

Использование многоуровневых вычислительных систем для обработки одноранговых пользовательских запросов в мобильных сетях

Р.Я. Пирмагомедов^{1, 2*}, А.А. Ахмед²

¹Университет Тампере,
Тампере, 33100, Финляндия

²Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича,
Санкт-Петербург, 193232, Российская Федерация

*Адрес для переписки: rustam.pirmagomedov@tuni.fi

Информация о статье

Поступила в редакцию 26.02.2020

Принята к публикации 22.06.2020

Ссылка для цитирования: Пирмагомедов Р.Я., Ахмед А.А. Использование многоуровневых вычислительных систем для обработки одноранговых пользовательских запросов в мобильных сетях // Труды учебных заведений связи. 2020. Т. 6. № 2. С. 79–86. DOI:10.31854/1813-324X-2020-6-2-79-86

Аннотация: Эволюция приложений Интернета Вещей диктует новые, более жесткие требования к скорости передачи и обработки информации. Для снижения задержки большую популярность набирают периферийные вычисления мультисервисного доступа. Такой подход позволяет обрабатывать данные пользователей ближе к их местоположению. Однако такие решения могут не учитывать специфику приложений, основанных на одноранговых пользовательских запросах (P2P). **Предмет исследования.** В статье исследуется время обработки P2P запросов в мобильных сетях. **Метод.** В качестве основного метода исследования использовалось имитационное моделирование. **Основные результаты.** Результаты работы позволяют сделать вывод о недостаточной эффективности периферийных вычислений мультисервисного доступа в их стандартном архитектурном исполнении. В то же время разработанное решение на основе многоуровневой архитектуры с многоадресной маршрутизацией обеспечивает значительно более низкое время обработки запросов. **Практическая значимость.** Предложенное авторами решение способно обеспечить работоспособность приложений, требующих минимальной сетевой задержки, как например, медицинские приложения, беспилотный транспорт, поиск различных поставщиков услуг с привязкой к географическим координатам.

Ключевые слова: 5G, P2P, одноранговые сервисы, IMT2030.

Введение

В последнее время наблюдается широкий интерес к одноранговым (P2P, от англ. Peer-to-Peer, равный к равному) услугам в мобильных сетях. Такие сервисы рассматривают взаимодействие пользователей в относительной близости, когда один пользователь ищет услуги, а другой готов их предоставить (например, приложения такси). Такие сервисы рассматриваются в качестве ключевого технологического компонента сетей IMT2030 [1], в которых доля самоорганизующихся одноранговых структур будет постепенно увеличиваться по отношению к инфраструктурным решениям.

В настоящее время такие сервисы реализуются при помощи централизованной архитектуры, обрабатывая пользовательские запросы на центральном сервере. Такой подход часто критикуется за высокие задержки и неэффективное использование сетевых ресурсов, поскольку запрос может быть обработан локально без пересылки данных через центральный сервер, что особенно важно для приложений реального времени [2].

Значительное количество недавних исследований было сосредоточено на обеспечении низкой задержки с использованием технологии граничных вычислений (MEC, от англ. Mobile Edge Computing) [3–5]. В основном улучшения были достигнуты за счет разгрузки вычислительных задач от ограни-

ченных пользовательских устройств к более мощным устройствам, расположенным совместно с базовыми станциями (БС) сотовых сетей, и делегирования доли серверных функций этим объектам [6].

Однако в известных публикациях в необходимой степени не были рассмотрены вопросы внедрения MEC для распределенных P2P-сервисов в мобильных сетях. Поэтому в данной работе мы решаем эту актуальную задачу. В частности, в статье рассмотрен сценарий услуги, при которой требуется взаимодействие между пользователями в реальном времени. В рамках рассмотренного сценария решается задача снижения среднего времени обработки P2P запроса.

Архитектура многоуровневой вычислительной системы в сетях мобильной связи пятого поколения

Архитектура сетей связи пятого поколения (5G), стандартизированная в пятнадцатом релизе консорциума 3GPP (*от англ.* 3rd Generation Partnership Project), отличается высокой гибкостью и возможностью предоставления различных услуг в рамках одной сети доступа. В отличие от предыдущих поколений, в 5G обеспечивается динамическое программное управление радиоресурсами во всех частотных диапазонах и явное разделение уровня пользовательских данных и уровня сигнализации. Общий вид архитектуры радиодоступа 5G приведен на рисунок 1.

