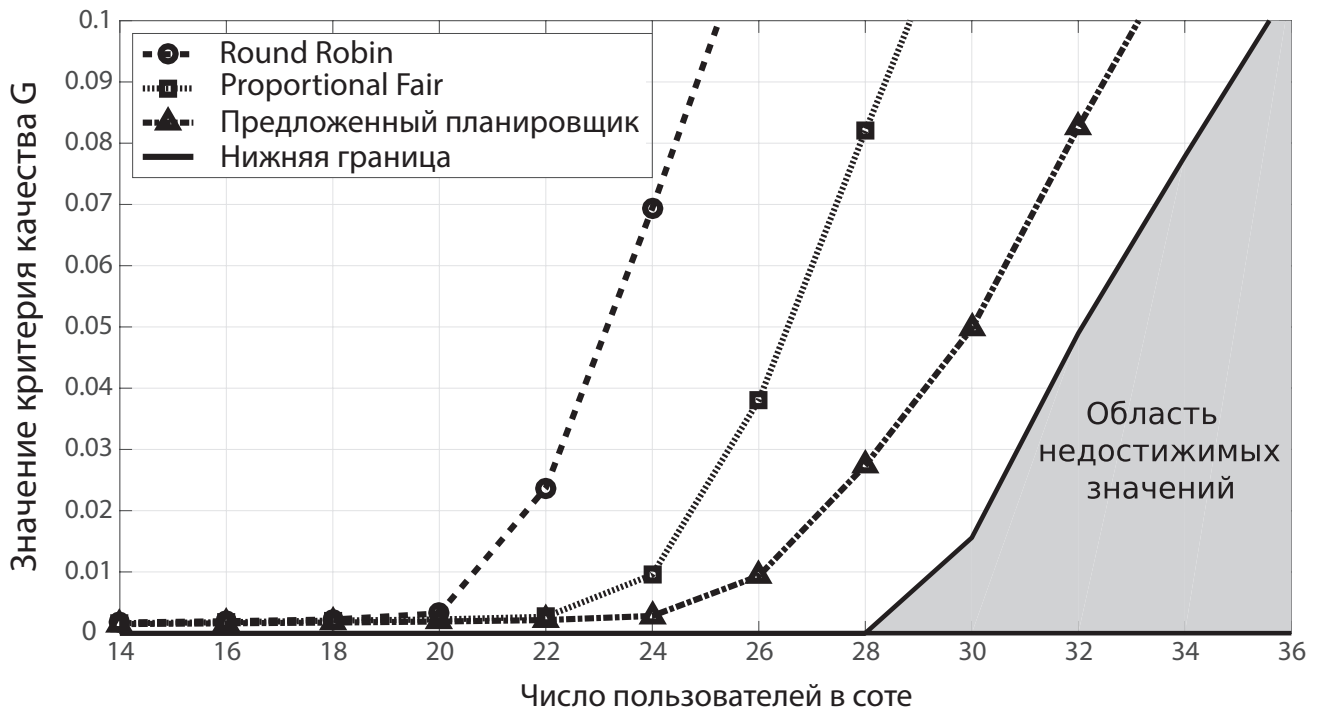


Рисунок 3.9 — Сравнение предложенного алгоритма планирования с известными планировщиками и найденной нижней границей

На рисунке 3.9 приведена нижняя граница, характеризующая максимальную производительность алгоритмов планирования в настоящем сценарии моделирования. Заштрихованная область характеризует значения, которые не могут быть достигнуты любым планировщиком, удовлетворяющим допущениям введенных в подразделе 2.7: не существует планировщика, который при 30-ти пользователей в соте обеспечивает значения критерия качества восприятия G равный 0.01. Из рассмотрения рисунка 3.9 следует, что предложенный алгоритм планирования показывает высокую производительность в сравнении со стандартными алгоритмами и нижней границей.

3.8 Выводы по разделу

Настоящий раздел был посвящен вопросам оптимизации передачи неадаптивных видеопоследовательностей в беспроводных централизованных сетях. В рамках раздела был выбран критерий качества восприятия: нормированное отношение длительности буферизации к просмотру (определение 3.2), характерный для неадаптивного видео, и произведен его анализ. На основании данного критерия и модели системы, введенной в разделе 2 была поставлена оптимизационная задача нелинейного программирования (3.4), решение которой характеризует максимально возможную производительность алгоритмов планирования при передаче неадаптивного видео. Далее было предложено решение данной задачи, путем ее сведения к обобщенной задаче о непрерывном рюкзаке, решение которой было найдено в настоящем разделе. Полученное решение характеризует нижнюю границу нормированного отношения длительности буферизации к просмотру для всевозможных алгоритмов планирования, удовлетворяющих введенным допущениям в подразделе 2.7, при передаче неадаптивных видеопотоков.

На основе найденного решения, обладающего низкой вычислительной сложностью, был предложен эвристический планировщик, с высокой производительностью в сравнении с существующими планировщиками и найденной нижней границей.

Основные результаты раздела могут быть сформулированы следующим образом:

- Проведен анализ факторов, влияющих на восприятие неадаптивных видеопотоков, и предложен критерий качества восприятия на основе нормированного отношения длительности буферизации к просмотру (подраздел 3.2);
- Произведена постановка оптимизационной задачи нелинейного программирования, характеризующей нижнюю границу для введенного критерия. Найдено решение с низкой вычислительной сложностью и предложен алгоритм расчета нижней границы для нормированного отношения длительности буферизации к просмотру по всевозможным алгоритмам планирования, удовлетворяющим допущениям из подраздела 2.7 (подразделы 3.3, 3.4 и 3.5);

- Предложен эвристический алгоритм планирования для неадаптивных видеопотоков, реализующий концепцию совместного планирования распределения ресурсов на базовой станции. Показано, что производительность предложенного планировщика превосходит известные решения и сравнима с нижней границей. Результаты настоящего раздела демонстрировались в системе моделирования NS-3, для стандарта связи LTE, соответствующего модели системы передачи информации, введенной в разделе 2 (подразделы 3.6 и 3.7).

4 ОПТИМИЗАЦИЯ ПЕРЕДАЧИ АДАПТИВНЫХ ВИДЕОПОТОКОВ В БЕСПРОВОДНЫХ ЦЕНТРАЛИЗОВАННЫХ СЕТЯХ

4.1 Вводные замечания

Ранее в разделе 3 было проведено исследование максимально возможной производительности алгоритмов распределения ресурсов радиоканала на базовой станции в беспроводных централизованных сетях для передачи неадаптивных видеопотоков. Несмотря на широкое использование неадаптивных видеоплееров для передачи видеоданных, в настоящее время все большую популярность приобретают адаптивные технологии передачи видео по протоколу HTTP (подраздел 1.4). Наиболее ярким представителем подобных технологий является стандарт Dynamic Adaptive Streaming over HTTP (DASH) [3; 6].

Данная технология позволяет, за счет хранения на видео контент сервере каждого видео в нескольких наборах качествах, пользовательскому устройству подобрать качество под конкретные условия в сети передачи видеоданных. Следствием данного факта является возможность снижения нагрузки на сеть передачи информации за счет адаптации видеоряда. Важным исключительным качеством данной технологии является инвариантность по отношению к сети передачи информации, обеспечиваемая совместным использованием сетевых протоколов HTTP (уровень приложения) и TCP (транспортный уровень). Однако, использование настоящей технологии вносит небольшую избыточность при передаче видео из-за использования кэширования пакетов протоколом TCP.

В настоящий момент широкое распространение стандарта DASH обеспечивается его использованием сервисом хранения видеоконтента YouTube, который является самым крупным хранилищем видео в сети Интернет.

Настоящий раздел посвящен аналитическому исследованию производительности беспроводных централизованных систем передачи информации с доминированием передачи видео по протоколу прикладного уровня HTTP с возможностью его адаптации под специфичные условия сети передачи информации. Данный раздел неотступно следует системе допущений, введенной в подразделе 2.7. В начале раздела предлагается критерий качества обслуживания для адаптивных видеопотоков. Следующим шагом формулируется оптимизационная задача

и предлагается нижняя граница для введенного критерия по всем возможным алгоритмам распределения ресурсов беспроводного канала и адаптации видеоряда. В завершении предлагается численный пример, демонстрирующий производительность стандартных алгоритмов планирования распределения ресурсов беспроводного канала по сравнению с полученной нижней границей.

Основные результаты данного раздела опубликованы в работе.

4.2 Критерии качества восприятия адаптивного видеопотока

В подразделе 3.2 был проведен анализ основных объективных показателей воспроизведения видео на оценку качества восприятия видеопотока, описанных ранее в подразделах 1.5 и 2.6:

- Битовая скорость потока, характеризующая качество видео;
- Длительность буферизации, включающая в себя длительности всех начальных и повторных буферизаций во время просмотра;
- Гладкость воспроизведения видеоряда.

Рассмотрим каждый из них с позиции адаптивной передачи видеоданных.

Как было сказано ранее в подразделе 1.4, принципиальным отличием адаптивных видеоплееров от неадаптивных является наличие механизма подбора репрезентации видео в зависимости от характеристик канала передачи информации (адаптация репрезентации). Основной целью механизма адаптации репрезентации видеоряда является минимизация общей длительности буферизации (начальные и повторные буферизации в течении просмотра видео), так как данный фактор имеет сильное влияние на качество восприятия видео пользователем. Однако, исходя из описания критериев качества восприятия в подразделе 1.5, следует, что помимо общей длительности буферизации, определяющую роль при оценке качества восприятия исполняет битовая скорость потока, характеризующая качество видеоряда. Это демонстрирует сложную взаимосвязь между битовой скоростью потока и длительностью буферизации при построении управления для адаптивных видеоплееров.

Продemonстрируем данный факт на примере. Пусть для некоторого мобильного пользователя скорость получения информации равняется 0.65 Мбит/с и по-

стоянна во времени. Исходя из таблицы 1.1, данная скорость получения информации может обеспечить просмотр пользователю с качеством видео 360p (0.45 Мбит/с) без возникновения повторных буферизаций во время просмотра, или в качестве 480p (0.7 Мбит/с) с наличием повторными буферизаций, длительностью 0.05 от длительности просмотра. Однако, битовая репрезентация задает верхнюю границу удовлетворенности пользователя при просмотре: для данного примера при просмотре видео в идеальных условиях (минимально возможная длительность начальной буферизации и отсутствие повторных) в качестве 360p значение MOS (методология U-vMOS) не может превышать значения 3, а для 480p – 3.64. Основываясь на данном факте, обеспечение пользователю просмотра видео в качестве 360p без повторных буферизаций не приводит к его достаточной удовлетворенности качеством обслуживания, так как пользователь считается удовлетворенным если величина MOS превышает значение 3.

