

Построение модели управления пропускной способностью телекоммуникационных сетей

С.А. Васильев, С.К. Канзитдинов, И.В. Левичев, Акоста Д. Перес

Аннотация — Внедрение технологий 5G/6G требует особого внимания разработчиков новых продуктов в области беспроводной связи, поскольку эти технологии смогут обеспечить более высокие скорости передачи данных, меньшую задержку, а также большую энергоэффективность по сравнению с используемыми в настоящее время технологиями 4G. Разработчики неизбежно сталкиваются с рядом технических проблем при проектировании архитектуры 5G/6G, которая должна справляться с более сложной многопользовательской средой и использованием каналов на более высоких частотах. Современные исследования управления нагрузкой в сетях 5G/6G позволяют понять проблемы, с которыми надо будет решать при внедрении технологии 5G/6G.

В нашей статье мы предлагаем экономико-математическую модель управления пропускной способностью телекоммуникационных сетей, принимая во внимание, что наиболее серьезным в телекоммуникационной отрасли на сегодняшний день является вопрос о взаимоотношениях между поставщиками сетевых услуг и контент-провайдерами. На основе построенной нами модели проводится анализ вопроса, следует ли поставщиками сетевых услуг предлагать контент-провайдерами различные классы обслуживания по передаче данных конечным клиентам, которые задают уровни качества обслуживания QoS. Было показано, что сегментирование по уровням качества обслуживания QoS может быть более эффективным в краткосрочной перспективе, поскольку оно лучше распределяет существующую пропускную способность телекоммуникационных сетей, а в долгосрочной перспективе такое сегментирование может обеспечить более высокие инвестиционные стимулы по причине предполагаемого роста спроса на услуги сетей следующих

поколений 5G/6G из-за появления новых поставщиков контента, сервисы которых могут быть чувствительны к перегрузкам сетей.

Ключевые слова— управление нагрузкой в сетях, сети 5G/6G, ценообразование в телекоммуникационной отрасли, управление качеством обслуживания (QoS).

I. ВВЕДЕНИЕ

Телекоммуникационные сети играют жизненно важную роль в преодолении пространственного разделения людей. Сочетание роста численности населения, урбанизации и устойчивого экономического роста привело к серьезным проблемам, которые связаны с серьезной перегруженностью инфраструктурных объектов связи во многих странах мира. Учитывая серьезность этой проблемы, разработчики телекоммуникационного оборудования предприняли ряд решений, которые заключаются в создании новых мобильных сетей 5G/6G с высокой пропускной способностью. В последнее время проблема управления пропускной способностью сетей считается крайне серьезной в телекоммуникационной отрасли [1]-[5].

В нашей работе мы предлагаем экономико-математическую модель управления пропускной способностью телекоммуникационных сетей, принимая во внимание, что наиболее серьезным в телекоммуникационной отрасли на сегодняшний день является вопрос о взаимоотношениях между поставщиками сетевых услуг и контент-провайдерами [6]-[12]. На основе построенной нами модели проводится анализ вопроса, следует ли поставщиками сетевых услуг предлагать контент-провайдерами различные классы обслуживания по передаче данных конечным клиентам, которые задают уровни качества обслуживания QoS. Было показано, что сегментирование по уровням качества обслуживания QoS может быть более эффективным в краткосрочной перспективе, поскольку оно лучше распределяет существующую пропускную способность телекоммуникационных сетей, а в долгосрочной перспективе такое сегментирование может обеспечить более высокие инвестиционные стимулы по причине предполагаемого роста спроса на услуги сетей следующих поколений 5G/6G из-за появления новых поставщиков контента, сервисы которых могут быть чувствительны к перегрузкам сетей.

Статья получена 14 ноября 2022.

Васильев Сергей Анатольевич, кандидат физико-математических наук, Российский университет дружбы народов, доцент кафедры прикладной информатики и теории вероятностей, ORCID: 0000-0003-1562-0256 (vasilyev-sa@rudn.ru).

Канзитдинов Шахмурад Канзитдинович, Российский университет дружбы народов, аспирант кафедры прикладной информатики и теории вероятностей, ORCID: 0000-0002-3972-7739 (shahkanzitdinov@mail.ru).

Левичев Игорь Владимирович, Российский университет дружбы народов, аспирант кафедры прикладной информатики и теории вероятностей, ORCID: 0000-0002-8566-9461 (levichev-iv.rudn@mail.ru).

Перес Акоста Даниэль, Российский университет дружбы народов, аспирант кафедры прикладной информатики и теории вероятностей, ORCID: 0000-0003-4937-1796 (peresacostadaniel@mail.ru).

Публикация выполнена при поддержке Программы стратегического академического лидерства РУДН (получатель С.А.Васильев, разработка математической модели, реализация численной схемы, проведение численного анализа).

А. Допущения модели

Наиболее серьезным в телекоммуникационной отрасли на сегодняшний день является вопрос о взаимоотношениях между поставщиками сетевых услуг (Network Service Provider, NSP) и контент-провайдером (Content Provider, CP). В данной работе предлагается экономико-математическая модель управления пропускной способностью телекоммуникационных сетей и на основе этой модели рассматривается вопрос о том, что следует ли NSP предлагать CP различные классы обслуживания по передаче данных конечным клиентам, которые задают уровни качества обслуживания QoS (Quality of Service, QoS). В соответствии с принципом сетевого нейтралитета (Network Neutrality, NN), по которому NSP не отдадут предпочтения одному целевому предназначению перед другим или одним классам приложений (например, World Wide Web) перед другими (например, онлайн-игры, IP-телефония, видеоконтент и др.), даже если распределение QoS по уровням предлагается на недискриминационной основе. Предполагается, что применение концепции NN может обеспечить равные условия для конкуренции между CP и, таким образом, приведет к большему разнообразию контента. При заданной пропускной способности ускорение приоритетного трафика может привести к замедлению остального трафика. Таким образом, те CP, которые не готовы платить за приоритетный трафик, оказываются в невыгодном положении. Эта проблема вызывает беспокойство в части перспективности применения концепции NN. В долгосрочной перспективе инвестиции в инфраструктуру телекоммуникационных сетей новых поколений 5G/6G, вероятно, будут выше при применении концепции NN, поскольку NSP вынуждены будут обеспечивать достаточную пропускную способность, чтобы CP имели возможность внедрять новый контент и поддерживать качество обслуживания потребителей на приемлемом уровне.