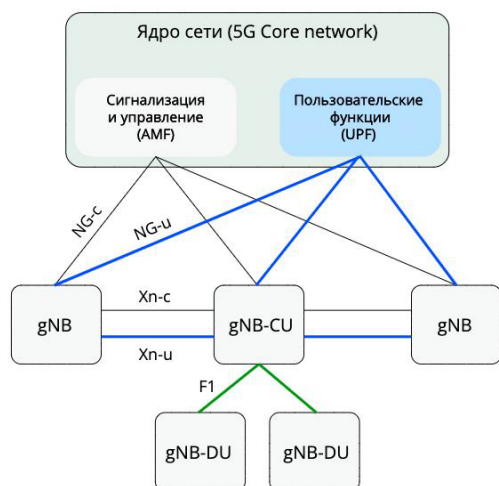


Рис. 1 Архитектура сетей радиодоступа пятого поколения

Fig. 1 The Fifth-Generation Radio Access Network Architecture

Архитектура состоит из ядра сети, узлов доступа (gNB, *от англ.* Next Generation Node B), которые с некоторыми оговорками можно считать базовыми станциями, и интерфейсов взаимодействия. Для взаимодействия gNB друг с другом используются Xn-интерфейсы, для взаимодействия с ядром сети – NG-интерфейсы. Каждый из этих интерфейсов также подразделяется на пользовательский и сигнальный. Xn-интерфейсы в основном используются для обеспечения мобильности пользователей

между базовыми станциями, поддержки множественных соединений пользователя с несколькими базовыми станциями одновременно и сопутствующего управления радио ресурсами. NG-интерфейсы используются для обеспечения функций UPF/AMF. Сигнализация и управление сетью доступа обеспечивается функциями AMF ядра сети, в то время как управление пользовательскими данными обеспечивается сетевой UPF (*от англ.* User Plane Function – функция передачи данных пользователей).

Вычислительные ресурсы MEC часто целесообразно размещать в точках агрегации доступа (например, gNB-CU). Это обусловлено, в первую очередь, наличием инфраструктуры, оставшейся от более ранних поколений сети и оказавшейся невостребованной в процессе эволюции технологий. Например, площадки, использовавшиеся под размещение мультиплексоров цифровых абонентских линий (DSLAM). Эти площадки имеют необходимую инфраструктуру, включая энергоснабжение, кондиционирование, системы безопасности и каналы связи. В дополнение к этим площадкам, размещение MEC серверов возможно и непосредственно на вынесенных базовых станциях (gNB-DU или gNB). На рынке существует достаточно большое количество соответствующих решений. Таким образом, на уровне доступа возможна организация двух уровней MEC-систем.

Кроме того, одной из отличительных особенностей систем мобильной связи пятого поколения является виртуализация сетевых функций (NFV, *от англ.* Network Functions Virtualization), реализуемая в рамках пользовательских функций ядра сети (UPF). Благодаря NFV аппаратные ресурсы могут динамически перераспределяться между сервисами, реализованными в виде программных модулей. Это также актуально и для MEC систем, в которых потребности/запросы могут меняться динамически во временной области. Таким образом, уровень ядра сети можно рассматривать в качестве точки интеграции пользовательских приложений. Данная особенность активно используется сервис-провайдерами для повышения качества услуг, например, размещения серверов распределения контента (CDN, *от англ.* Content Delivery Network) компанией Netflix в сети оператора AT&T.

Описание рассматриваемого сценария

Чтобы проиллюстрировать вычислительные функции, необходимые для реализации сервиса, предположим, что в рассматриваемой области существует m пользователей, способных удовлетворить P2P-запрос с различной степенью качества, и n пользователей этот запрос инициирующих. Приложение обрабатывает каждый полученный запрос путем поиска устройства из множества m для каждого запроса, полученного от n , решая задачу многокритериальной оптимизации, в результате чего формируется матрица коэффициентов D :

$$D = \begin{pmatrix} d_{11} & \cdots & d_{1n} \\ \vdots & \ddots & \vdots \\ d_{m1} & \cdots & d_{mn} \end{pmatrix}.$$

Наилучшее решение для каждого запроса определяется поиском наилучшего коэффициента в соответствующем столбце. При реализации приложений, подобных рассмотренному, следует учитывать, что чрезмерное количество запросов может привести к перегрузке центрального сервера [7]. Для снижения нагрузки на сервер можно использовать технологию MEC [5, 7]. В рассматриваемом нами сценарии MEC будет включать в себя три уровня вычислительных устройств, условно обозначенных как первый уровень, второй уровень и центральный сервер (облако).

Территория, находящаяся в зоне обслуживания сервера, разделена на зоны обслуживания устройств второго уровня, каждая из которых, в свою очередь, разделены на несколько зон первого уровня. Вычислительные узлы первого уровня обладают относительно низкими вычислительными возможностями: одно такое устройство обрабатывает запросы с территории покрытия одной базовой станции. Вычислительные устройства второго уровня развертываются в точке агрегации потоков от нескольких базовых станций и обеспечивают обработку пользовательских запросов в случае, если они не могут быть удовлетворены на первом уровне системы. В случае, если запрос не может быть обработан на втором уровне, его обработка будет делегирована на центральный сервер.