Приведенный выше пример демонстрирует важную специфику организации управления полезной скоростью получения информации для адаптивных видеоплееров: *безусловная минимизация общей длительности буферизации не приводит к максимизации удовлетворенности пользователя просмотром видео*. Следствием из данного факта является необходимость учета битовой скорости потока в критерии качества восприятия для адаптивных видеоплееров.

Из анализа работы адаптивных видеоплееров, проведенному в подразделе 1.4, следует, что гладкость воспроизведения видеоряда обеспечивается алгоритмом выбора репрезентации, установленном на видеоплеере и ограничено сверху некоторой величиной. Подробное описание данного факта приведено в подразделе 2.6.

Обобщая описанное выше, при формировании критерия качества восприятия адаптивного видеопотока необходимо учитывать два фактора воспроизведения: длительность буферизаций и битовую скорость потока.

В качестве основания для построения критерия качества восприятия в настоящем разделе было выбрано отношение длительности буферизации к просмотру (определение 2.2), характеризующее длительность буферизаций:

$$q_i = \lim_{T \rightarrow \infty} \frac{b_i^T}{w_i^T},$$

где b_i^T – общая длительность буферизации пользователя i за время T , w_i^T – общая длительность просмотра видео пользователем i за время T .

Данный критерий позволяет оценить удовлетворенность пользователя:

$$q_i = \begin{cases} 0, & \text{пользователь } i \text{ удовлетворен просмотром} \\ (0, \infty) & \text{пользователь } i \text{ наблюдает негативные эффекты при просмотре} \end{cases}.$$

Чем выше значение q_i , тем ниже качество восприятия видео пользователем (присутствует обратная корреляция между критерием качества q_i и удовлетворенностью пользователем просмотром). Впервые критерий качества q_i был рассмотрен в работе [40].

В качестве критерия качества, описывающий систему в целом, было выбрано среднее значение отношение длительности буферизации к просмотру всех пользователей в системе:

$$\bar{q}(\mathcal{A}, \mathcal{B}) = \frac{1}{N} \left(\sum_{i=1}^N q_i(\mathcal{A}, \mathcal{B}) \right), \quad (4.1)$$

где $\mathcal{A}(\cdot)$ – алгоритм планирования распределения ресурсов на базовой станции, $\mathcal{B}(\cdot)$ – алгоритм адаптации битовой репрезентации.

В отличие от критерия качества восприятия, представленного в подразделе 3.2, наличие адаптации видео ряда вносит более сложную зависимость качества восприятия от параметров системы. Она выражается в наличии обратной связи при взаимной работе алгоритмов $\mathcal{A}(\cdot)$ и $\mathcal{B}(\cdot)$: исходя из скорости получения информации обеспеченной алгоритмом $\mathcal{A}(\cdot)$, алгоритм $\mathcal{B}(\cdot)$ выбирает битовую скорость для заказываемых сегментов, которая влияет на загрузку беспроводного канала, и как следствие на распределение его ресурсов алгоритмом $\mathcal{A}(\cdot)$ и скорость передачи информации. Данный факт еще раз демонстрирует сложность оптимизации передачи адаптивных видеопотоков в централизованных беспроводных сетях.

В качестве оценки качества репрезентации видеопотока для системы в целом, в настоящем разделе предлагается рассматривать среднюю битовую скорость видеопотока всех пользователей, присутствующих в системе:

$$\bar{R}(\mathcal{A}, \mathcal{B}) = \frac{1}{N} \left(\sum_{i=1}^N \tilde{R}_i(\mathcal{A}, \mathcal{B}) \right), \quad (4.2)$$

где \tilde{R}_i – математическое ожидание битовой скорости видеопотока пользователя i (подраздел 2.9).

В данном подразделе проведен анализ основных факторов, влияющих на качество восприятия адаптивного видеопотока. Результатом данного анализа является выбор двух критериев качества: отношение длительности буферизации к просмотру (4.1) и средняя битовая скорость видеопотока всех пользователей в системе (4.2). В дальнейшем, на основе данных критериев будет сформирована оптимизационная задача, для исследования максимальной производительности беспроводных централизованных сетей связи при передаче адаптивных видеопоследовательностей.

4.3 Постановка оптимизационной задачи

В данном подразделе предлагается и проводится анализ оптимизационной задачи для исследования максимально возможной производительности беспроводных централизованных сетей связи при передаче адаптивных видеопоследовательностей. В подразделе 4.2 было произведено рассмотрение факторов, влияющих на качество восприятия адаптивного видео, и выделены два основных: (4.1) и (4.2). Настоящий подраздел построен следующим образом: в начале предлагается обобщенный критерий качества, объединяющий критерии (4.1) и (4.2). На основе сформированного обобщенного критерия качества восприятия предлагается оптимизационная задача. В завершении подраздела производится ее анализ.

Важным выводом из подраздела 4.2 является необходимость условной оптимизации фактора длительности буферизации: $\bar{q}(\mathcal{A}, \mathcal{B})$, при передачи адаптивных видеопотоков. В качестве второго фактора, влияющего на качество восприятия, была выделена средняя битовая скорость видео по всем пользователям в системе: R_{avg} , выступающая в роли ограничения при оптимизации фактора длительности буферизации. Основываясь на данных факторах, в настоящем разделе обобщенный критерий качества обслуживания для адаптивного видеоряда вводится в следующем виде:

$$Q = \inf_{\mathcal{A}, \mathcal{B}: \bar{R}(\mathcal{A}, \mathcal{B}) \geq R_{avg}} \bar{q}(\mathcal{A}, \mathcal{B}). \quad (4.3)$$

Таким образом, целью настоящего раздела является нахождение нижней границы отношения общей длительности буферизации к просмотру по всем возможным алгоритмам планирования распределения ресурсов на базовой станции $\mathcal{A}(\cdot)$ и алгоритмам выбора репрезентации $\mathcal{B}(\cdot)$, удовлетворяющих допущениям введенных в подразделе 2.7, таким что среднее средняя битовая скорость видеопотоков, просматриваемых в данной системе, больше или равно некоторой величине R_{avg} . Важно отметить, что в настоящем критерии число алгоритмов $\mathcal{A}(\cdot)$ и $\mathcal{B}(\cdot)$ в отдельности бесконечно и несчетно, следовательно, число их комбинаций так же бесконечное и несчетное число. Основываясь на данном факте, в рамках настоящего раздела будет произведен поиск именно нижней границы (инфинума) по всевозможным алгоритмам $\mathcal{A}(\cdot)$ и $\mathcal{B}(\cdot)$.

Важно отметить, что в настоящем разделе используется система допущений из подраздела 2.7 без дополнительных изменений.

Так как критерий качества восприятия q_i имеет обратную корреляцию со степенью удовлетворенности пользователя просмотром, следовательно, формируемая оптимизационная задача относится к задачам минимизации. Важность роли битовой скорости, при оценке качества восприятия, определяет необходимость условной оптимизации с ограничением снизу на среднюю битовую скорость потока, просматриваемую пользователями в системе, некоторой величиной. Данное ограничение позволяет гарантировать достаточное качество репрезентации видео для обеспечения удовлетворительных значений качества восприятия. Из подраздела 1.3 следует, что битовая скорость каждого пользователя может принимать значения в некотором ограниченном отрезке $[R_{min}, R_{max}]$. Неизменность системы допущения позволяет использовать утверждение 2.1 в виде, представленном в подразделе 2.9. Обобщая сказанное выше, оптимизационная задача для критерия качества (4.3), описывающая максимальную производительность передачи адаптивных видеопотоков в беспроводных централизованных сетях, принимает следующий вид:

Минимизировать: $\frac{1}{N} \sum_{i=1}^N q_i$

При условии:

$$\begin{cases} \sum_{i=1}^N (1 - \nu_i^R \nu_i^C) \frac{\tilde{R}_i \tilde{C}_i^{-1}}{q_i + \gamma_i} - 1 \leq 0 \\ -\frac{1}{N} \sum_{i=1}^N \tilde{R}_i + R_{avg} \leq 0 \\ \tilde{R}_i \in [R_{min}, R_{max}], i = \overline{1, N} \\ -q_i \leq 0, i = \overline{1, N} \end{cases} . \quad (4.4)$$

Проведем анализ задачи (4.4) на качественном уровне. В настоящей оптимизационной задаче $2N$ переменных: $q = (q_1, q_2, \dots, q_N)$ и $\tilde{R} = (\tilde{R}_1, \tilde{R}_2, \dots, \tilde{R}_N)$. Из вида первого ограничения следует, что оптимизационная задача (4.4) относится к классу задач нелинейного программирования. Важнейшим этапом анализа оптимизационных задач, определяющий возможные методы ее решения, является определение подкласса задачи нелинейного: задачи выпуклого, вогнутого или невыпуклого программирования. Завершающая часть данного подраздела посвящена определению подкласса задачи (4.4).

Оптимизационная задача (4.4) обладает целевой функцией линейного вида и линейными ограничениями типа неравенства, исключая первое ограничение, основываясь на данном факте, тип задачи будет определяться видом первого ограничения. Стоит отметить, что данное ограничение это функция от $2N$ переменных. Для его анализа рассмотрим случай, когда в системе находится только один абонент ($N = 1$), в данном случае число переменных уменьшается до двух. Для упрощения последующих математических выкладок в выражении (4.5) номер абонента был опущен, так как в системе присутствует только он один.

$$(1 - \nu^R \nu^C) \frac{\tilde{R} \tilde{C}^{-1}}{q + \gamma} - 1. \quad (4.5)$$

Определене вида функции от нескольких переменных возможно на основе матрицы Гессе [32]. В зависимости вида матрицы Гессе возможно определить вид функции следующим образом: если матрица Гессе является

- Положительно определенной во всех точках, то функция выпуклая;
- Отрицательно определенной во всех точках, то функция вогнутая;
- Не является знакоопределенной, то функция общего вида (невыпуклая).