В работе рассматривается модель телекоммуникаций отрасли с рынками: рынок, на котором присутствуют NSP и CP, а также рынок, на котором присутствуют CP и клиенты (Network Content Clients, NCC), которые являются потребителями контента, предлагаемого CP. При этом предполагается, что каждый из участников рынка ценит растущее присутствие другой стороны в отрасли и не предпочитает перегрузок в сети - повышенного сетевого трафика (Network Congestion, NC), а также, что NSP может использовать монопольное доминирование на рынке (например, из-за отсутствия у клиентов альтернативных NSP или по причине очень высоких затрат на подключение к конкурентам). Таким образом, единственный способ для CP связаться с NCC - это воспользоваться сетями NSP.

Мы будем считать, что клиенты контент-сервисов, предоставляемых CP, пользуются услугами разных NSP, причем каждый NSP может выступать в роли локального монополиста, но в дальнейшем будет полезно исследовать вопрос о взаимосвязи между CP и одним NSP, особенно если этот NSP является крупным и может использовать доминирующее положение на рынке.

Также предполагается, что CP производят оплату NSP за доступ в сеть, когда имеется факт использования сети NSP для доставки контента NCC.

В. Поставщики контента

Мы предлагаем модель, в которой имеется бесконечно много поставщиков контента, где множество CP представляет собой континуум. Предполагается, что какую бы услугу ни предлагали CP, они предоставляют ее для NCC платно посредством оформления подписки, а также получают выручку косвенно через онлайн-рекламу. В предлагаемой модели выручка CP от контента и рекламы будет зависеть от среднего объема трафика, который создают NCC в процессе пользования контентом, т.е. выручки от контента, а также выручки от рекламы за клик и индивидуального показателя посещаемости NCC контент-ресурсов, который определяется путем мониторинга перегрузок NC, порождаемого клиентами, когда те пользуются контентом CP.

В рамках рассматриваемой модели делается следующее предположение.

Предположение 1. Каждый CP получает одинаковый средний сетевой трафик от каждого клиента в определенный момент времени t , для которого введем обозначение $\lambda(t)$, причем $\lambda(t)$ не зависит от бизнес-модели CP и, следовательно, от чувствительности CP к перегрузкам NC.

Можно рассматривать величину $\lambda(t)$ в качестве показателя активности клиентов (количество просмотренного контента, количество "кликов", которые клиент совершает на веб-сайте каждого CP, и так далее).

Также предполагается, что клиенты равномерно распределяют свою сетевую активность между доступными CP. Предполагается, что $\lambda(t)$ не является постоянной величиной в долгосрочной перспективе, и, таким образом, по мере увеличения на рынке числа CP активность NCC будет возрастать. Это должно привести к усилению конкурентной борьбы среди CP, т.е. по мере того, как на рынок выходит все больше CP, доход каждого CP снижается. В конечном итоге, руководство компаний CP должно прийти к выводу, только часть сетевой активности клиентов $\lambda(t)$ может обеспечивать выручку от подписок на контент и от рекламы. Этот показатель известен как показатель кликабельности (Click-Through Rate, CTR). Мы предполагаем, что скорость перехода клиентов от одного контента к другому будет уменьшаться по мере перегрузок NC.

Более того, предполагается, что бизнес-модель каждого CP обладает чувствительностью к тому, в какой степени перегрузка NC влияет на показатель CTR. Нужно заметить, что показатель CTR зависит от среднего времени ожидания $w(t)$ клиента в процессе работы с контентом.

Клиенты, которые заходят на сайт CP, будут довольны сервисом даже при высокой загрузке сети и с большей вероятностью нажмут на рекламные объявления или подпишутся на контент. Напротив, потребители Web-сервиса с высокой чувствительностью к перегрузке (например, потокового ТВ) могут по-

прежнему заходить на Web-сайт СР, но при наличии перегрузки сети они будут менее удовлетворены сервисом СР и, следовательно, с меньшей вероятностью будут просматривать контент и рекламу.

Эту индивидуальную чувствительность к перегрузке сети обозначим функцией $\theta(t)$, и предположим, что соответствующая скорость перехода по сайтам СР равна $(1 - \theta(t)w(t))$, где $w(t)$ обозначает воспринимаемый СР к среднему уровню перегрузки в сети.

Как было отмечено выше, существует континуальное множество контент-провайдеров СР с единичной нормировкой (0 – первая компания в ряду СР, 1 – последняя компания в ряду СР) и функцией распределения $F(\theta) : [0,1] \rightarrow [0,1]$. Пусть r – средняя выручка СР за одну подписку на контент или за один клик клиента по рекламе, которая будет зависеть от количества активных СР на рынке. В этом случае выручка каждого СР при условии реализации принципа NN в отрасли равна:

$$\Gamma_N(\theta(t)) = \begin{cases} (1 - \theta(t)w_N(t))\lambda(t)\bar{\eta}r & \text{если активен,} \\ 0 & \text{иначе,} \end{cases} \quad (1)$$

где функция $\bar{\eta}(t)$ обозначает рыночную долю интернет-клиентов NCC в равновесии. При NN все СР фиксируют одинаковый уровень сетевой перегрузки $w_N(t)$. Однако в режиме распределения QoS по уровням СР могут выбрать класс передачи с приоритетом $w_{Q_1} < w_N$ по цене p за клик. Введем функцию $\bar{\theta}(t)$, которая будет характеризовать чувствительностью к перегрузке тех СР, которые проявляют безразличие между проявлением активности на рынке и отсутствием активности на рынке. Таким образом, функция $F(\bar{\theta})$ отражает количество всех активных СР на рынке (разнообразие контента на рынке).