Модель мобильности пользователей

В результате неравномерного распределения пользователей на рассматриваемой площади, обусловленного спецификой их мобильности, возможно возникновение повышенной нагрузки на отдельных элементах инфраструктуры, что должно быть учтено в разрабатываемой модели. В данной работе мы не рассматриваем влияние специфических событий, например, таких, как массовые спортивные мероприятия, и ограничимся рассмотрением традиционных моделей мобильности пользователей, которые в условиях городской среды характеризуется мобильностью пешеходов и транспорта. В данной работе мы примем допущение, что мобильность отдельно взятого пользователя не зависит от поведения других пользователей, равно как и от предыдущего состояния системы или ее отдельных элементов, что позволит воспользоваться случайными моделями мобильности в решаемой задаче для отдельного пользователя и далее масштабировать их до необходимого количества [8].

Для учета мобильности пешеходов воспользуемся моделью RWP (от англ. Random Way-Point). В соответствии с этой моделью пользователь выбирает произвольное направление движения (точку

на рассматриваемой плоскости) и скорость, с которой он будет двигаться. При этом скорость может быть представлена как постоянной величиной, так и ожидаемым значением случайной величины, заданной на определенном интервале. При достижении выбранной точки процедура повторяется. Таким образом, процесс движения пользователя представлен некоторым количеством итераций (N), для каждой из которых движение характеризуется следующим образом:

$$\{(X_{n-1}, X_n, V_n)\}, \quad n \in N,$$

где X_n – конечная точка движения для итерации n ; X_{n-1} – начальная точка движения (конечная точка движения для итерации $n - 1$); V_n – скорость.

Точка X_n выбирается таким образом, что угол между вектором $[X_{n-1}, X_n]$ и абсциссой равномерно распределен в диапазоне $[0, 2\pi]$, а расстояние $L_n = \|X_n - X_{n-1}\|$ – положительная случайная величина. В соответствии с [8] примем, что L_n для всех пешеходов определяется распределением Релея со средним:

$$E[L] = \frac{1}{2\sqrt{\lambda}},$$

где λ – параметр, характеризующий ожидаемое значение случайной величины L_n . Большие значения λ характеризуют более короткие расстояния между точками X_n и X_{n-1} .

Время движения (T) между точками X_n и X_{n-1} определяется скоростью V_n . В данной работе примем, что скорость является неизменной. В таком случае время определяется следующим образом:

$$E[T] = E\left[\frac{L}{V}\right] = \frac{1}{2V\sqrt{\lambda}}.$$

Далее определим пространственное распределение пользователей. Для пользователя, осуществляющего движение из X_n в X_{n-1} , определим вероятность его нахождения в зоне обслуживания некоторой базовой станции в процессе этого движения. В соответствии с теоремой, доказанной в [9], такое распределение на рассматриваемой территории характеризуется функцией плотности вероятности, которая может быть определена как:

$$f(r, \theta) = \frac{\sqrt{\lambda}}{\pi r} e^{-\lambda \pi r^2}.$$

Физический смысл этой функции можно интерпретировать следующим образом: пусть $dA(r, \theta)$ – это выпуклая область на рассматриваемой площади (зона обслуживания некоторой базовой станции) с центром в точке (r, θ) в системе полярных координат, тогда вероятность $P(dA(r, \theta))$ того, что пользователь находится в этой области в процессе движения из X_n в X_{n-1} может быть аппроксимирована следующим образом:

$$P(dA(r, \theta)) = f(r, \theta) \cdot |dA(r, \theta)|,$$

где $|dA(r, \theta)|$ – это площадь интересующей нас области.

Анализ вероятностей для всех пар пользователь-БС при известной интенсивности запросов, генерируемых каждым пользователем, позволит определить ожидаемую интенсивность запросов, поступающих на каждую базовую станцию. В случае необходимости расчета вероятности переключения пользователя между БС, рассмотренная модель может быть также дополнена выводами, полученными в [10, 11]. Для описания мобильности транспорта широко используется модель MG (от англ. Manhattan Grid – сетка Манхэттена). В соответствии с данной моделью движение осуществляется вдоль «вертикальных» и «горизонтальных» улиц. Конечными точками движения на каждой итерации являются пересечения этих улиц (перекрестки). В этих точках пользователь может выбрать новое направление движения, при этом относительный угол нового направления движения кратен $\pi/2$ и выбор каждого из четырех возможных направлений движения равновероятен.

Допустим, скорость движения транспорта в данной модели также будет постоянной. Тогда в соответствии с [12], функция плотности вероятности, характеризующая пространственное распределение пользователей (в Декартовой системе координат), может быть определена как:

$$f(x, y) = \frac{3}{16}(2x - x^2 + 2y - y^2),$$

где x и y – размеры рассматриваемой площадки.

При условии, что плотность возможных маршрутов равномерно распределена по рассматриваемой территории, и при невозможности выхода пользователей за ее границы, плотность пользователей будет выше в области геометрической середины (рисунок 2). На горизонтальных осях рисунка (x, y) отложены отклонения (в метрах) относительно середины рассматриваемой области, которая имеет координаты $(0, 0)$.