Матрица Гессе $H(\cdot)$ для функции (4.5) принимает следующий вид:

$$H(\tilde{R}, q) = \begin{bmatrix} 0 & -\frac{(1-\nu^R\nu^C)\tilde{C}^{-1}}{(q+\gamma)^2} \\ -\frac{(1-\nu^R\nu^C)\tilde{C}^{-1}}{(q+\gamma)^2} & \frac{(1-\nu^R\nu^C)\tilde{C}^{-1}\tilde{R}}{(q+\gamma)^3} \end{bmatrix}$$

Так как матрица Гессе является симметричной относительно главной диагонали, то, для определения ее знакоопределенности, возможно использовать критерий Сильвестра: симметричная матрица является положительно определенной тогда и только тогда, когда все ее угловые миноры положительны, в случае неотрицательной определенности – больше или равны нулю. Отрицательная знакоопределенность симметричной матрицы задается чередованием знаков угловых миноров: отрицательный, положительный и т.д. В иных случаях матрица не является знакоопределенной [32].

Найдем значения угловых миноров Δ матрицы $H(\tilde{R}, q)$:

$$\Delta_1 = 0 \cdot \frac{(1 - \nu^R\nu^C) \tilde{C}^{-1}\tilde{R}}{(q + \gamma)^3} = 0.$$

$$\Delta_2 = 0 \cdot \frac{(1 - \nu^R\nu^C) \tilde{C}^{-1}\tilde{R}}{(q + \gamma)^3} - \left(-\frac{(1 - \nu^R\nu^C) \tilde{C}^{-1}}{(q + \gamma)^2} \right)^2 = - \left(\frac{(1 - \nu^R\nu^C) \tilde{C}^{-1}}{(q + \gamma)^2} \right)^2 < 0.$$

Так как значение $(1 - \nu^R\nu^C)$ и сумма значений q и γ отличны от нуля (система допущений, подраздел 2.7), то значение второго углового минора (Δ_2) строго отрицательно. Таким образом, основываясь на критерии Сильвестра, функция (4.5) является функцией общего вида (невыпуклой). Следовательно, оптимизационная задача (4.4) относится к подклассу невыпуклых задач.

Критично важно отметить, что для задач невыпуклого программирования не существует стандартных методов или подходов к решению [32; 33]. Следовательно, решение оптимизационной задачи (4.4) является нетривиальной задачей. В настоящем разделе будет предложено решение данной задачи, которое будет основано на двухступенчатой оптимизации. Дальнейшая часть раздела, посвященная решению настоящей задачи, будет построена следующим образом. На первом шаге будет представлено решение задачи выпуклого программирования на основе теоремы Каруша-Куна-Таккера (ККТ) с использованием метода «*water-filling*», описанного в [32]. Далее будет предложен алгоритм вычисления нижней

границы для оптимизационной задачи (4.4), который использует решение задачи на основе теоремы ККТ.

4.4 Решение вспомогательной оптимизационной задачи выпуклого программирования на основе теоремы Каруша-Куна-Таккера

В настоящем подразделе представлено решение задачи выпуклого программирования при наличии ограничения типа неравенств на основе теоремы Каруша-Куна-Таккера (ККТ) [32; 33]. Представленное решение было описано в работе [40].

Рассмотрим оптимизационную задачу следующего вида:

$$\begin{aligned} &\text{Минимизировать: } Q^* = \frac{1}{N} \sum_{i=1}^N q_i \\ &\text{При условии:} \\ &\begin{cases} \sum_{i=1}^N \frac{\tilde{a}_i}{q_i + \gamma_i} - L^* \leq 0 \\ -q_i \leq 0, i = \overline{1, N} \end{cases}, \end{aligned} \quad (4.6)$$

где $\tilde{a}_i, i = \overline{1, N}$ и L^* являются положительными и отличными от нуля числами.

Утверждение 4.1. *Оптимальное значение Q^* может быть получено на основе выражения (4.7):*

$$Q^* = \begin{cases} 0, & \text{если } \sum_{i=1}^N \left(\frac{\tilde{a}_i}{\gamma_i} \right) \leq L^* \\ \frac{1}{N} \sum_{i=1}^N \max(\sqrt{\mu_0 \tilde{a}_i N} - \gamma_i, 0), & \text{иначе} \end{cases}, \quad (4.7)$$

где μ_0 – решение уравнения (4.8):

$$\sum_{i=1}^N \frac{\tilde{a}_i}{\max(\sqrt{\mu_0 \tilde{a}_i N}, \gamma_i)} - L^* = 0. \quad (4.8)$$

Доказательство. Из вида первого ограничения следует, что оптимизационная задача (4.6) относится к классу задач нелинейного выпуклого программирования.

ния. Следовательно, для данной задачи могут быть найдены условия существования минимума ККТ, и существует единственная точка на многомерной плоскости, удовлетворяющая настоящим условиям. Теорема ККТ позволяет свести задачу выпуклого программирования к поиску седловой точки функции Лагранжа на множестве определенном ограничениями на неотрицательность значений множителей Лагранжа, принадлежность решения области допустимых значений [33].

Изначально для оптимизационной задачи (4.6) найдем функцию Лагранжа $\mathcal{L}(\cdot)$:

$$\mathcal{L}(\mathbf{q}, \boldsymbol{\mu}) = \frac{1}{N} \sum_{i=1}^N q_i + \mu_0 \left[\left(\sum_{i=1}^N \frac{\tilde{a}_i}{q_i + \gamma_i} \right) - L^* \right] - \mu_i q_i, \quad (4.9)$$

где $\mathbf{q} = (q_1, q_2, \dots, q_N)$ и $\boldsymbol{\mu} = (\mu_0, \mu_1, \dots, \mu_N)$ являются векторами длины N и $(N + 1)$ соответственно. Вектор $\boldsymbol{\mu}$ используется для обозначения множителей Лагранжа.

На основе (4.9), условия существования решения задачи (4.6) Каруша-Куна-Таккера представлены следующей системой:

$$\left\{ \begin{array}{ll} \frac{1}{N} - \mu_0 \frac{\tilde{a}_i}{(q_i + \gamma_i)^2} - \mu_i = 0, i = \overline{1, N} & \text{Стационарность} \\ -q_i \leq 0, i = \overline{1, N} & \text{Выполнимость } 1 \\ \sum_{i=1}^N \frac{\tilde{a}_i}{q_i + \gamma_i} - L^* \leq 0 & \text{Выполнимость } 2 \\ \mu_j \geq 0, j = \overline{0, N} & \text{Неотрицательность} \\ \mu_i q_i = 0, i = \overline{1, N} & \text{Дополняющая нежесткость } 1 \\ \mu_0 \left[\left(\sum_{i=1}^N \frac{\tilde{a}_i}{q_i + \gamma_i} \right) - L^* \right] = 0 & \text{Дополняющая нежесткость } 2 \end{array} \right. \quad (4.10)$$

Теорема ККТ утверждает, что решение оптимизационной задачи (4.6) удовлетворяет всем условиям, представленным в системе (4.10). Одновременно с этим, выпуклость задачи обеспечивает единственность решения настоящей системы. Важно отметить, что условия ККТ могут быть применены и к задачам невыпуклого программирования, однако, в таком случае может существовать несколько возможных решений систем, аналогичных (4.10), так как в таких задачах возможно будет существовать множество локальных экстремумов. Данный факт приводит к необходимости организации перебора по полученным решениям

для поиска глобального экстремума при использовании подхода на основе условий ККТ для задач невыпуклого программирования.

В обозначениях к системе (4.6) использовался адаптированный перевод англоязычных терминов:

- «*Stationarity*» – **Стационарность**. Равенство градиента функции Лагранжа нулю;
- «*Primal feasibility*» – **Выполнимость**. Ограничения типа неравенства в каноническом виде;
- «*Dual feasibility*» – **Неотрицательность**. Ограничение на положительность значений множителей Лагранжа;
- «*Complimentary slackness*» – **Дополняющая нежесткость**.

Продолжим рассмотрение условий существования экстремума (4.6). Из условия **Стационарность** выразим значение μ_i :

$$\mu_i = \frac{1}{N} - \mu_0 \frac{\tilde{a}_i}{(q_i + \gamma_i)^2}, i = \overline{1, N}.$$

Следовательно, условие **Дополняющая нежесткость 1** преобразуется к виду:

$$q_i \left[\frac{1}{N} - \mu_0 \frac{\tilde{a}_i}{(q_i + \gamma_i)^2} \right] = 0, i = \overline{1, N}. \quad (4.11)$$

Уравнение (4.11) задает зависимость значений q_i от значения только одного множителя Лагранжа μ_0 . Проведем последовательный анализ выражения (4.11) для конкретного пользователя i в зависимости от значений μ_0 .

Пусть $\mu_0 \in \left[0, \frac{\gamma_i^2}{a_i N}\right)$, тогда выражение $\left[\frac{1}{N} - \mu_0 \frac{\tilde{a}_i}{(q_i + \gamma_i)^2}\right]$, равное μ_i , принимает значения строго больше нуля. Следовательно, при данных значениях μ_0 , выражение (4.11) имеет единственное решение $q_i = 0$.

Если $\mu_0 \in \left[\frac{\gamma_i^2}{a_i N}, \infty\right)$, то значение в квадратных скобках (4.11), равное μ_i , принимает отрицательное значение при $q_i = 0$, однако, в соответствии с условием **Неотрицательность**, $\mu_i \geq 0$. Следовательно, при данных значениях μ_0 решением уравнения (4.11) является $q_i = \sqrt{\mu_0 \tilde{a}_i N} - \gamma_i$. Объединяя два полученных выше результата анализа, решение уравнения (4.11) определяется следующим выражением:

$$q_i = \max \left(\sqrt{\mu_0 \tilde{a}_i N}, \gamma_i \right). \quad (4.12)$$

Таким образом, задача нахождения значений вектора \mathbf{q} может быть сведена к нахождению одного значения множителя Лагранжа μ_0 . Значение μ_0 может быть получено из решения условия **Дополняющая нежесткость 2**, подставив в него (4.12) получим следующее уравнение:

$$\mu_0 \left[\sum_{i=1}^N \frac{\tilde{a}_i}{\max(\sqrt{\mu_0 \tilde{a}_i N}, \gamma_i)} - L^* \right] = 0.$$

Данное уравнение имеет два возможных решения:

- $\mu_0 = 0$. В данном случае $\forall i : q_i = 0$, если при данных значениях вектора \mathbf{q} выполняется условие **Выполнимость 2**.
- μ_0 находится из решения уравнения (4.8).

Настоящий факт завершает доказательство. □

Описанный подход к нахождению решения системы условий существования экстремума в англоязычной литературе известен под названием *water-filling* и представлен в [32].