С другой стороны, СР, которые могут принадлежать к приоритетному классу обслуживания сетевого трафика, будут иметь более высокий уровень загруженности сети $w_{Q_2} > w_N$. В этом случае выручка каждого СР

$$\Gamma_N(\theta(t)) = \begin{cases} (1 - \theta(t)w_{Q_2}(t))\lambda\bar{\eta}r, & (A) \\ (1 - \theta(t)w_{Q_1}(t))\lambda\bar{\eta}r - \lambda\bar{\eta}p, & (B) \\ 0, & (C) \end{cases} \quad (2)$$

где применяется выражение (А), если СР выбирает приоритетный класс обслуживания сетевого трафика (А-класс), где применяется выражение (Б), если СР выбирает класс «лучшее из возможного» (В-класс), где класс «лучшее из возможного» – best-effort class (в контексте качества обслуживания это один из негарантированных классов качества, в том смысле, что качество не гарантируется, но поддерживается лучшим из того, что осталось) и выражение (В) применяется в остальных случаях.

Множество СР, которые безразличны в выборе между классами А и В в режиме многоуровневого QoS, будем обозначать функцией $\tilde{\theta}(t)$, тогда доля СР,

выбирающих класс А в режиме многоуровневого QoS, определяется выражением $\beta = 1 - F(\tilde{\theta}) / F(\bar{\theta})$.

Мы предполагаем наличие конкурентного рекламного рынка, который косвенно стимулирует конкуренцию между СР и уровень валовой выручки от рекламы у СР зависит от количества активных СР $r(F(\bar{\theta}))$, где $\partial r / \partial F < 0$.

С. Клиенты

Предполагается, что клиенты контент-сервисов ценят возможность подключения к Интернету, а также наличие большого количества СР на рынке. В частности, мы предполагаем, что возможность пользоваться сетевыми сервисами обеспечивает клиентам базовую полезность $U = b > 0$, но при этом каждый дополнительный СР добавляет предельную полезность в размере $v > 0$ к базовой полезности клиента. Но с другой стороны, перегруженность сети снижает полезность от использования услуг СР для клиентов.

Мы предполагаем далее, что полезность клиентов от использования сети определяется только средним уровнем ее загруженности, которая задается выражением $\hat{w}_Q = \beta w_{Q_1} + (1 - \beta)w_{Q_2}$ или выражением $\hat{w}_Q = \hat{w}_N = w_N$, которое имеет место при $w_Q = w_N$ в краткосрочной перспективе, когда $\mu_Q = \mu_N$, $\bar{\theta}_Q = \bar{\theta}_N$. В долгосрочной перспективе будет удобно проводить различие между \hat{w}_Q и \hat{w}_N для обозначения среднего времени ожидания в каждом режиме соответственно.

Так как клиенты чувствительны к перегрузкам сети, то можно было бы оценить уровень перегрузки сети таким образом:

$$\hat{w}_N(t) = \int_0^{\bar{\theta}} w_N(t)\theta(t)F(\theta)d\theta,$$

$$\hat{w}_N(t) = \int_0^{\bar{\theta}} w_{Q_2}(t)\theta(t)F(\theta)d\theta + \int_{\bar{\theta}}^{\bar{\theta}} w_{Q_1}(t)\theta(t)F(\theta)d\theta.$$

Хотя это оценка более строгая, она качественно не изменит наш анализ, а покажет преимущества режима многоуровневого QoS, когда это имеет место. Причина в том, что режим многоуровневого распределения QoS более эффективно распределяет перегрузку сети таким образом, чтобы $\hat{w}_Q \leq \hat{w}_N$.

Таким образом, можно предположить, что функция полезности клиента определяется таким образом:

$$U = \begin{cases} b + v\bar{\theta} - \iota\hat{w} - a & \text{если в сети,} \\ 0 & \text{в противном случае,} \end{cases} \quad (3)$$

где параметр $\iota > 0$ обозначает предельные альтернативные издержки по времени из-за перегрузки сети и a плату за доступ в сеть, взимаемую провайдером.

Д. Перегрузка в телекоммуникационных сетях

Загруженность телекоммуникационных сетей определяется по среднему времени ожидания клиентов после запроса контента. Используем для описания таких процессов модель теории массового обслуживания $M / M / 1$, чтобы увязать подход в предлагаемой модели

о взаимосвязи между средним временем ожидания клиента, сетевым трафиком и пропускной способностью.

В случае NN сети модель $M/M/1$ показывает, что каждый клиент имеет среднее время ожидания, которое можно выразить с помощью функции $w_N(t) = 1/(\mu(t) - \Lambda(t))$. Показатель $\mu(t)$ представляет среднюю скорость, с которой обрабатываются запросы на обслуживание, которая интерпретируется как общая пропускная способность. Тогда функция $\Lambda(t) = \lambda(t) \bar{\eta}(t) F(\bar{\theta}(t))$ будет определять среднюю скорость, с которой совокупные запросы на получение контента от клиентов поступают в сеть NSP, которая интерпретируется как сетевой трафик. Чтобы система массового обслуживания была стабильной, мы должны предположить, что $\mu > \Lambda$.