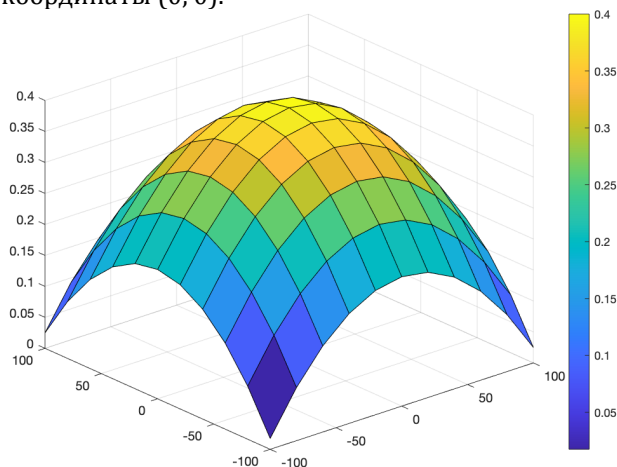


Рис. 2. Распределение плотности вероятности нахождения пользователя в соответствии с моделью MG

Fig. 2. The Probability Distribution Function for a Spatial Location of a User Following MG Model

Таким образом, для учета влияния фактора неравномерного распределения пользователей на нагрузку в системе, необходимо проанализировать совокупную плотность вероятности для заданного количества пользователей в зоне каждой из БС.

Модель обслуживания запросов

Количество запросов на обслуживание, поступающих на БС, характеризуется количеством пользователей, находящихся в зоне обслуживания, и ожидаемой интенсивностью запросов, генерируемых одним пользователем. Когда запрос принимается узлом более высокого уровня, область, обработанная на более низких уровнях, исключается из диапазона анализа. Такой подход позволяет более эффективно использовать вычислительные ресурсы. Если обработка одного запроса в узлах одного уровня занимает одинаковое количество времени, а вычислительные ресурсы, доступные на узлах, пропорциональны вычислительной сложности задачи, выполняемой на каждом уровне (размер базы данных), то время обработки запроса t также будет одинаковым для каждого вычислительного объекта.

Поскольку система допускает несколько запросов пользователя одновременно, то на вычислительных устройствах может возникнуть очередь. Также для рассматриваемой ситуации справедливо допущение о детерминированном времени обслуживания запроса на серверах, так как процедуры обслуживания представляет собой поиск в базе данных, при этом размер базы данных на устройствах одного класса можно принять примерно одинаковым. При этом интенсивность запросов, поступающих на вход системы, в аналогичных задачах моделируется Пуассоновским случайным процессом [13, 14]. Таким образом рассматриваемая система может быть описана с помощью модели M/D/1. Средняя задержка, вызванная ожиданием в очереди на устройстве первого уровня tq_1 , может быть определена из [15] следующим образом:

$$tq_1 = \frac{\lambda_1}{2 \cdot \mu_1(\mu_1 - \lambda_1)},$$

где μ_1 – скорость обслуживания на устройствах первого уровня, для $\mu = 1/t$ (для всех узлов одного уровня μ равны); λ_1 – средняя скорость поступления запросов от пользователей на устройство первого уровня, являющаяся произведением интенсивности запросов, генерируемых пользователями, и вероятности их нахождения в зоне базовой станции, рассмотренной в предыдущем разделе.

Если запрос может быть обработан на устройстве первого уровня с вероятностью P_1 , то интенсивность поступления запросов на устройство второго уровня λ_2 определяется как:

$$\lambda_2 = (1 - P_1)\lambda_1 N_1,$$

где N_1 – количество устройств первого уровня в зоне обслуживания одного устройства второго

уровня. Средняя задержка, вызванная ожиданием в очереди на устройстве второго уровня tq_2 , может быть выражена как:

$$tq_2 = \frac{(1 - P_1)\lambda_1 N_1}{2 \cdot \mu_2(\mu_2 - (1 - P_1)\lambda_1 N_1)}.$$

Наконец, если запрос может быть обработан на устройстве второго уровня с вероятностью P_2 , интенсивность поступления запросов на сервер λ_{server} определяется как:

$$\lambda_{\text{server}} = (1 - P_2)(1 - P_1)\lambda_1 N_1 N_2,$$

где N_2 – это число устройств второго уровня в рассматриваемой системе.

Средняя задержка, вызванная ожиданием в очереди на сервере tq_{server} , может быть выражена как:

$$tq_{\text{server}} = \frac{(1 - P_2)(1 - P_1)\lambda_1 N_1 N_2}{2 \cdot \mu_{\text{server}}(\mu_{\text{server}} - (1 - P_2)(1 - P_1)\lambda_1 N_1 N_2)}.$$

В соответствии с построенной аналитической моделью, среднее время обслуживания запроса T_q будет распределено следующим образом:

$$T_q = \begin{cases} tq_1 + tn_1 + t, & P_1 \\ tq_1 + tq_2 + tn_2 + 2t, & (1 - P_1)P_2, \\ tq_1 + tq_2 + tn_{\text{server}} + 3t, & (1 - P_1)(1 - P_2) \end{cases}$$

где tn_1 , tn_2 , tn_{server} – сетевые задержки между пользователем и вычислительными устройствами соответствующих уровней.