Обобщая сказанное выше, решение оптимизационной задачи (4.6) может быть получено из решения уравнения (4.8). Важно отметить, что (4.8) обладает монотонной завистью от значения μ_0 . Для нахождения решения данного уравнения может быть предложена двухэтапная численная процедура (алгоритм 4.1).

Алгоритм 4.1 : Численное решение уравнения (4.8)

$$1: \text{Найти отрезок } [k, r] : \begin{cases} \sum_{i=1}^N \frac{\tilde{a}_i}{\max(\sqrt{k \tilde{a}_i N}, \gamma_i)} \geq L^* \\ \sum_{i=1}^N \frac{\tilde{a}_i}{\max(\sqrt{r \tilde{a}_i N}, \gamma_i)} \leq L^* \end{cases}.$$

алгоритмом экспоненциального шага: удвоение значения r , до тех пор пока не найдется значение, удовлетворяющее условиям.

- 2: Найти решение уравнения (4.8) на отрезке $[k, r]$ алгоритмом дихотомии (деления отрезка пополам).
-

Графическое представление алгоритма 4.1 приведено на рисунке 4.1.

Следующим этапом рассмотрения решения оптимизационной задачи (4.6) является оценка сложности полученного решения. Численный алгоритм 4.1 обладает логарифмической сложностью: $O(N \log_2[r - k])$, $r \rightarrow \infty$, $k \rightarrow \infty$, $N \rightarrow \infty$. Однако, значения r и k заранее неизвестны и зависят от входных условий задачи. В ходе тестирования алгоритма 4.1, применительно к задачам передачи

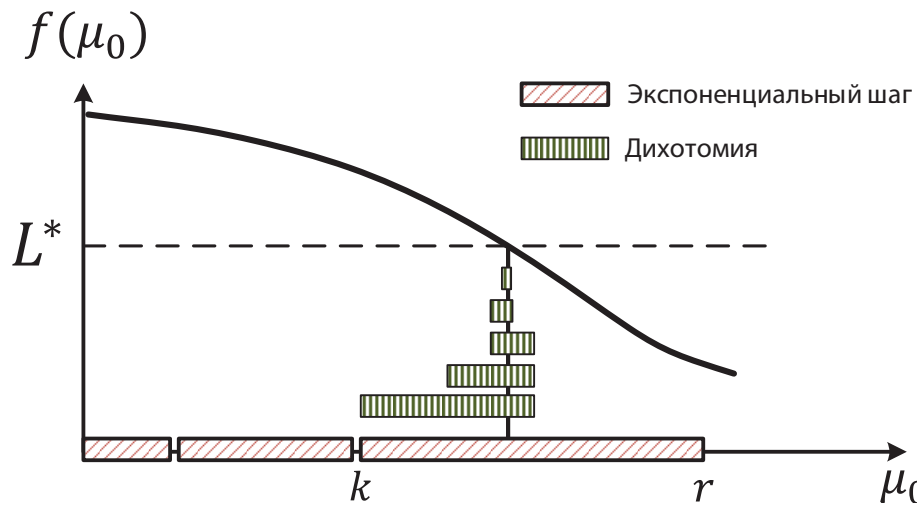


Рисунок 4.1 — Графическое представление алгоритма численного нахождения решения

видео, значения разности $[r - k]$ не превышали N . На основе данного факта справедливо следующее: сложность алгоритма 4.1, нахождения решения оптимизационной задачи (4.6), является сопоставимой со сложностью алгоритма 3.1: $O(N \log_2 N)$, $N \rightarrow \infty$ (подраздел 3.4, таблица 3.2).

4.5 Нижняя граница отношения длительности буферизации к просмотру при передаче адаптивных видеопотоков

Ранее в подразделе 4.3 была предложена оптимизационная задача невыпуклого программирования (4.4), определяющая максимально возможную производительность беспроводных централизованных систем для передачи адаптивных видеопотоков. Данная оптимизационная задача учитывает два основных фактора, влияющие на качество восприятия адаптивного видео: отношение общей длительности буферизации к просмотру (q_i) и средняя битовая скорость просмотренного видео (\tilde{R}_i), представленные в подразделе 4.2. Сложность получения решения задачи (4.4) обусловлена невыпуклостью ее ограничения, что приводит к отсутствию стандартных методов и подходов к решению. Оптимизационная задача (4.4) имеет следующий вид:

Минимизировать: $\frac{1}{N} \sum_{i=1}^N q_i$

При условии:

$$\begin{cases} \sum_{i=1}^N (1 - \nu_i^R \nu_i^C) \frac{\tilde{R}_i \tilde{C}_i^{-1}}{q_i + \gamma_i} - 1 \leq 0 \\ -\frac{1}{N} \sum_{i=1}^N \tilde{R}_i + R_{avg} \leq 0 \\ \tilde{R}_i \in [R_{min}, R_{max}], i = \overline{1, N} \\ -q_i \leq 0, i = \overline{1, N} \end{cases}.$$

Предлагаемый подход для решения задачи (4.4) состоит в разделении ее решения на два этапа.

На первом этапе будет рассмотрена промежуточная задача (4.13):

Максимизировать: $L = \sum_{i=1}^N l_i$

При условии:

$$\begin{cases} \sum_{i=1}^N l_i \tilde{R}_i - 1 \leq 0 \\ -\frac{1}{N} \sum_{i=1}^N \tilde{R}_i + R_{avg} \leq 0 \\ l_i \in \left[0, \frac{(1 - \nu_i^R \nu_i^C) \tilde{C}_i^{-1}}{\gamma_i} \right], \quad i = \overline{1, N} \\ \tilde{R}_i \in [R_{min}, R_{max}], \quad i = \overline{1, N} \end{cases}. \quad (4.13)$$

Оптимизационная задача (4.13) является задачей с линейной целевой функцией и одним невыпуклым ограничением квадратичного вида. Важным отличием оптимизационной задачи (4.13) от (4.4) состоит в виде первого ограничения: ограничение в задаче (4.4) является *ограничением общего вида* (подраздел 4.3), когда ограничение в задаче (4.4) является *ограничением квадратичного вида*. Это небольшое отличие оптимизационных задач, имеет колоссальное влияние на методику решения. Для задач квадратичного программирования, с невыпуклым ограничением квадратичного вида существует эффективный алгоритм решения, построенный на основе метода многогранной аппроксимации (в англоязычной литературе «*polyhedral underestimating method*»), в сочетании со стандартными численными методами решения задач выпуклого программирования [41]. Описание данного численного метода решения задач квадратичного программирования

ния с невыпуклым квадратичным ограничением является очень громоздким и не приводится в рамках настоящей работы, так как не является результатом, полученным автором.

Оптимизационная задача (4.13) сохраняет все ограничения задачи (4.4), но вводит обозначение:

$$l_i = (1 - \nu_i^R \nu_i^C) \frac{\tilde{C}_i^{-1}}{q_i + \gamma_i}.$$

В соответствие с введенным обозначением происходит преобразование ограничений оптимизационной задачи (4.4). Так как $q_i, i = \overline{1, N}$ может принимать значения в $[0, \infty)$, то значения $l_i, i = \overline{1, N}$ должны принимать значения на отрезке $\left[0, \frac{(1 - \nu_i^R \nu_i^C) \tilde{C}_i^{-1}}{\gamma_i}\right]$, что отражено в задаче (4.13).

Взаимосвязь решений задач (4.4) и (4.13) описывается утверждением 4.2.

Утверждение 4.2. Для оптимальных значений $\{q_i^*\}$ задачи (4.4) и оптимального решения L^* задачи (4.13) справедливо следующее неравенство:

$$\sum_{i=1}^N \frac{\tilde{a}_i}{q_i^* + \gamma_i} \leq L^*,$$

где $\tilde{a}_i = (1 - \nu_i^R \nu_i^C) \tilde{C}_i^{-1}$.

Доказательство. Доказательство утверждения 4.2 построено на основе метода от противного. Предположим, что это не так, тогда найдется

$$\sum_{i=1}^N \frac{\tilde{a}_i}{q_i^* + \gamma_i} > L^*.$$

Следовательно, существует $l_i^o = \frac{\tilde{a}_i}{q_i^* + \gamma_i}, i = \overline{1, N}$, удовлетворяющие ограничениям оптимизационной задачи (4.13). Тогда $\sum_{i=1}^N l_i^o > L^*$, однако, из данного соотношения следует, что значение L^* не было решением задачи (4.13). Таким образом, было найдено противоречие. Настоящий факт завершает доказательство. \square

Используя результат утверждения 4.2 предлагается постановка оптимизационной задачи (4.4) с ослабленными ограничениями:

$$\text{Минимизировать: } \frac{1}{N} \sum_{i=1}^N q_i$$

При условии:

$$\begin{cases} \sum_{i=1}^N \frac{\tilde{a}_i}{q_i + \gamma_i} - L^* \leq 0 \\ -q_i \leq 0, i = \overline{1, N} \end{cases}.$$

Данная постановка задачи повторяет оптимизационную задачу (4.6), представленную в подразделе 4.4. Обозначения, используемые в настоящей постановке и задаче (4.6), согласованы. Основываясь ее решении, так же приведенному в подразделе 4.4, и обобщая ранее приведенные результаты, предлагается следующая теорема.

Теорема 4.1. *Нижняя граница отношения длительности буферизации к просмотру при передаче адаптивных видеопотоков может быть вычислена на основе алгоритма 4.2.*

Алгоритм 4.2 : Вычисление нижней границы отношения длительности буферизации к просмотру при передаче адаптивных видеопотоков

- 1: Вычисление L^* – решение оптимизационной задачи (4.13);
- 2: Нахождение значения μ являющегося решением уравнения:

$$\sum_{i=1}^N \frac{\tilde{a}_i}{\max(\sqrt{\mu \tilde{a}_i N}, \gamma_i)} - L^* = 0;$$

- 3: Подстановка значения μ , полученного на предыдущем шаге, в выражение для вычисления нижней границы:

$$Q \geq \begin{cases} 0, & \text{если } \sum_{i=1}^N \left(\frac{\tilde{a}_i}{\gamma_i} \right) \leq L^* \\ \frac{1}{N} \sum_{i=1}^N \max(\sqrt{\mu \tilde{a}_i N} - \gamma_i, 0), & \text{иначе} \end{cases}.$$

Завершающим этапом настоящего подраздела является анализ сложности вычисления нижней границы отношения длительности буферизации к просмотру

при передаче адаптивных видеопотоков. При рассмотрении сложности каждого из шагов алгоритма 4.2, наибольшей вычислительной сложностью обладает шаг 1, так как на данном шаге для поиска решения оптимизационной задачи (4.13) используются численные методы нахождения минимума целевой функции при ограничениях типа неравенства. Сложность подобных методов и их сравнение с возможными аналогами было приведено в таблице 3.2 подраздел 3.4.