В режиме многоуровневого QoS CP предлагается выбор клиентом между классами передачи контента А (приоритетный класс обслуживания сетевого трафика) или В (класс обслуживания сетевого трафика «лучшее из возможного»). В модели $M/M/1$ это приводит к введению дополнительной очереди, которая обрабатывает запрос CP в рамках класса А или В. Однако в каждом классе очередь формируется в порядке «живой очереди» (способ FIFO). Для данного случая модель массового обслуживания $M/M/1$ представляет среднее время ожидания w_{Q_1} в классе А и среднее время

ожидания w_{Q_2} в классе В в виде:

$$w_{Q_1}(t) = 1/(\mu(t) - \beta\Lambda), w_{Q_2}(t) = w_{Q_1}\mu(t)/(\mu(t) - \Lambda).$$

Легко видеть, что соотношение $w_{Q_1} < w_N < w_{Q_2}$ всегда выполняется, если предполагать фиксированную пропускную способность $\mu = \mu_Q = \mu_N$ и $\beta < 1$. Это важная особенность нашей модели, поскольку она формально показывает, что обслуживание некоторых CP с приоритетом (в краткосрочной перспективе) однозначно приведет к ухудшению качества обслуживания для остальных CP в классе В.

Е. Поставщики сервисов по доступу в сеть

Предполагается, что NSP присутствует на рынках доступа в сеть как NCC, так и PC, на которых NSP имеет монопольную власть. NSP взимает плату с NCC за доступ в размере a . При NN режиме работы сети плата за доступ NCC является единственным источником дохода для NSP. В долгосрочной перспективе NSP также устанавливает уровень пропускной способности μ сети. Как упоминалось ранее, NCC и CP не предпочитают перегрузки сети. Уровень перегрузки сети определяется средним временем ожидания $w(t)$ контента NCC, которое, опять же, контролируется NSP посредством выбора пропускной способности сети.

Следовательно, при режиме NN прибыль NSP составляет $\Pi_N(t) = \bar{\eta}(t)a - c(\mu(t))$, где $c(\mu(t))$ обозначает затраты на расширение пропускной способности сети. При многоуровневом режиме QoS у NSP есть дополнительная стратегическая переменная

p - цена, которую NSP взимает с CP за передачу сетевого трафика с приоритетом. Провайдер будет выбирать p таким образом, чтобы максимизировать свою дополнительную прибыль от продажи CP приоритетного доступа. Таким образом, в случае многоуровневого режима QoS функция прибыли провайдера $\Pi_Q(t) = \bar{\eta}(t)a + \beta\Lambda p - c(\mu)$.

II. РАЗНООБРАЗИЕ КОНТЕНТА В КРАТКОСРОЧНОМ ПЕРИОДЕ

А. Режим сетевой нейтральности NN

Мы считаем, что инвестиционные решения NSP в отношении пропускной способности сети были таким, что в краткосрочной перспективе величина $\mu(t)$ может рассматриваться как экзогенная переменная, которая не влияет на максимизацию прибыли NSP.

Рассмотрим два сетевых режима в краткосрочной перспективе, когда пропускная способность сети $\mu(t)$ является экзогенной переменной и одинаковой в обоих режимах.

Во-первых, очевидно, что NSP установит абонентскую плату за доступ клиентов в размере $a = b + v\bar{\theta} - i\hat{w}$, такую, чтобы все клиенты подключались к сети ($\bar{\eta} = 1$). При применении принципа сетевого нейтралитета все CP ожидают одинаковый уровень перегрузки сети w_N . Следовательно, для последнего CP, которому еще выгодно войти на рынок, имеется условие для входа в отрасль: $\bar{\theta}_N = 1/w_N = \mu - \mu F(\bar{\theta})$.

Следовательно, увеличение сетевого трафика $\lambda(t)$ на единицу времени оказывает неблагоприятное влияние на перегрузку сети $\partial w_N / \partial \lambda > 0$ и разнообразие контента $\partial \bar{\theta}_N / \partial \lambda < 0$. Этот вывод очень выжжен при анализе влияния режима сетевой нейтральности на рынок, поскольку служит негативным фактором для NSP, который заставляет предполагать увеличение трафика сети и необходимых дополнительных инвестиций в сетевую инфраструктуру.

В. Режим повышенного качества обслуживания

При реализации режима многоуровневого QoS NSP может уменьшить перегрузку для наиболее чувствительных к перегрузке CP за счет внедрения дифференцированных классов передачи данных.

Выделим три типа CP:

1) CP, бизнес-модель которых относительно нечувствительна к перегрузке сети, эти CP останутся в В-классе и не будут оплачивать абонентскую плату.

2) CP, бизнес-модель которых достаточно чувствительна к перегрузке сети и они выберут А-класс с ценой обслуживания p .

3) CP, бизнес-модель которых чрезвычайно чувствительна к перегрузке сети и они не будут проявлять активность, так как вход на рынок им не выгоден.

Так как функция чувствительностью к перегрузке CP, для которых безразличны состояния 1) и 2)

обозначается $\tilde{\theta}(t)$, а СР, для которых безразличны состояния 2) и 3) обозначается $\bar{\theta}_Q(t)$, то очевидно, что должно выполняться неравенство $0 \leq \tilde{\theta}(t) \leq \bar{\theta}_Q(t)$.

При условии рыночного равновесия последний СР, которому еще выгодно войти на рынок, находится таким образом:

$$\bar{\theta}_Q(t) = \frac{1-p}{w_{Q_1}(t)} = \frac{p-r}{r} \left(\mu(t) - \lambda(t) \left(F(\bar{\theta}_Q) - F(\tilde{\theta}) \right) \right),$$

$$\frac{\partial \bar{\theta}_Q}{\partial p} = -\frac{1}{r} \left(\mu(t) - \lambda(t) \left(F(\bar{\theta}_Q) - F(\tilde{\theta}) \right) \right) +$$

(Эффект первого порядка)

$$+ \lambda(t) \frac{p-r}{r} \left(\frac{\partial F(\tilde{\theta})}{\partial \tilde{\theta}} \frac{\partial \tilde{\theta}}{\partial p} - \frac{\partial F(\bar{\theta}_Q)}{\partial \bar{\theta}_Q} \frac{\partial \bar{\theta}_Q}{\partial p} \right).$$

(Эффект второго порядка)

где легко видеть, что увеличение тарифа p на приоритетную передачу оказывает однозначно негативное влияние первого порядка при увеличении разнообразия контента.