При необходимости, данное решение может быть выведено для общего случая с i иерархических уровней вычислительных устройств. Помимо инфраструктурных сетей мобильной связи, разработанная модель может быть также применена в ячеистых сетях. В этом случае обработка запроса может осуществляться на центральных узлах кластеров.

Численные результаты

Чтобы оценить преимущества от использования многоуровневой архитектуры обработки одноранговых запросов, была проведена серия экспериментов с использованием инструментария имитационного моделирования (программный комплекс WinterSIM); с набором параметров для частного случая представлен в таблице 1.

В ходе эксперимента в качестве базового сценария рассматривалась централизованная архитектура, при которой все запросы обрабатываются на одном сервере, а в качестве альтернативы – многоуровневая архитектура с МЕС [16–18]. Гистограмма времени обработки запросов, полученная в ходе эксперимента, приведена на рисунке 3. На графике наблюдается некоторый разброс результатов для случая с многоуровневой архитектурой, что обусловлено различной сетевой задержкой до серверов, на которых происходила обработка запросов.

Результаты эксперимента показывают, что многоуровневая архитектура позволяет сократить среднее

время обработки, если пользователи находятся в зоне одного устройства второго уровня. В противном случае, обработка запроса с ее использованием потребует почти вдвое большего времени по сравнению с централизованной архитектурой.

ТАБЛИЦА 1. Параметры имитационного моделирования

TABLE 1. Simulation Parameters

Параметр	Значение
Модель мобильности пользователей	RWP + MG
Длительность эксперимента	100 сек
Размер буфера на устройствах первого уровня	30 запросов
Размер буфера на устройствах второго уровня	120 запросов
Размер буфера на сервере	960 запросов
Сетевая задержка до сервера первого уровня (microcloud), tn_{micro}	1 мс
Сетевая задержка до сервера второго уровня (minicloud), tn_{mini}	6 мс
Сетевая задержка до центрального сервера, tn_s	45 мс
Время обработки запроса, t	50 мс

С целью улучшения характеристик была проведена серия дополнительных экспериментов с измененным типом маршрутизации.

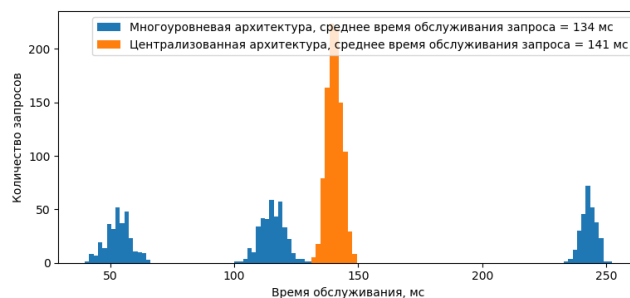


Рис. 3. Распределение времени обработки P2P запросов для различных сетевых архитектур

Fig. 3. Histogram of P2P Request Processing Time for Different Architectures

В частности, пользовательский запрос отправлялся мультивещанием на устройства первого, второго и центрального уровней. Гистограмма времени обработки запроса для данного случая представлена на рисунке 4.

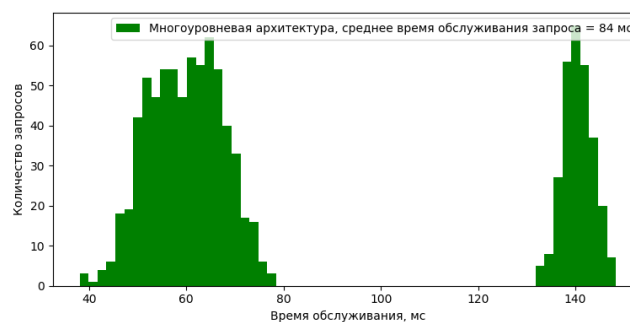


Рис. 4. Многоуровневая архитектура с многоадресной маршрутизацией P2P запросов

Fig. 4. The Request Processing Time When Using Multicasting Over the Multi-Level Architecture

Как видно из графика, наблюдается некоторый разброс в значениях времени обработки запросов, что, как и в первом случае, объясняется разницей в сетевой задержке до соответствующих серверов. В целом, мультитещательная схема отправки запросов позволяет добиться значительных преимуществ перед классической облачной архитектурой, а также многоуровневой архитектурой с однонаправленной схемой отправки запросов.

Среднее время обработки Р2Р-запросов (в мс) для всех архитектурных схем, рассмотренных в ходе эксперимента, следующее: централизованная – 141; многоуровневая – 134; многоуровневая с многоадресной маршрутизацией – 84. Как видно из результатов, стандартная МЕС архитектура не обеспечивает значительного улучшения времени обработки запросов, что в первую очередь обусловлено ее последовательным, итеративным алгоритмом поиска. В случае использования МЕС архитектуры совместно с многоадресной маршрутизацией, время обработки запросов сокращается за счет распараллеливания процесса поиска оптимального решения.