Таким образом, алгоритм 4.2 обладает высокой вычислительной сложностью и не может быть реализован в реальном масштабе времени при организации управления распределением ресурсов беспроводного канала на базовой станции. Однако, данное свойство не снижает значимости полученного результата для аналитической оценки производительности беспроводных централизованных сетей при передаче адаптивного видео. Настоящий результат может быть использован при планировании сетей беспроводной связи (расположение и характеристики устанавливаемых базовых станций) и создания алгоритмов планирования ресурсов беспроводного канала для существующих и будущих стандартов беспроводной связи.

4.6 Численный пример

Настоящий подраздел демонстрирует результаты сравнения нижней границы для отношения длительности буферизации к просмотру, представленной в подразделе 4.5, с известными алгоритмами планирования (подраздел 2.5).

Демонстрация результатов настоящего раздела производится в соответствии с подразделом 3.7. Рассматривается работа одной соты, где пользователи находятся на одной линии по удалению от базовой станции. Состояние беспроводного канала изменяется во времени, так как в моделировании используется модель радиоканала с плоскими замираниями.

На рисунке 4.2 представлено сравнение производительности известных алгоритмов планирования (*Round Robin* и *Proportional Fair*) с нижней границей при зафиксированном числе абонентов в соте, равном 10 пользователям. По оси абсцисс отложена требуемая средняя скорость просматриваемого видеопотока в соте, а по оси ординат значение критерия качества восприятия Q . Заштрихованная

область под кривой нижней характеризует область, значения которой недостижимы для всевозможных алгоритмов планирования, удовлетворяющих введенным допущениям в подразделе 2.7, например, не существует алгоритма планирования, обеспечивающего значения критерия качества восприятия Q равном 0.05 при средней битовой скорости просматриваемых видеопотоков равной 3.75 Мбит/с, когда в сети находится 10 абонентов. Найденная нижняя граница для адаптивных видеопотоков сравнивается с уже известной нижней границей для неадаптивных, представленной в работе [40]. При моделировании и расчете характеристик для неадаптивного видео предполагалось, что все пользователи просматривают видео с битовой скоростью равной R_{avg} .

Из результатов, представленных на рисунке 4.2, следует:

- Использование адаптивной технологии передачи видеоданных позволяет достичь большей производительности беспроводной централизованной сети передачи информации.
- Планировщик *Proportional Fair* обладает большей производительностью в сравнении с алгоритмом *Round Robin*, и демонстрирует производительность близкую к нижней границе по критерию качества Q .

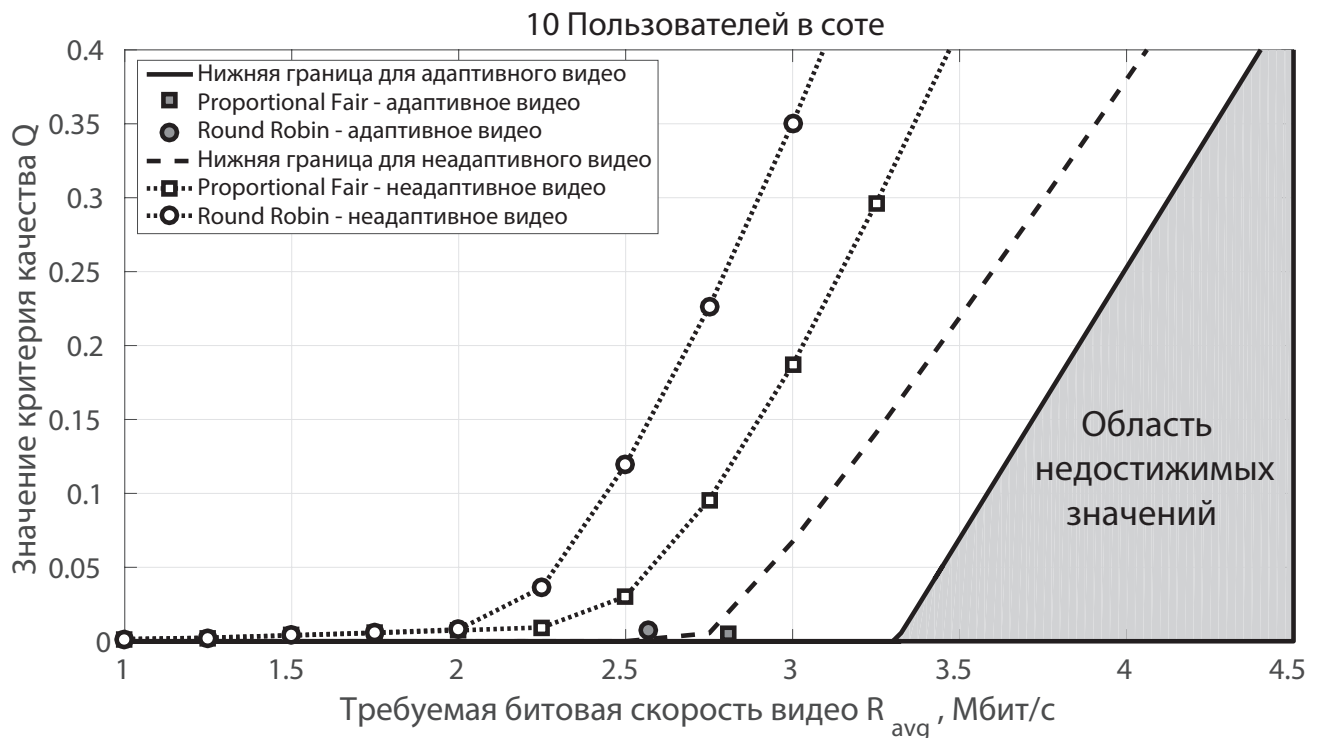


Рисунок 4.2 — Производительность известных алгоритмов планирования при передаче адаптивного видео в сравнении с нижней границей

На рисунке 4.3 представлено сравнение нижних границ для критерия Q при наличии различного числа абонентов в соте. С ростом числа абонентов при заданном уровне значений критерия Q система переходит в состояние перегрузки (определение 3.1) при меньшем значении требуемой средней битовой скорости видеопотоков в соте.

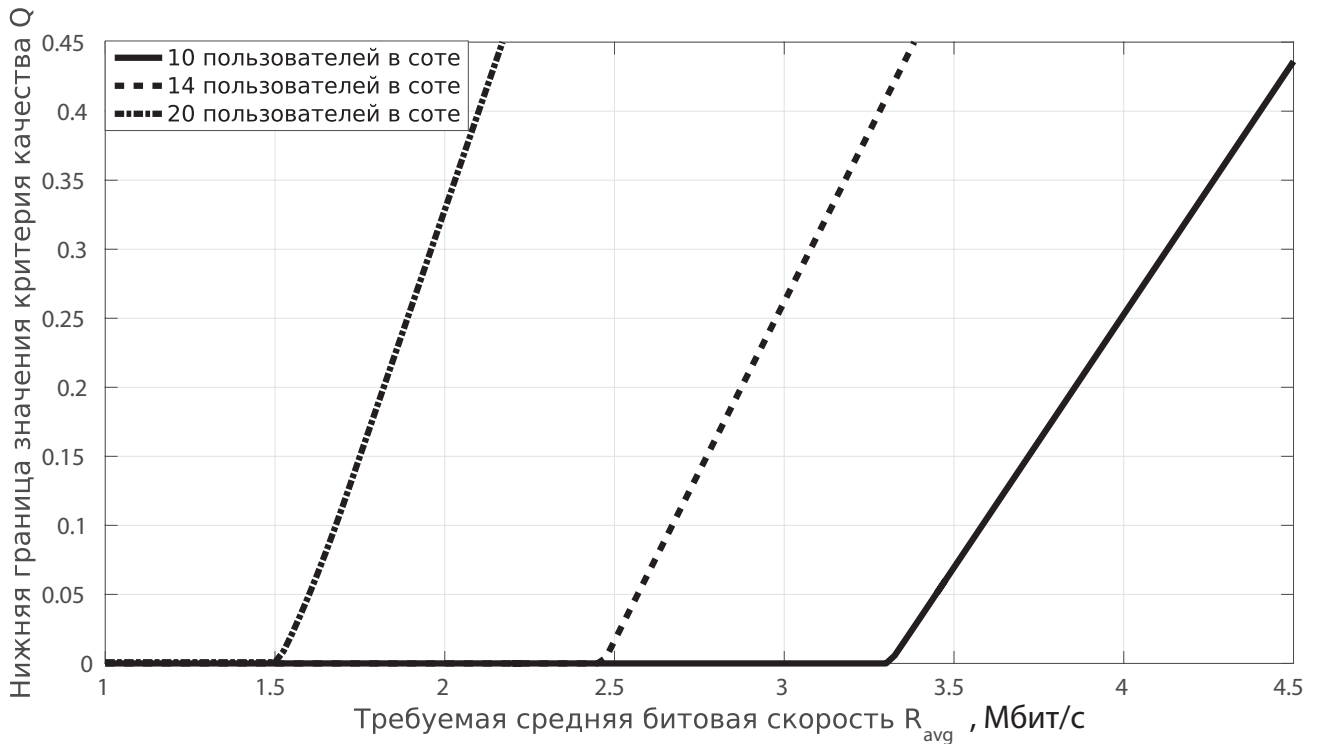


Рисунок 4.3 — Нижняя граница для критерия Q при различном числе абонентов в соте

Введем определение емкости соты:

Определение 4.1. Емкость соты – число пользователей в соте, при котором значение критерия Q равняется некоторой заданной величине.

Определение 4.1 взаимосвязано с определением перегрузки следующим образом: определение перегрузки задает границу значения критерия качества восприятия, при котором система перестает поддерживать требуемый уровень качества обслуживания. В свою очередь определение емкости соты детерминирует максимальное число пользователей, при котором система еще не перешла в состояние перегрузки.

Рисунок 4.4 иллюстрирует максимально возможную емкость соты для рассматриваемого сценария при значении критерия восприятия Q равном 0.01, 0.05

и 0.1 в зависимости от требуемой битовой скорости просматриваемых видеопотоков. Возможно отметить, что наибольшая емкость соты достигается при больших значениях критерия качества Q .