Для целей нашего анализа предполагается использовать функцию плотности для величины θ , которая иллюстрирует эффект неравномерного распределения θ . С этой целью рассмотрим функцию плотности вида:

$$f(\theta) = \alpha + 2\theta(1-\alpha), \alpha \in [0, 2]$$

и пусть F будет функцией распределения от функции f . Заметим, что при $\alpha = 1$ можно получить равномерное распределение $F(\theta) = \theta$. В противном случае, если $\alpha > 1$, то существует относительно большое количество нечувствительных к перегрузкам СР ($F(\theta) > \theta$), а если $\alpha < 1$, то существует относительно большое количество чувствительных к перегрузкам ($F(\theta) < \theta$) СР. Следовательно, изменение коэффициента α приводит к смещению некоторого числа СР с участка, чувствительного ($\alpha > 0,5$) к перегрузкам, на участок ($\alpha < 0,5$), нечувствительных к перегрузкам, и наоборот.

Легко показать, что при равномерном распределении ($\alpha = 1$) эффекты первого и второго порядка для СР точно компенсируются при любом уровне тарифов, так что тариф за приоритетную передачу не влияет на разнообразие контента при распределении по уровням QoS. Таким образом, при равномерном распределении СР применение принципа сетевого нейтралитета и распределение СР по уровням QoS будут в точности обеспечивать одинаковый уровень разнообразия контента, т. е. $\bar{\theta}_Q = \bar{\theta}_N = \mu / (\lambda + 1) = 1 / w_N$.

Следующее предложение позволит провести анализ разнообразия контента.

Предложение 1 (Разнообразие контента). Если чувствительность СР к перегрузке распределена равномерно, распределение QoS по уровням не влияет на разнообразие контента в краткосрочной перспективе: количество активных СР такое же, как и при применении принципа сетевого нейтралитета. В обоих

режимах количество активных СР обратно пропорционально среднему уровню перегрузки в сети. Однако, если количество СР, чувствительных к перегрузке, сравнительно невелико (велико), то разделение QoS на уровни приведет к большему (меньшему) разнообразию контента.

Легко убедиться, что для $\alpha = 1$

$$F(\bar{\theta}_N) = \bar{\theta}_N(t) = \frac{\mu(t)}{\lambda(t)+1}, F(\bar{\theta}_Q) = \bar{\theta}_Q(t) = \frac{\mu(t)}{\lambda(t)+1},$$

$$F(\tilde{\theta}) = \tilde{\theta}(t) = \frac{p}{\lambda(t)(r-p)} \bar{\theta}_Q(t),$$

что доказывает *Предложение 1*. Кроме того, из этого следует, что

$$p = \left(1 + \sqrt{\frac{1}{1+\lambda(t)}} \right) r = \left(1 - \sqrt{\frac{\bar{\theta}_Q(t)}{\mu(t)}} \right) r.$$

Так как в общем случае $\alpha \neq 1$, то функция $\bar{\theta}_Q(t)$ будет зависеть от тарифа p , который не будет превышать максимальный тариф p_{max} , который можно найти из уравнения $\Gamma_Q(\bar{\theta}_Q) = \Gamma_Q(\tilde{\theta})$, решая которое, получим:

$$p_{max} = r \left(1 + \frac{\lambda\alpha + 1 - \sqrt{(\lambda\alpha + 1)^2 + 4\mu\lambda(1-\alpha)}}{2\mu\lambda(1-\alpha)} \right).$$

Предложение 2. Индивидуальная чувствительность к перегрузке сети СР $\theta(t)$, равномерно распределена таким образом, что $F(\theta(t)) = \theta(t)$ в каждый момент времени.

При таких предположениях многоуровневое распределение QoS не приведет ни к большему, ни к меньшему разнообразию контента. Однако в соответствии с режимом многоуровневости QoS NSP может дополнительно извлекать плату из СР путем продажи приоритетного доступа класса А. В краткосрочной перспективе это будет достигнуто за счет максимизации доходов от приоритетных продаж ($\Lambda\beta p$), что достигается при цене

$$p(t) = \left(1 - \sqrt{\frac{\bar{\theta}_Q(t)}{\mu(t)}} \right) r = \left(1 - \frac{w_{Q_1}(t)}{w_Q(t)} \right) r.$$

Это демонстрирует, что интернет-провайдер может извлекать у СР часть валовой выручки r от контента или рекламы в зависимости от снижения уровня перегрузки до приоритетного класса А по сравнению со средним уровнем перегрузки в сети.

Предложение 2 (Предпочтительный режим NSP). NSP всегда будет предпочитать режим многоуровневого QoS, потому что он может получать дополнительную прибыль, продавая СР услугу приоритетной передачи класса А.

Легко видеть, что при $F(\bar{\theta}_N) = F(\bar{\theta}_Q)$ и для $\alpha = 1$ получим, что

$$\Pi_Q - \Pi_N = \Lambda(t)\beta(t)p = \mu(t)r \left(1 + \frac{1}{1+\lambda} - \frac{2}{\sqrt{1+\lambda}} \right) > 0,$$

который всегда больше нуля для любых $\mu, r, \lambda > 0$ в любой момент времени.

III. ДОЛГОСРОЧНЫЕ РЕЗУЛЬТАТЫ ИНВЕСТИРОВАНИЯ В ПРОПУСКНУЮ СПОСОБНОСТЬ СЕТИ И МИНИМАЛЬНЫЕ СТАНДАРТЫ КАЧЕСТВА QoS

A. Инвестиционные стимулы для NSP

В этом разделе будет проведен анализ эффективности долгосрочных инвестиций в пропускную способность сети. В модели пропускная способность представлена средней скоростью обслуживания $\mu(t)$, при которой могут обрабатываться запросы клиентов. Увеличение $\mu(t)$ позволяет NSP обрабатывать больше запросов на обслуживание от CP одновременно.