Заключение

Результаты работы показывают, что стандартная многоуровневая архитектура граничных вычислений не обеспечивает значительного улучшения времени обработки запросов одноранговых запросов в мобильных сетях. Однако производительность стандартной многоуровневой архитектуры может быть резко повышена при использовании многоадресной маршрутизации (запросы направляются одновременно на все уровни системы). Предложенная многоуровневая архитектура с многоадресной маршрутизацией улучшает работоспособность приложений, основанных на одноранговых запросах и требующих низкой задержки.

В условиях возрастающей роли самоорганизующихся сетей и одноранговых сервисов, в качестве следующего этапа развития данной работы, следует рассмотреть предоставление одноранговых сервисов в беспроводных ячеистых сетях. Это позволит обеспечить доступность одноранговых сервисов в условиях ограниченности либо отсутствия доступа к мобильной сети.

Список используемых источников

1. Zhang Z., Xiao Y., Ma Z., Xiao M., Ding Z., et al. 6G Wireless Networks: Vision, requirements, architecture, and Key Technologies // IEEE Vehicular Technology Magazine. 2019. Vol. 14. Iss. 3. PP. 28–41. DOI:10.1109/MVT.2019.2921208
2. Pirmagomedov R., Hudoev I., Shangina D. Simulation of Medical Sensor Nanonetwork Applications Traffic // Proceedings of the 19th International Conference on Distributed Computer and Communication Networks (DCCN 2016, Moscow, Russia, 21–25 November 2016). Communications in Computer and Information Science (CCIS). Vol. 678. Cham: Springer, 2016. PP. 430–441. DOI:10.1007/978-3-319-51917-3_38
3. Yousefpour A., Fung C., Nguyen T., Kadiyala K., Fatemeh J., et al. All one needs to know about fog computing and related edge computing paradigms: A complete survey // Journal of Systems Architecture. 2019. Vol. 98. PP. 289–330. DOI:10.1016/j.sysarc.2019.02.009
4. Hu Y.C., Patel M., Sabella D., Sprecher N., Young V. Mobile Edge Computing: A key technology towards 5G // ETSI White Paper. 2015. No. 11. 16 p.
5. Ren J., Yu G., He Y., Li G.Y. Collaborative Cloud and Edge Computing for Latency Minimization // IEEE Transactions on Vehicular Technology. 2019. Vol. 68. Iss. 5. PP. 5031–5044. DOI:10.1109/TVT.2019.2904244
6. Pirmagomedov R., Blinnikov M., Glushakov R., Muthanna A., Kirichek R., et al. Dynamic data packaging protocol for Real-Time Medical Applications of Nanonetworks // Proc. of the 17th Internet of Things, Smart Spaces, and Next Generation Networks and Systems (NEW2AN 2017), 10th Conference on Internet of Things and Smart Spaces (ruSMART 2017), Third International Workshop on Nano-scale Computing and Communications (NsCC 2017), St. Petersburg, Russia, 28–30 August 2017. Lecture Notes in Computer Science (LNCS). Vol. 10531. Springer: Cham, 2017. PP. 196–205. DOI:10.1007/978-3-319-67380-6_18
7. Mach P., Becvar Z. Mobile Edge Computing: A Survey on Architecture and Computation Offloading // IEEE Communications Surveys & Tutorials. 2017. Vol. 19. Iss. 3. PP. 1628–1656. DOI:10.1109/COMST.2017.2682318
8. Поселенцева Д. Ю., Замятина Е. Б. Опыт исследования алгоритмов маршрутизации и передачи данных в ad-hoc-сетях // Вестник пермского университета. Серия: Математика. Механика. Информатика. 2019. № 4(47). С. 76–85. DOI:10.17072/1993-0550-2018-4-76-85
9. Lin X., Ganti R.K., Fleming P.J., Andrews J.G. Fundamentals of mobility in cellular networks: Modeling and analysis // Proceedings of IEEE Global Communications Conference (GLOBECOM, Anaheim, USA, 3–7 December 2012). IEEE, 2012. PP. 5433–5438. DOI:10.1109/GLOCOM.2012.6503985
10. Hyytiä E., Lassila P., Virtamo J. A Markovian Waypoint Mobility Model with Application to Hotspot Modeling // Proceedings of IEEE International Conference on Communications (Istanbul, Turkey, 11–15 June 2006). IEEE, 2006. Vol. 3. PP. 979–986. DOI:10.1109/ICC.2006.254875
11. Hyytiä E., Virtamo J. Random waypoint mobility model in cellular networks // Wireless Networks. 2007. Vol. 13. Iss. 2. PP. 177–188. DOI:10.1007/s11276-006-4600-3
12. Crescenzi P., Di Ianni M., Marino A., Rossi G., Vocca P. Spatial Node Distribution of Manhattan Path Based Random Waypoint Mobility Models with Applications // Proceedings of the 16th International Colloquium on Structural Information and Communication Complexity (SIROCCO 2009, Piran, Slovenia, 25–27 May 2009). Lecture Notes in Computer Science (LNCS). Vol. 5869. Berlin, Heidelberg: Springer, 2009. PP. 154–166. DOI:10.1007/978-3-642-11476-2_13
13. Tyagi R.R., Aurzada F., Lee Ki-Dong, Reisslein M. Connection establishment in LTE-A Networks: Justification of Poisson Process Modeling // IEEE Systems Journal. 2015. Vol. 11. Iss. 4. PP. 2383–2394. DOI:10.1109/JSYST.2014.2387371