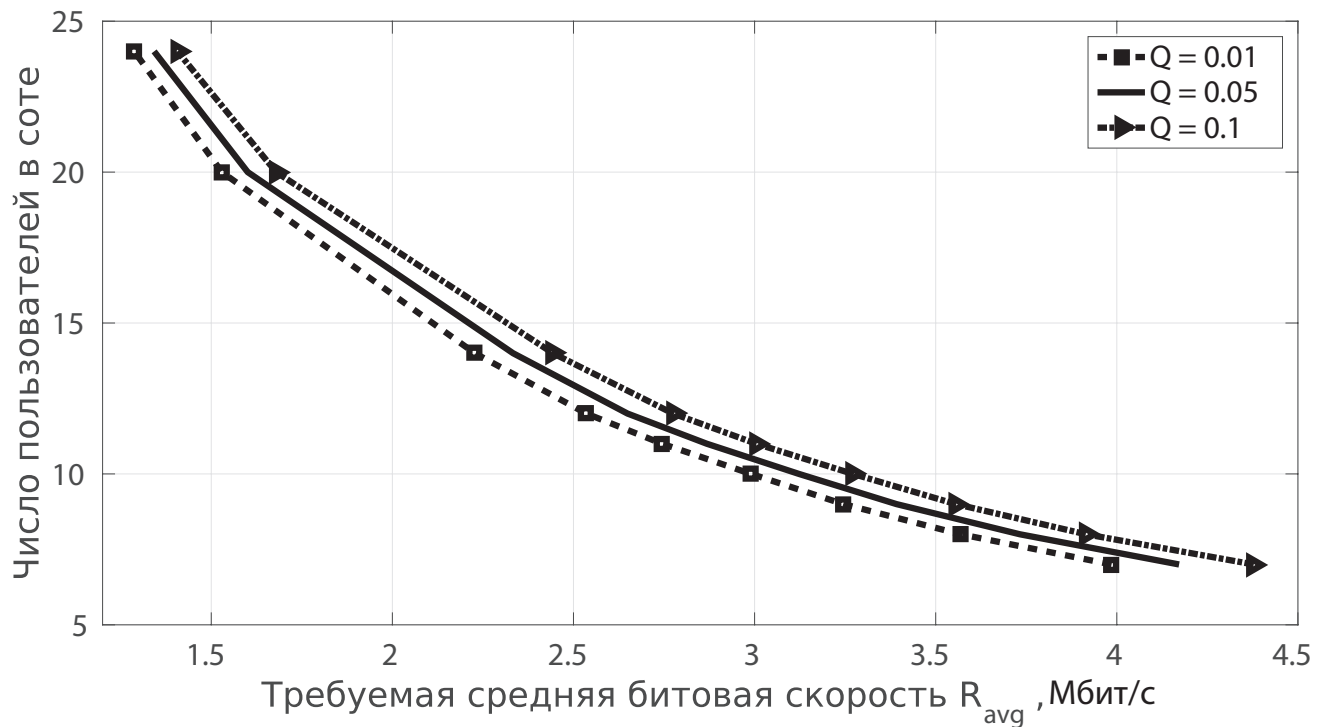


Рисунок 4.4 — Максимальная емкость соты для различных значений критерия качества Q

4.7 Выводы по разделу

Настоящий раздел рассматривал передачу адаптивных видеопотоков по протоколу HTTP в беспроводных централизованных сетях связи. В начале раздела был предложен критерий качества восприятия адаптивного видео, учитывающий два основных фактора: длительность буферизации и битовая скорость просматриваемого потока. На основе данного критерия была сформулирована оптимизационная задача невыпуклого программирования, определяющая максимальную производительность систем передачи информации для введенного критерия. Для поставленной оптимизационной задачи была найдена нижняя граница по всевозможным алгоритмам планирования распределения ресурсов беспровод-

ного канала и адаптации видеопотока, удовлетворяющих допущениям, введенным допущениям в подразделе 2.7. В заключении раздела была продемонстрировано сравнение существующих алгоритмов планирования с найденной нижней границей и определена максимально возможная емкость соты для рассматриваемого сценария при передаче адаптивных видеопотоков.

В качестве результатов настоящего раздела могут быть выделены следующие:

- Предложен критерий качества восприятия адаптивного видео: отношение длительности буферизации к просмотру при ограничении снизу на среднее значение битовой скорости просматриваемого видеопотока (подраздел 4.2);
- Для предложенного критерия произведена постановка оптимизационной задачи невыпуклого программирования. Для данной постановки найдена нижняя граница для данного критерия по всевозможным алгоритмам планирования и адаптации видеопотока, удовлетворяющих введенным допущениям в подразделе 2.7, с использованием двухступенчатой оптимизации (подразделы 4.3, 4.4 и 4.5);
- Произведена демонстрация производительности существующих алгоритмов планирования в сравнении с найденной нижней границей и определена максимально возможная емкость соты для рассматриваемого сценария (подраздел 4.6).

Настоящий раздел является заключительным в диссертационном исследовании, и для демонстрации общей картины проделанной работы в приложении В представлен граф основных зависимостей между подразделами (рисунок В.1). Нумерация узлов графа соответствует нумерации подразделов, направленные ребра (сплошные линии) обозначают прямые ссылки между подразделами (например, подраздел 3.3 ссылается на подразделы 2.7 и 2.9), пунктирные линии обозначают логическое соединение материалов подразделов.

Возможно отметить, что опорными подразделами настоящей работы являются 2.7 и 2.9, представляющие систему допущений и основополагающую взаимосвязь между характеристиками сети передачи информации и воспроизведением видеоряда соответственно. Основные результаты работы, представленные в подразделах 3.5 и 4.5, и их демонстрации 3.7 и 3.7 соответственно, непосредственно опираются на систему допущений и основополагающую взаимосвязь. В

свою очередь, материалы подразделов 2.7 и 2.9 сформированы на основе анализа передачи видео по протоколу HTTP (раздел 1) и современных беспроводных централизованных телекоммуникационных систем (подразделы 2.2, 2.4 и 2.5).

Важным моментом проведенного исследования является связь между системой передачи видеоданных и замкнутой системой массового обслуживания с конечным числом абонентов (подраздел 2.8). Из материалов подраздела 2.8 следует, что исследуемая система является обобщением замкнутой СМО с конечным числом абонентов, в которой каждая генерируемая заявка уникальна и длительности пауз между их отправками не являются экспоненциальными случайными величинами, а интенсивность обслуживания зависит от характеристик всего множества заявок, находящихся в буфере на обслуживающем устройстве. Более того, интенсивность обслуживания изменяется во времени. Настоящие отличия не позволяют применить классические методы анализа СМО и данное диссертационное исследование предлагает важные аналитические результаты для подобных систем массового обслуживания.

ЗАКЛЮЧЕНИЕ

В диссертационной работе рассмотрена передача видеоданных по протоколу НТТР в беспроводных централизованных системах связи. Основное внимание уделено аналитическим исследованиям максимально возможной производительности подобных систем при доминировании передачи видео.

Основные результаты работы заключаются в следующем.

1. Предложена модель беспроводной централизованной сети связи при передаче видеоданных по протоколу НТТР. Найдена взаимосвязь между объективными характеристиками сети и качеством воспроизведения видео. Предложенная модель и найденная взаимосвязь позволяют производить аналитические исследования производительности беспроводных централизованных сетей при доминировании передачи видеоданных.
2. Предложены алгоритмы вычисления граничных значений максимально возможной производительности беспроводных централизованных сетей для следующих критериев качества восприятия:
 - Нормированное отношение длительности буферизации к просмотру при передаче неадаптивных видеопоследовательностей;
 - Отношение длительности буферизации к просмотру с учетом средней битовой скорости видео при передаче адаптивного видео.

Полученные результаты могут быть использованы в качестве опорных значений при разработке новых алгоритмов планирования, учитывающие тип передаваемого трафика и требования к его обслуживанию, для существующих и последующих стандартов беспроводной связи.

3. Предложен алгоритм планирования распределения ресурсов беспроводного канала, позволяющий минимизировать нормированное отношение длительности буферизации к просмотру при передаче неадаптивного видео. Предложенный алгоритм позволяет увеличить емкость соты на 7 – 14% в сравнении с известными решениями и демонстрирует производительность близкую к найденной аналитической границе.

СПИСОК ЛИТЕРАТУРЫ

1. *Cisco*. Cisco VNI Mobile Forecast (2016 – 2021). — <http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html>. — 2016.
2. *Пастушок, И.А.* Обзор передачи и оценок качества восприятия видеоданных при использовании технологии адаптивной передачи видео по протоколу HTTP / И.А. Пастушок // *Информационно-управляющие системы*. — 2017. — Июнь. — Т. 3, № 88. — С. 1337–1347.
3. *ISO*. — ISO/IEC 23009-1:2014, Information technology, Dynamic adaptive streaming over HTTP (DASH); Progressive Download and Dynamic Adaptive Streaming Over HTTP (3GP-DASH), Part 1: Media presentation description and segment formats, 2014. — May.
4. *YouTube*. Рекомендуемые настройки кодирования. — <https://support.google.com/youtube/answer/1722171?hl=ru>. — 2016.
5. *Huawei*. Video as a Basic Service of LTE Networks: Mobile vMOS Defining Network Requirements. — <http://www.huawei.com/minisite/4-5g/en/industryjsdc-j.html>. — 2015.
6. *3GPP*. — 3GPP TS 26.247 v14.0.0, Transparent End-to-End Packet Switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming Over HTTP (3GP-DASH), 16 edition, 2016. — December.
7. *Dash.js*. Dash Industry Forum. — <https://github.com/Dash-Industry-Forum/dash.js/wiki>. — 2016.
8. *Pu, Wei*. Video Adaptation Proxy for Wireless Dynamic Adaptive Streaming over HTTP / Wei Pu, Zixuan Zou, Chang Wen Chen // *Proceedings of 19th International Packet Video Workshop*. — 2012. — Pp. 65–70.
9. *Азгальдов*. Экспертные методы в оценке качества товаров / Азгальдов, Райхман // *Экономика*. — 1974. — Р. 151.