Формально инвестиционное решение NSP - это отдельный этап принятия стратегических решений, которые определяют развитие технологии передачи данных. Действительно, NSP сначала выбирает уровень пропускной способности $\mu(t)$ сети, а затем устанавливает тариф за доступ клиентам a и приоритетную цену p для CP.

NSP установит оптимальный уровень пропускной способности при значении, когда предельные доходы от расширения пропускной способности, $MR = \partial\Pi / \partial\mu$ будут равны предельным издержкам $MC = \partial c(\mu) / \partial\mu$, т.е. $\partial\Pi / \partial\mu = \partial c(\mu) / \partial\mu$. Следовательно, оптимальный уровень пропускной способности NSP будет выше, если маржинальные доходы от расширения пропускной способности будут выше. В обоих сетевых режимах можно выделить следующие два предельных эффекта расширения пропускной способности, которые будут влиять на доходы провайдера:

- 1) Стимулирование разнообразия: где $v\partial F(\bar{\theta}) / \partial\mu$ - эффект дополнительной предельной выручки NSP от входа новых CP, чувствительных к перегрузке.
- 2) Стимул перегрузки сети: где $-idw / \partial\mu$ - эффект влияние предельной выручки NSP за абонентский доступ, возникающей в результате изменения общего уровня перегрузки в сети.

Кроме того, в предположении равномерного распределения чувствительности CP к перегрузке эти инвестиционные стимулы всегда являются положительными и идентичными в обоих сетевых режимах. Следовательно, различия в инвестициях между двумя режимами могут быть только результатом дополнительного инвестиционного стимула, который NSP имеет только при распределении QoS по уровням:

- 3) Стимул для получения приоритетного дохода: где $\partial(\beta\lambda p) / \partial\mu$ - эффект получения предельной выручки NSP от продажи приоритетного доступа CP.

Следовательно, стимул 3) является определяющим для сравнения инвестиционных стимулов при многоуровневом QoS и в случае применения принципа сетевого нейтралитета.

Таким образом можно сформулировать следующее предложение.

Предложение 3 (Инвестиционные стимулы). Если чувствительность NSP к перегрузке равномерно

распределена, оптимальный уровень пропускной способности сети будет выше при многоуровневом распределении QoS.

Стимулы для инвестирования NSP в пропускную способность сети выше при многоуровневом QoS, если предельная выручка от приоритетных продаж больше нуля, при условии, что функция выручки NSP вогнута, а функция затрат на расширение пропускной способности выпуклая. Чтобы гарантировать, что функция выручки NSP является вогнутой, должно быть выполнено неравенство $\partial^2\Pi_Q / \partial\mu^2 \leq 0$.

Таким образом, условие второго порядка задается выражением:

$$\frac{\partial^2\Pi_Q}{\partial\mu^2} = -\frac{t(1+\lambda)}{\mu^2} + \left(\frac{\partial^2 r(\bar{\theta})}{\partial\bar{\theta}^2} \frac{\bar{\theta}}{2} + \frac{\partial r(\bar{\theta})}{\partial\bar{\theta}} \right) \frac{\partial\bar{\theta}}{\partial\mu} \times \left(1 + \frac{1}{1+\lambda} - \frac{2}{\sqrt{1+\lambda}} \right) < 0.$$

Поскольку неравенство

$$B = 1 + \frac{1}{1+\lambda} - \frac{2}{\sqrt{1+\lambda}} \geq 0$$

Выполняется при любых $\lambda \geq 0$, функция выручки NSP будет вогнутой, если

$$\frac{\partial^2\Pi_Q}{\partial\mu^2} \leq 0 \begin{cases} A \leq 0, & \text{то всегда,} \\ A > 0, & \text{если } \frac{t(1+\lambda)^2}{\mu^3 B} \geq A. \end{cases}$$

B. Минимальные стандарты качества QoS

Необходимо отметить, что в телекоммуникаций отрасли минимальный стандарт качества (Minimum Quality Standards, MQS) может быть подходящим инструментом регулирования рынка телекоммуникаций в этом контексте. MQS будет способствовать повышению более оптимальному распределению ресурсов в условиях конкуренции на рынке телекоммуникаций. Таким образом, требуя при выполнении MQS, чтоб имело место равенство $w_N(\mu_N^*) = w_{Q_2}(\mu_{MQS})$, регулятор рынка неявно определяет новый уровень пропускной способности $\mu_{MQS} > \mu_N^*$. В соответствии с *Предложением 3* порядок соответствующего уровня пропускной способности таков $\mu_Q^{**} > \mu_Q^* > \mu_N^*$. Так как имеет место неравенство $\mu_{MQS} > \mu_N^*$, то таким образом имеется возможность провести различие между тремя возможными состояниями пропускной способности:

1) если $\mu_Q^* \geq \mu_{MQS}$ и MQS не является обязательным условием для выбора пропускной способности NSP, тогда это состояние неэффективно;

2) если $\mu_Q^{**} \geq \mu_{MQS} > \mu_Q^*$ и применяется MQS для повышения уровня пропускной способности сети NSP, тогда возможно достижение потенциально эффективного уровня инвестиций;

3) если $\mu_{MQS} > \mu_Q^{**}$ и применяется политика MQS, то это может привести к чрезмерным инвестициям в сетевую инфраструктуру.

Таким образом, MQS эффективны только в одном из трех случаев.

IV. ЗАКЛЮЧЕНИЕ

Необходимо отметить, что применение принципа сетевого нейтралитета стала темой изучения для многих исследователей, но эффект от такого регулирования телекоммуникационной отрасли до сих пор неясен. В научных статьях часто встречаются противоречивые результаты в отношении последствий применения принципа сетевого нейтралитета и его влияния на разнообразие контента и инвестиций в пропускную способность сетей. Наш анализ был сосредоточен на отношениях между СР и NSP-монополистом и проведено сравнение различных состояний рынка телекоммуникаций при применении принципа сетевого нейтралитета, режима распределения QoS по уровням качества обслуживания, при котором СР могут платить за передачу своих данных с приоритетом на недискриминационной основе.