14. Chlebus E., Brazier J. Nonstationary Poisson modeling of web browsing session arrivals // Information Processing Letters. 2007. Vol. 102. Iss. 5. PP. 187–190. DOI:10.1016/j.ipl.2006.12.015
15. Cooper R.B. Introduction to queueing theory. New York: North Holland, 1981.
16. Ateya A.A., Vybornova A., Kirichek R., Koucheryavy A. Multilevel cloud based Tactile Internet system // Proceedings of the 19th International Conference on Advanced Communication Technology (ICACT, Bongpyeong, South Korea, 19–22 February 2017). IEEE, 2017. PP. 105–110. DOI:10.23919/ICACT.2017.7890067
17. Атея А.А., Выборнова А.И., Кучерявый А.Е. Многоуровневая облачная архитектура для услуг тактильного интернета // Электросвязь. 2017. № 2. С. 26–30.
18. Ateya A.A., Sayed M.S., Abdalla M.I. Multilevel Hierarchical Clustering protocol for wireless sensor networks // Proceedings of International Conference on Engineering and Technology (ICET, Cairo, Egypt, 19–20 April 2014). IEEE, 2014. DOI:10.1109/ICEngTechnol.2014.7016756

* * *

Multi-Level Peer-to-Peer Requests Processing in Mobile Networks

R. Pirmagomedov^{1, 2}, A.A. Ahmed²,

¹Tampere University,
Tampere, 33100, Finland

²The Bonch-Bruевич Saint Petersburg State University of Telecommunications,
St. Petersburg, 193232, Russian Federation

Article info

DOI:10.31854/1813-324X-2020-6-2-79-86

Received 26th February 2020

Accepted 22nd June 2020

For citation: Pirmagomedov R., Ahmed A.A. Multi-Level Peer-to-Peer Requests Processing in Mobile Networks. *Proc. of Telecom. Universities*. 2020;6(2):79–86. (in Russ.) DOI:10.31854/1813-324X-2020-6-2-79-86

Abstract: *The evolution of the Internet of Things application dictates new, more stringent requirements for the speed of transmission and processing of information. To reduce latency, peripheral multi-service access computing is gaining in popularity. This approach allows users to process data closer to their location. However, such solutions may not take into account the specifics of applications based on peer-to-peer user requests (P2P). Subject of study. The article explores the processing time of P2P requests in mobile networks. Method. The applied methodology includes methods of mathematical and simulation modeling. The main results. The results of the work allow us to conclude that the peripheral computing of multiservice access in their standard architectural design is insufficient. At the same time, the developed solution based on a multilevel architecture with multicast routing provides significantly lower query processing time. Practical significance. The solution proposed by the authors is able to ensure the operability of applications requiring minimal network latency, such as medical applications, unmanned vehicles, search for various service providers with reference to geographical coordinates.*

Keywords: 5G, P2P, IMT2030.

References

1. Zhang Z., Xiao Y., Ma Z., Xiao M., Ding Z., et al. 6G Wireless Networks: Vision, requirements, architecture, and Key Technologies. *IEEE Vehicular Technology Magazine*. 2019;14(3):28–41. DOI:10.1109/MVT.2019.2921208
2. Pirmagomedov R., Hudoev I., Shangina D. Simulation of Medical Sensor Nanonetwork Applications Traffic. *Proceeding of the 19th International Conference on Distributed Computer and Communication Networks, DCCN 2016, 21–25 November 2016, Moscow, Russia. Communications in Computer and Information Science (CCIS, vol. 678)*. Cham: Springer; 2016. p.430–441. DOI:10.1007/978-3-319-51917-3_38
3. Yousefpour A., Fung C., Nguyen T., Kadiyala K., Fatemeh J., et al. All one needs to know about fog computing and related edge computing paradigms: A complete survey. *Journal of Systems Architecture*. 2019;98:289–330. DOI:10.1016/j.sysarc.2019.02.009
4. Hu Y.C., Patel M., Sabella D., Sprecher N., Young V. Mobile Edge Computing: A key technology towards 5G // ETSI White Paper. 2015. No. 11. 16 p.