10. *Laghari, K. U. Rehman*. Toward total quality of experience: A QoE model in a communication ecosystem / K. U. Rehman Laghari, K. Connelly // *IEEE Communications Magazine*. — 2012. — April. — Vol. 50, no. 4. — Pp. 58–65.
11. *International Telecommunication Union*. — ITU-T Recommendation P.800.1: Mean Opinion Score (MOS) terminology. — International Telecommunication Union, 2006. — July.
12. *International Telecommunication Union*. — ITU-T P.1201 (10/2012)), Parametric non-intrusive assessment of audiovisual media streaming quality. — International Telecommunication Union, 2012. — November.
13. *Huawei*. Video Experience-based Bearer Network Technical White Paper. — <http://www-file.huawei.com/~media/CORPORATE/PDF/white%20paper/video-experience-based-bearer-network-technical-whitepaper.pdf>. — 2016.
14. *Huawei*. U-vMOS, Benchmarking Video Experience. — <http://www.huawei.com/minisite/ubbf2015/ppt/PPT-111-vMOS-Video%20Experience%20Standard%20-%20Ken%20Wang%20-%2020150908%20V1.pdf>. — 2016.
15. *Huawei*. U-vMOS, Benchmarking Video Experience. — <http://developer.huawei.com/ru/ict/Products/TelecoSoftware/VideoExperienceCapability/U-vMOS/SDK>. — 2016.
16. *Пастушок, И.А.* Эвристический алгоритм выбора подоптимальных значений характеристик видео потока и параметров алгоритма распределения ресурсов радиоканала / И.А. Пастушок, А.В. Борисовская // Сборник докладов Научной сессии ГУАП. — 2015. — С. 209–219.
17. Quality-of-Experience driven Adaptive HTTP Media Delivery / A. El Essaili, D. Schroeder, D. Staehle et al. // *IEEE ICC 2013 - Communication QoS, Reliability and Modeling Symposium*. — 2013. — Pp. 2480–2485.
18. *Chen, Yan*. QoS Requirements of Network Applications on the Internet / Yan Chen, Toni Farley, Nong Ye // *Inf. Knowl. Syst. Manag.* — 2004. — January. — Vol. 4, no. 1. — Pp. 55–76.
19. QoS Requirements For Multimedia Services / Rafael Asorey CACHEDA, Daniel Castro García, Antonio Cuevas et al. // *Resource Management in Satellite Networks*:

- Optimization and Cross-Layer Design / Ed. by Giovanni Giambene. — Boston, MA: Springer US, 2007. — Pp. 67–94.
20. ITU-T. — ITU-T Y.1540 (11/2007), Internet protocol data communication service – IP packet transfer and availability performance parameters, 2007. — November.
 21. Sesia, Stefania. LTE - the UMTS long term evolution : from theory to practice / Stefania Sesia, Issam Toufik, Matthew Baker. — Chichester: Wiley, 2009. <http://opac.inria.fr/record=b1130916>.
 22. Vetro, A. Video transcoding architectures and techniques: an overview / A. Vetro, C. Christopoulos, Huifang Sun // *IEEE Signal Processing Magazine*. — 2003. — Mar. — Vol. 20, no. 2. — Pp. 18–29.
 23. Xin, Jun. Digital Video Transcoding / Jun Xin, Chia-Wen Lin, Ming-Ting Sun // *Proceedings of the IEEE*. — 2005. — Jan. — Vol. 93, no. 1. — Pp. 84–97.
 24. Scherr, A.A. An analysis of time-shared computer systems / A.A. Scherr. — MIT Press, 1967.
 25. Adiri, I. A time sharing queue with a finite number of customers / I. Adiri, B. Avi-Itzhak // *Journal of the Association for Computing Machinery*. — 1969. — Vol. 16. — Pp. 315–323.
 26. Jaiswal, N. Optimal operating policies for the finite-source queueing process / N. Jaiswal, P. Simha // *Oper. Res.* 20. — 1972. — Vol. 20. — Pp. 698–707.
 27. Veran, M. Modelling and performance evaluation methodology / M. Veran. — Springer, 1984.
 28. Pastushok, I. Lower bound and optimal scheduling for mean user rebuffering percentage of HTTP progressive download traffic in cellular networks / I. Pastushok, A. Turlikov // 2016 XV International Symposium Problems of Redundancy in Information and Control Systems (REDUNDANCY). — 2016. — Sept. — Pp. 105–111.
 29. Пастушок, И.А. Решение обобщенной задачи о непрерывном рюкзаке / И.А. Пастушок // Сборник докладов Научной сессии ГУАП. — 2017. — С. SET ME!!!

30. Video capacity and QoE enhancements over LTE / S. Singh, O. Oyman, A. Papatthanassiou et al. // 2012 IEEE International Conference on Communications (ICC). — 2012. — June. — Pp. 7071–7076.
31. Introduction to Algorithms, Third Edition / Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein. — 3rd edition. — The MIT Press, 2009.
32. *Boyd, Stephen*. Convex Optimization / Stephen Boyd, Lieven Vandenberghe. — Cambridge University Press, 2009.
33. *Моисеев, Н.Н.* Методы оптимизации / Н.Н. Моисеев, Ю.П. Иванова, Е.М. Столярова. — Наука, 1978.
34. IEEE. — IEEE Std 802.16-2012, IEEE Standard for Air Interface for Broadband Wireless Access Systems, 2012. — June.
35. 3GPP. — 3GPP TS 36.101 v14.3.0, Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception, 2017. — March.
36. *MathWorks*. LTE System Toolbox. — <https://www.mathworks.com/products/lte-system.html>. — 2015.
37. *OPNET*. OPNET LTE SIMULATION. — <http://opnetprojects.com/opnet-lte-simulation>. — 2015.
38. *NS-3*. LTE Module. — <https://www.nsnam.org/docs/models/html/lte.html>. — 2017.
39. *ElSawy, Hesham*. Stochastic Geometry for Modeling, Analysis, and Design of Multi-Tier and Cognitive Cellular Wireless Networks: A Survey / Hesham ElSawy, Ekram Hossain, Martin Haenggi // *IEEE Communnications Survey & Tutorials*. — 2013. — Vol. 15. — Pp. 996–1019.
40. *Bakin, E*. Lower bound of mean user waiting/watching time ratio in wireless video streaming networks / E. Bakin, G. Evseev, A. Turlikov // 2014 IEEE Global Communications Conference. — 2014. — Pp. 1170–1175.

41. *Zheng, Xiao Jin*. Convex relaxations for nonconvex quadratically constrained quadratic programming: matrix cone decomposition and polyhedral approximation / Xiao Jin Zheng, Xiao Ling Sun, Duan Li // *Mathematical Programming*. — 2011. — Vol. 129, no. 2. — Pp. 301–329. <http://dx.doi.org/10.1007/s10107-011-0466-y>.
42. *Oyman, Ozgur*. Quality of Experience for HTTP Adaptive Streaming Services / Ozgur Oyman, Sarabjot Singh // *IEEE Communication Magazine*. — 2012. — April. — Vol. 4. — Pp. 20–27.
43. *Bui, N*. Anticipatory quality-resource allocation for multi-user mobile video streaming / N. Bui, S. Valentin, J. Widmer // 2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). — 2015. — April. — Pp. 245–250.
44. Quality of experience-driven adaptation scheme for video applications over wireless networks / A. Khan, L. Sun, E. Jammeh, E. Ifeachor // *IET Communications*. — 2010. — July. — Vol. 4, no. 11. — Pp. 1337–1347.
45. Mobile web QoE study for smartphones / J. Hosek, M. Ries, P. Vajsar et al. // 2013 IEEE Globecom Workshops (GC Wkshps). — 2013. — Dec. — Pp. 1157–1161.
46. Predicting user QoE satisfaction in current mobile networks / J. Hosek, P. Vajsar, L. Nagy et al. // 2014 IEEE International Conference on Communications (ICC). — 2014. — June. — Pp. 1088–1093.
47. *Chen, Yu-Chieh*. A Multi-level QoE Framework for Smartphone Video Streaming Applications / Yu-Chieh Chen, Jen-Wei Chang, Hung-Yu Wei // The 6th IEEE International Workshop on Management of Emerging Networks and Services. — 2014.

СПИСОК РИСУНКОВ

1.1	Обобщенная структура систем передачи видеоданных	12
1.2	Модель поведения пользователя при просмотре видео	17
1.3	Машина состояний видеоплеера	18
1.4	Логика работы неадаптивного видеоплеера	20
1.5	Логика работы адаптивного видеоплеера	22
1.6	Логика адаптации репрезентации опорного видеоплеера Dash.js . .	24
1.7	Схема вычисления оценки качества передачи видео в методологии ITU MOS	30
1.8	Влияние объективных характеристик на критерий качества восприятия U-vMOS	40
2.1	Структура современных беспроводных централизованных систем .	45
2.2	Временная диаграмма поведения пользователя при просмотре видео	47
2.3	Структура беспроводного канала	50
2.4	Функциональная структура базовой станции и пользовательского устройства	53
2.5	Модель производящего устройства	58
2.6	Структура модели передачи видеоданных в централизованных беспроводных сетях	59
2.7	Система передачи видеоданных как замкнутая система массового обслуживания	62
3.1	Вид решений X и Y	77
3.2	Вид целевой функции промежуточной оптимизационной задачи . .	80
3.3	Возможные значения w_s	80
3.4	Вид решений X , Y и Y^*	84
3.5	Концептуальная схема предлагаемого алгоритма планирования . .	91
3.6	Сравнение скорости сходимости методов оценки среднего значения	94
3.7	Логика алгоритма совместного планирования распределения ресурсов для неадаптивных видеопотоков	99
3.8	Сценарий моделирования	102

3.9	Сравнение предложенного алгоритма планирования с известными планировщиками и найденной нижней границей	103
4.1	Графическое представления алгоритма численного нахождения решения	118
4.2	Производительность известных алгоритмов планирования при передаче адаптивного видео в сравнении с нижней границей	123
4.3	Нижняя граница для критерия Q при различном числе абонентов в соте	124
4.4	Максимальная емкость соты для различных значений критерия качества Q	125
В.1	Графическое представление зависимостей подразделов в диссертационном исследовании	143

СПИСОК ТАБЛИЦ

1.1	Рекомендованные настройки репрезентации сервисом YouTube . .	15
1.2	Основные параметры адаптивного видеоплеера Dash.js	27
1.3	Фактор качества видеоданных (sQuality) в методологии U-vMOS .	32
1.4	Фактор ожидания начала воспроизведения (sInteraction) в методологии U-vMOS	33
1.5	Фактор качества воспроизведения (sView) в методологии U-vMOS	33
1.6	Коэффициенты для расчета значение MOS в методологии U-vMOS	33
1.7	Влияние объективных показателей качества обслуживания при передачи видео на критерий качества восприятия U-vMOS	39
3.1	Решение промежуточной оптимизационной задачи	83
3.2	Сравнение сложности методов решения обобщенной задачи о непрерывном рюкзаке	86
A.1	Аппроксимации функции MOS для передачи видеоданных	137
B.1	Параметры моделирования	140