Было показано, что распределение QoS по уровням, будет более эффективным режимом, если сбалансирована доля компаний СР, чувствительных к перегрузкам, и компаний СР, нечувствительных к перегрузкам. В этом случае NSP вкладывает больше средств в инфраструктуру сетей связи и тем самым позволяет в краткосрочной перспективе вводить новые, чувствительные к перегрузкам точки доступа и поэтому, особенно очень чувствительным к перегрузкам СР, будет лучше при многоуровневом QoS.

Политика регулятора при обеспечении MQS, которая требует от NSP гарантировать соответствующий уровень пропускной способности сети для класса А при многоуровневом распределении QoS, который, по крайней мере, не уступает уровню пропускной способности сети при максимальной нагрузке при сетевом нейтралитете, недостаточна для стимулирования эффективных инвестиций в инфраструктуру сети со стороны NSP. Однако, если у NSP есть стимул снижать качество для класса В, то обеспечение MQS может быть подходящим инструментом для смягчения пагубных последствий, которые эта практика может оказать на разнообразие контента.

Необходимо отметить, что возражения против многоуровневости QoS оправданы, но имеются веские аргументы в пользу многоуровневой сети. Потенциальные опасности многоуровневого режима QoS, такие как ухудшение качества QoS, могут быть преодолены с помощью обязательств NSP по обеспечению минимальных стандартов качества. В условиях конкуренции NSP будут пытаться привлечь клиентов, предлагая им большее разнообразие контента и более низкий средний уровень загрузки сети, чем у их конкурентов, а это повысит их инвестиционные стимулы по развитию пропускной способности сетей.

БЛАГОДАРНОСТИ

Публикация выполнена при поддержке Программы стратегического академического лидерства РУДН (получатель С.А.Васильев, разработка математической модели, реализация численной схемы, проведение численного анализа).

БИБЛИОГРАФИЯ

- [1] J. Chen, Y. Jing, "Multiple bottleneck topology TCP/AQM switching network congestion control with input saturation and prescribed performance," *ISA Transactions*, 2022.
- [2] X. Zhang, N. Huang, L. Sun, X. Zheng, Z. Guo, "Modeling congestion considering sequential coupling applications: A network-cell-based method," *Physica A: Statistical Mechanics and its Applications*, vol. 604, 2022, p.127668.
- [3] J. Zeng, Y. Xiong, F. Liu, J. Ye, J. Tang, "Uncovering the spatiotemporal patterns of traffic congestion from large-scale trajectory data: A complex network approach," *Physica A: Statistical Mechanics and its Applications*, vol. 604, 2022, p. 127871
- [4] S. Khan, A. Hussain, S. Nazir, F. Khan, A. Oad, M. Dahman Alshehri, "Efficient and reliable hybrid deep learning-enabled model for congestion control in 5G/6G networks," *Computer Communications*, vol.182, 2022, p. 31-40.
- [5] S. Han, "Congestion-aware WiFi offload algorithm for 5G heterogeneous wireless networks," *Computer Communications*, vol. 164, 2020, p. 69-76.
- [6] R. Roostaei, Z/ Dabiri, Z. Movahedi, "A game-theoretic joint optimal pricing and resource allocation for Mobile Edge Computing in NOMA-based 5G networks and beyond," *Computer Networks*, vol. 198, 2021, p. 108352.
- [7] I. Nikolic, N. Galli, "Patent pools in 5G: The principles for facilitating pool licensing," *Telecommunications Policy*, vol. 46, No 4, 2022, p. 102287.
- [8] R. Fedrizzi, R. Behraves, C. Costa, F. Granelli, "An adaptive GA-based slice provisioning method for vertical industries in 5G and beyond networks," *Computer Networks*, vol. 218, 2022, p. 109397.
- [9] D. Kim, "A 2020 perspective on "A dynamic model for the evolution of the next generation Internet – Implications for network policies": Towards a balanced perspective on the Internet's role in the 5G and Industry 4.0 era.," *Electronic Commerce Research and Applications*, vol. 41, 2020, p. 100966.
- [10] W. Lehr, F. Queder, J. Haucap, "5G: A new future for Mobile Network Operators, or not?," *Telecommunications Policy*, vol. 45, No 3, 2021, p. 102086.
- [11] J. Kraemer, L. Wiewiorra, "Network Neutrality and Congestion Sensitive Content Providers: Implications for Content Variety, Broadband Investment and Regulation," *Information Systems Research*, vol. 23 No. 4, p. 1303-1321.
- [12] L. A. Sevastianov, S. A. Vasilyev, "Economical analysis of congested networks," *2014 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, 2014, p. 526-533

Васильев Сергей Анатольевич, кандидат физико-математических наук, доцент кафедры прикладной информатики и теории вероятностей, Федеральное государственное автономное образовательное учреждение высшего образования "Российский университет дружбы народов", 117198, г. Москва, ул. Миклухо-Маклая, д. 6, **ORCID:** <https://orcid.org/0000-0003-1562-0256>, vasilyev-sa@rudn.ru

Канзитдинов Шахмурад Канзитдинович, аспирант кафедры прикладной информатики и теории вероятностей, Федеральное государственное автономное образовательное учреждение высшего образования "Российский университет дружбы народов", 117198, г. Москва, ул. Миклухо-Маклая, д. 6, **ORCID:** <https://orcid.org/0000-0002-3972-7739>, shahkazitdinov@mail.ru

Левичев Игорь Владимирович, аспирант кафедры прикладной информатики и теории вероятностей, Федеральное государственное автономное образовательное учреждение высшего образования "Российский университет дружбы народов", 117198, г. Москва, ул. Миклухо-Маклая, д. 6, **ORCID:** <https://orcid.org/0000-0002-8566-9461>, levicev-iv.rudn@mail.ru