5. Ren J., Yu G., He Y., Li G.Y. Collaborative Cloud and Edge Computing for Latency Minimization. *IEEE Transactions on Vehicular Technology*. 2019;68(5):5031–5044. DOI:10.1109/TVT.2019.2904244
6. Pirmagomedov R., Blinnikov M., Glushakov R., Muthanna A., Kirichek R., et al. Dynamic data packaging protocol for Real-Time Medical Applications of Nanonetworks. *Proceedings of the 17th Internet of Things, Smart Spaces, and Next Generation Networks and Systems (NEW2AN 2017), 10th Conference on Internet of Things and Smart Spaces (ruSMART 2017), 3d International Workshop on Nano-scale Computing and Communications (NsCC 2017), 28–30 August 2017, St. Petersburg, Russia. Lecture Notes in Computer Science (LNCS, vol. 10531)*. Springer: Cham; 2017. p.196–205. DOI:10.1007/978-3-319-67380-6_18
7. Mach P., Becvar Z. Mobile Edge Computing: A Survey on Architecture and Computation Offloading. *IEEE Communications Surveys & Tutorials*. 2017;19(3):1628–1656. DOI:10.1109/COMST.2017.2682318
8. Poselentseva D.Yu., Zamyatina E.B. An experience of routing algorithms and a data transfer algorithms using simulation. *Vestnik permskogo universiteta. Seriya: Matematika. Mekhanika. Informatika*. 2019;4(47):76–85. (in Russ.) DOI:10.17072/1993-0550-2018-4-76-85
9. Lin X., Ganti R.K., Fleming P.J., Andrews J.G. Fundamentals of mobility in cellular networks: Modeling and analysis. *Proceedings of IEEE Global Communications Conference, GLOBECOM, 3–7 December 2012, Anaheim, USA*. IEEE; 2012. p.5433–5438. DOI:10.1109/GLOCOM.2012.6503985
10. Hyttia E., Lassila P., Virtamo J. A Markovian Waypoint Mobility Model with Application to Hotspot Modeling. *Proceedings of IEEE International Conference on Communications, 11–15 June 2006, Istanbul, Turkey*. IEEE; 2006. vol.3. p.979–986. DOI:10.1109/ICC.2006.254875
11. Hyttiä E., Virtamo J. Random waypoint mobility model in cellular networks. *Wireless Networks*. 2007;13(2):177–188. DOI:10.1007/s11276-006-4600-3
12. Crescenzi P., Di Ianni M., Marino A., Rossi G., Vocca P. Spatial Node Distribution of Manhattan Path Based Random Waypoint Mobility Models with Applications. *Proceedings of the 16th International Colloquium on Structural Information and Communication Complexity, SIROCCO 2009, 25–27 May 2009, Piran, Slovenia. Lecture Notes in Computer Science (LNCS, vol. 5869)*. Berlin, Heidelberg: Springer; 2009. p.154–166. DOI:10.1007/978-3-642-11476-2_13
13. Tyagi R.R., Aurzada F., Lee Ki-Dong, Reisslein M. Connection establishment in LTE-A Networks: Justification of Poisson Process Modeling. *IEEE Systems Journal*. 2015;11(4):2383–2394. DOI:10.1109/JSYST.2014.2387371
14. Chlebus E., Brazier J. Nonstationary Poisson modeling of web browsing session arrivals. *Information Processing Letters*. 2007;102(5):187–190. DOI:10.1016/j.ipl.2006.12.015
15. Cooper R.B. *Introduction to queueing theory*. New York: North Holland; 1981.
16. Ateya A.A., Vybornova A., Kirichek R., Koucheryavy A. Multilevel cloud based Tactile Internet system. *Proceedings of the 19th International Conference on Advanced Communication Technology, ICACT, 19–22 February 2017, Bongpyeong, South Korea*. IEEE; 2017. p.105–110. DOI:10.23919/ICAICT.2017.7890067
17. Ateya A.A., Vybornova A.I., Koucheryavy A.E. Multilevel cloud architecture for tactile internet applications. *Electrosvyaz*. 2017;2:26–30. (in Russ.)
18. Ateya A.A., Sayed M.S., Abdalla M.I. Multilevel Hierarchical Clustering protocol for wireless sensor networks. *Proceedings of International Conference on Engineering and Technology, 19–20 April 2014, ICET, Cairo, Egypt*. IEEE; 2014. DOI: 10.1109/ICEngTechnol.2014.7016756

Сведения об авторах:

ПИРМАГОМЕДОВ
Рустам Ярахмедович

кандидат технических наук, научный сотрудник факультета информационных технологий и коммуникаций университета Тампере, доцент кафедры сетей связи и передачи данных Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича,
rustam.pirmagomedov@tuni.fi
 <https://orcid.org/0000-0002-4860-8598>

АЛЬ АХМЕД
Арам

аспирант кафедры сетей связи и передачи данных Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича,
aramkurdian@gmail.com