ПРИЛОЖЕНИЕ А

АППРОКСИМАЦИИ ФУНКЦИИ MOS ДЛЯ ПЕРЕДАЧИ ВИДЕОДАНЫХ

Таблица А.1

Аппроксимации функции MOS для передачи видеоданных

Формула	Описание	Источ- ник
Процент замирания воспроизведения потока (Stalling, Rebuffering Percentage)		
$g_i = \lim_{t \rightarrow \infty} \frac{rb_i(t)}{rb_i(t) + d_i(t)}$ $Minimize : \sum_i g_i$	g_i – процент времени, когда пользователь наблюдал прерывание воспроизведения потока; $rb_i(t)$ – длительность замираний воспроизведения потока за время t ; $d_i(t)$ – длительность просмотренной последовательности за время t .	[28; 42]
$q_i = \lim_{t \rightarrow \infty} \frac{w_i(t)}{d_i(t)}$ $Minimize : \sum_i q_i$	$w_i(t)$ – длительность ожидания в течении времени t , учитывающее начальные буферизации и прерывания воспроизведения потока.	[40]
Взвешенные суммы характеристик воспроизведения		
$MOS_i = f_{RATE}(R_i)$ $Maximize : \sum_i MOS_i$	$R_i(t)$ – битовая скорость просматриваемого потока.	[17]

$f_i = k_1 \ln R_i - k_2 \frac{rb_i(t)}{d_i(t)}$ $Maximize : \sum_i f_i$	$R_i(t)$ – битовая скорость просматриваемого потока; k_1, k_2 – весовые коэффициенты приоритета функции от битовой скорости и замираний воспроизведения соответственно, причем $k_1 \ll k_2$.	[16]
$Minimize : W\lambda - \theta$	λ – средний процент времени прерывания воспроизведения по всем пользователям; θ – среднее значение скорости битовой репрезентации по всем пользователям; W – весовой коэффициент приоритета прерывания воспроизведения, сравним со значением θ .	[43]
$QoE_i[t] = W_1$ $\log(R_i[t] - W_2(Buf_i[t] - Buf_{thr})^2 - W_3(R_i[t] - R_i[t-1])^2)$ $Maximize : \sum_i QoE_i[t]$	W_1, W_2, W_3 – весовые коэффициенты; $R_i[t]$ – битовая скорость в момент времени t ; $Buf_i[t]$ – уровень буфера в момент времени t ; Buf_{thr} – пороговое значение уровня буфера.	[8]
$MOS_i = \frac{a_1 + a_2 Fr R_i + a_3 \log(R_i^{TCP})}{1 + a_4 PL_i + a_5 (PL_i)^2}$ $Maximize : \sum_i MOS_i$	a_j – коэффициенты, зависящие от характеристик просматриваемого контента; $Fr R_i$ – частота кадров потока; R_i^{TCP} – скорость получения информации; PL_i – вероятность потери пакета.	[44]
$MOS_i = \frac{b-a}{1+c_0(R_i^{TCP})^{-c_1}(c_2)^{SUD_i}} + a$ $Maximize : \sum_i MOS_i$	a, b – нормирующие константы; c_j – константы, подобранные в ходе эксперимента; SUD_i – длительность ожидания начала воспроизведения.	[45; 46]

$\begin{cases} MOS_{Res}(Resl_i) = 1.475 \log(Resl_i) - 6.15 \\ MOS_{Reb}(RebRate_i) = 0.738 e^{-RebRate_i} \\ MOS_{SUD}(SUD_i) = -0.02 SUD_i + 2.53 \end{cases}$ $MOS_i = 0.174 MOS_{Res}(Resl_i) MOS_{Reb}(RebRate)_i \cdot MOS_{SUD}(SUD_i)$ $Maximize : \sum_i MOS_i$	<p style="text-align: center;">$Resl_i$ – разрешение репрезентации; $RebRate_i$ – отношение времени просмотра к длительности опустошения буфера.</p>	[47]
---	---	------

ПРИЛОЖЕНИЕ Б

ПАРАМЕТРЫ МОДЕЛИРОВАНИЯ

Таблица Б.1

Параметры моделирования

Название параметра	Значение
Общие положения	
Максимальная удаленность пользователей от базовой станции X_{max}	700 м
Типы трафика	Только видео
Длительность моделирования	360000 с
Длительность интервала планирования	1 мс
Беспроводной канал	
Модель затухания распространения сигнала	Окамура-Хата
Окружение	Городская застройка
Тип города	Большой город
Мощность передатчика базовой станции	44 дБм
Мощность передатчика пользовательского устройства	23 дБм
Коэффициент усиления антенны базовой станции и пользовательского устройства	10 дБ
Несущая частота	2 ГГц
Ширина полосы передачи	10 МГц
Модель ошибки при передаче данных	Отсутствует
Плоские замирания	
Модель	Extended Pedestrian A (EPA)
Движение окружения	3 км/ч
Длительность генерируемой последовательности	10 с
Несущая частота	2 ГГц
TCP протокол	

Стандарт	TCP NewReno
Размер фрагментации пакета	1440 Б
Минимальное значения таймаута	1 с
Начальный размер окна	4320 Б
Ограничение начального роста окна быстрого старта	Неограничен
Размер буфера на приемной и передающей стороне	6 МБ
Алгоритм планирования для неадаптивных видеопотоков	
Интервал усреднения оценки скорости передачи информации (\bar{w}_S)	1000 мс
Интервал усреднения оценки максимальной пропускной способности канала (\bar{w}_C)	1000 мс
Минимальная гарантированная скорость получения информации (S^{min})	100 Кб/с
Видео контент	
Длительность видео	300 с
Битовые скорости потока неадаптивных видеопоследовательностей	1 Мбит/с
Битовые скорости потока адаптивных видеопоследовательностей	[1, 4.5] Мбит/с
Видеоплеер	
Длительность сегмента видеоданных	2 с
Размер начальной буферизации	4 с
Максимальный размер буфера	10 с
Адаптация видеопотока	Соответствует параметрам, представленным в таблице 1.2
Поведение пользователя	
Начальная задержка при заказе видео	Равномерная случайная величина в отрезке [1, 60] с

Пауза между заказами видео	Усеченная экспоненциальная случайная величина в отрезке $[15, 45]$ с со средним значением 30 с
----------------------------	--

ПРИЛОЖЕНИЕ В

ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ЗАВИСИМОСТЕЙ РАЗДЕЛОВ

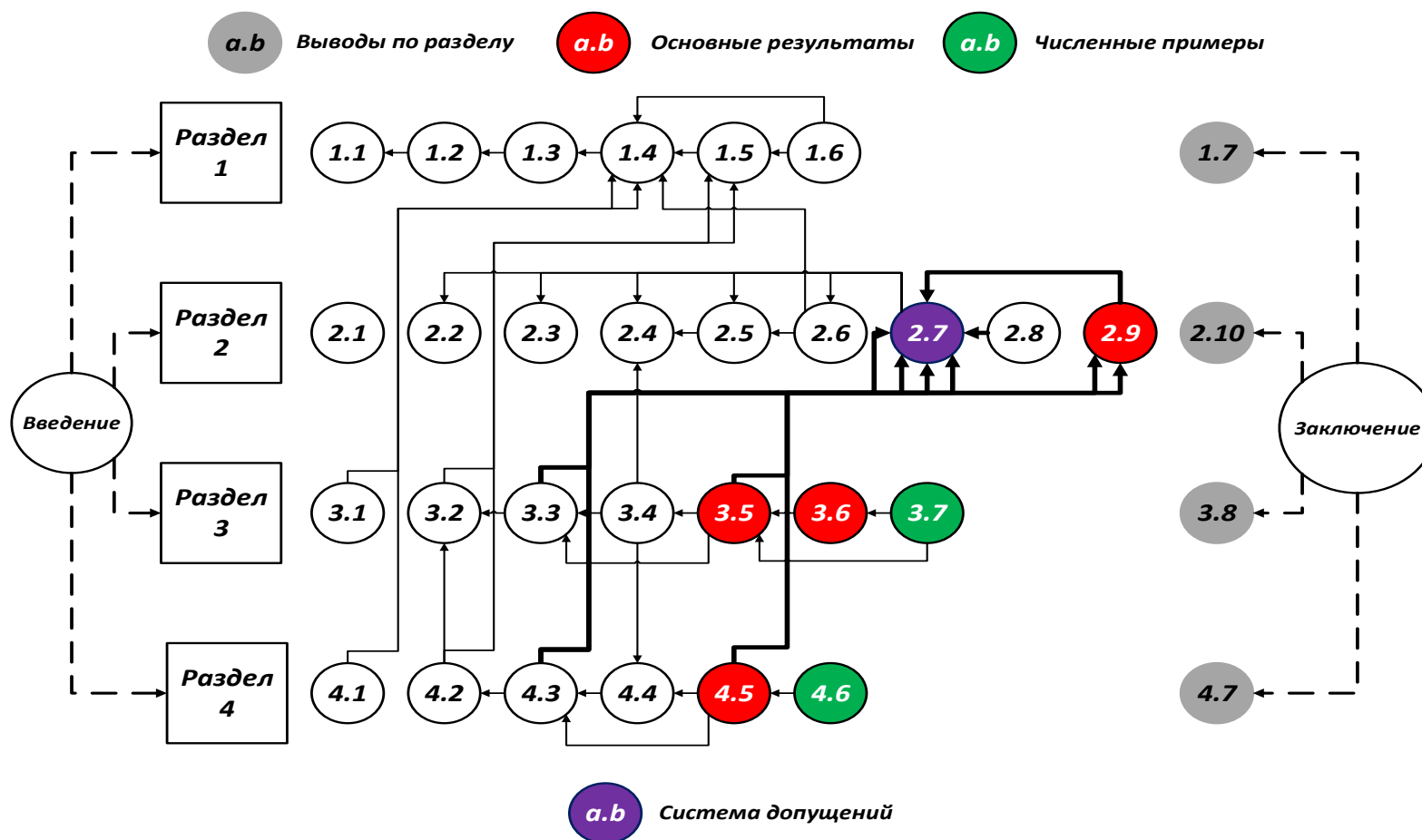


Рисунок В.1 — Графическое представление зависимостей подразделов в диссертационном исследовании