Перес Акоста Даниэль, аспирант кафедры прикладной информатики и теории вероятностей, Федеральное государственное автономное образовательное учреждение высшего образования "Российский университет дружбы народов", 117198, г. Москва, ул. Миклухо-Маклая, д. 6, **ORCID:** <https://orcid.org/0000-0003-4937-1796>, peresacostadaniel@mail.ru

Congestion control scheme model of telecommunication networks capacity

S.A. Vasilyev, S.K.Kanzitdinov, I.V. Levichev, D. Peres Acosta

Abstract— The implementation of 5G/6G technologies requires special attention of developers in the field of wireless communication, since these technologies will be able to provide higher data transfer speeds, lower latency, as well as greater energy efficiency compared to currently used 4G technologies. Developers inevitably face a number of technical problems when designing the 5G/6G architecture, which must cope with a more complex multi-user environment and the use of channels at higher frequencies. Modern studies of congestion control scheme models in 5G/6G networks allow us to understand the problems that will need to be solved when implementing 5G/6G technology.

In our article, we propose an economic and mathematical model of congestion control scheme for telecommunications networks, taking into account that the most serious issue in the telecommunications industry today is the relationship between network service providers and content providers. Based on the model we have built, we analyze the question of whether network service providers should offer content providers different classes of data transmission services to end customers who set QoS service quality levels. It has been shown that segmentation by QoS service quality levels can be more effective in the short term, since it better distributes the existing congestion control scheme for telecommunications networks, and in the long term, such segmentation can provide higher investment incentives due to the expected growth in demand for next-generation 5G/6G network services due to the emergence of new content providers., whose services may be sensitive to network congestion.

Keywords— network load management, 5G/6G networks, telecom pricing, quality of service (QoS) management.

ACKNOWLEDGMENT

This paper has been supported by the RUDN University Strategic Academic Leadership Program (recipient S.A. Vasilyev, mathematical model development, simulation model development, numerical analysis).

REFERENCES

- [1] J. Chen, Y. Jing, "Multiple bottleneck topology TCP/AQM switching network congestion control with input saturation and prescribed performance," *ISA Transactions*, 2022.
- [2] X. Zhang, N. Huang, L. Sun, X. Zheng, Z. Guo, "Modeling congestion considering sequential coupling applications: A network-cell-based method," *Physica A: Statistical Mechanics and its Applications*, vol. 604, 2022, p.127668.
- [3] J. Zeng, Y. Xiong, F. Liu, J. Ye, J. Tang, "Uncovering the spatiotemporal patterns of traffic congestion from large-scale trajectory data: A complex network approach," *Physica A: Statistical Mechanics and its Applications*, vol. 604, 2022, p. 127871
- [4] S. Khan, A. Hussain, S. Nazir, F. Khan, A. Oad, M. Dahman Alshehri, "Efficient and reliable hybrid deep learning-enabled model for congestion control in 5G/6G networks," *Computer Communications*, vol.182, 2022, p. 31-40.
- [5] S. Han, "Congestion-aware WiFi offload algorithm for 5G heterogeneous wireless networks," *Computer Communications*, vol. 164, 2020, p. 69-76.
- [6] R. Roostaei, Z/ Dabiri, Z. Movahedi, "A game-theoretic joint optimal pricing and resource allocation for Mobile Edge Computing in NOMA-based 5G networks and beyond," *Computer Networks*, vol. 198, 2021, p. 108352.
- [7] I. Nikolic, N. Galli, "Patent pools in 5G: The principles for facilitating pool licensing," *Telecommunications Policy*, vol. 46, No 4, 2022, p. 102287.
- [8] R. Fedrizzi, R. Behraves, C. Costa, F. Granelli, "An adaptive GA-based slice provisioning method for vertical industries in 5G and beyond networks," *Computer Networks*, vol. 218, 2022, p. 109397.
- [9] D. Kim, "A 2020 perspective on "A dynamic model for the evolution of the next generation Internet – Implications for network policies": Towards a balanced perspective on the Internet's role in the 5G and Industry 4.0 era," *Electronic Commerce Research and Applications*, vol. 41, 2020, p. 100966.
- [10] W. Lehr, F. Queder, J. Haucap, "5G: A new future for Mobile Network Operators, or not?," *Telecommunications Policy*, vol. 45, No 3, 2021, p. 102086.
- [11] J. Kraemer, L. Wiewiorra, "Network Neutrality and Congestion Sensitive Content Providers: Implications for Content Variety, Broadband Investment and Regulation," *Information Systems Research*, vol. 23 No. 4, p. 1303-1321.
- [12] L. A. Sevastianov, S. A. Vasilyev, "Economical analysis of congested networks," *2014 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, 2014, p. 526-533

Vasilyev Sergey, Candidate of Physico-mathematical Sciences, Associate Professor, Department of Applied Probability and Informatics, Peoples' Friendship University of Russia, 117198, Moscow, Miklukho-Maklaya str.6, **ORCID:** <https://orcid.org/0000-0003-1562-0256>, vasilyev-sa@rudn.ru

Kanzitdinov Shakhmurad Kanzitdinovich, Postgraduate Student, Department of Applied Probability and Informatics, Peoples' Friendship University of Russia, 117198, Moscow, Miklukho-Maklaya str.6, **ORCID:** <https://orcid.org/0000-0002-3972-7739>, shahkzanditdinov@mail.ru

Levichev Igor Vladimirovich, Postgraduate Student, Department of Applied Probability and Informatics, Peoples' Friendship University of Russia, 117198, Moscow, Miklukho-Maklaya str.6, **ORCID:** <https://orcid.org/0000-0002-8566-9461>, levichev-iv.rudn@mail.ru

Peres Acosta Daniel, Postgraduate Student, Department of Applied Probability and Informatics, Peoples' Friendship University of Russia, 117198, Moscow, Miklukho-Maklaya str.6, **ORCID:** <https://orcid.org/0000-0003-4937-1796>, peresacostadaniel@mail.ru