

Научная статья

УДК 621.391

<https://doi.org/10.31854/1813-324X-2024-10-4-27-37>

Задача маршрутизации в сети динамических туманных вычислений

✉ Артем Николаевич Волков, artem.nv@sut.ru

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

Аннотация

Актуальность. В условиях роста трафика, перехода к сетям ИМТ-2030 и услугам Телеприсутствия, особое место занимают задачи эффективного управления сетевыми и вычислительными ресурсами. Туманные вычисления, как следующая стадия декомпозиции архитектуры многоуровневых облачных вычислений, призваны в корне изменить модели и методы распределения вычислительных задач, повлияв, в том числе, на модели взаимодействия «пользователь-оператор». На данный момент существует целый пласт научных задач для раскрытия возможностей туманных вычислений. Их возможно разделить на ряд направлений, таких как: исследование моделей и методов реализации услуг сверхнадежных сетей связи с ультрамалыми задержками, определенные еще в сетях ИМТ-2020; исследование моделей и методов обеспечения качества обслуживания, включая качество восприятия; исследование методов живой миграции микросервисов, а также групп типовых микросервисов; исследование моделей и методов распределения ресурсов динамических туманных вычислений с обеспечением устойчивости форм туманных вычислений (кластера, туманности); одним из потенциально эффективных направлений является исследования в области объединения федеративного обучения (Federated Learning) с динамическими туманными вычислениями. В данной статье решается задача маршрутизации, которую можно отнести к направлению инфраструктурных исследований динамических туманных вычислений.

Постановка задачи: исследование эффективных методов определения маршрутов в сети динамических туманных вычислений, в том числе для задач миграции микросервисов услуг Телеприсутствия.

Цель работы: исследование и разработка эффективного метода поиска пути миграции микросервисов в сети туманных вычислений, который мог бы обеспечивать учет характеристик соединений (граней графа сети), вычислительные возможности, ограничения устройств туманных вычислений, и их особенности – динамичность устройств вычислений.

Используемые методы: для апробации предложенного метода была разработана модель в среде симулятора NS-3. Анализ полученных **результатов** показал эффективность предложенного метода в рамках поставленной задачи и различных сценариях применения.

Новизна: разработан метод миграции микросервисов в качестве нового протокола маршрутизации в среде динамических туманных вычислений, отличающийся от известных тем, что этот метод обеспечивает взаимодействие устройств туманных вычислений для миграции микросервисов, при этом позволяет достичь снижения потребляемой энергии устройствами туманных вычислений на 41 % и уменьшить долю потерянных пакетов в среднем до 34 %.

Практическая значимость: разработанный метод может быть использован при реализации туманных вычислений в условиях подвижности конечных устройств, в том числе с целью достижения требований перспективных услуг сетей ИМТ-2030.

Ключевые слова: ИМТ-2030, туманные вычисления, маршрутизация

Ссылка для цитирования: Волков А.Н. Задача маршрутизации в сети динамических туманных вычислений // Труды учебных заведений связи. 2024. Т. 10. № 4. С. 27–37. DOI:10.31854/1813-324X-2024-10-4-27-37. EDN:QWBVQY

Original research

<https://doi.org/10.31854/1813-324X-2024-10-4-27-37>

Routing Task in Dynamic Fog Computing Network

 Artem N. Volkov, artem.nv@sut.ru

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

Annotation

Relevance. In the context of traffic growth, transition to IMT-2030 networks and Telepresence services, the tasks of efficient management of network and computing resources occupy a special place. Fog computing as the next stage of decomposition of the architecture of multi access edge cloud computing is designed to radically change the models and methods of distributing computing tasks, influencing, among other things, the user-operator interaction models. At the moment, there is a whole layer of scientific problems for revealing the possibilities of fog computing. They can be divided into a number of areas, such as: study of models and methods for implementing services of ultra-reliable and ultra-low latency communications, defined in IMT-2020 networks; study of models and methods for ensuring quality of service, including quality of experience; study of methods for live migration of microservices, as well as groups of typical microservices; study of models and methods for distributing resources of dynamic fog computing while ensuring the stability of fog computing forms (clusters, nebulae); one of the potentially effective areas is research in the field of combining federated learning with dynamic fog computing. This paper solves a routing problem that can be attributed to the direction of infrastructure research in dynamic fog computing.

Problem statement: research and develop the effective methods for routes determination in a dynamic fog computing network, including tasks of migrating microservices of telepresence services.

Goal of the work: research and development of an effective method for ways determination to migrate microservices in communication networks using fog computing technologies, which could take into account not only the characteristics of connections (edges of the network graph), but also the computing capabilities and limitations of fog computing devices, as well as their features - the dynamics of computing devices.

Methods: in order to test the proposed method, the program model was developed in the NS-3 modeling environment.

Result. Analysis of the results showed the effectiveness of the proposed method within the framework of the task and various application scenarios.

Novelty. A microservice migration method has been developed as a new routing protocol in a dynamic fog computing environment, which differs from the known ones in that this method ensures the interaction of fog computing devices for migrating microservices, while achieving a reduction in energy consumption by fog computing devices by 41% and reducing the share of lost packages on average up to 34%.

Practical significance: The developed method can be used to implement fog computing in conditions of mobility of end devices in order to achieve the requirements of promising services of IMT-2030 networks.

Keywords: IMT-2030, fog computing, routing

For citation: Volkov A.N. Routing Task in Dynamic Fog Computing Network. *Proceedings of Telecommunication Universities*. 2024;10(4):27–37. (in Russ.) DOI:10.31854/1813-324X-2024-10-4-27-37. EDN:QWBVQY

Введение

В течение последних 10-ти лет активно развивалась архитектура многоуровневых облачных вычислений (МЕС, аббр. от англ. Multi Access Edge Computing), эффективность внедрения в сети которой была доказана как теоретически (во многих научных публикациях), так и на практике. Последующее масштабное внедрение многоуровневой облачной архитектуры осталось только вопросом времени,

где повышение эффективности распределения ресурсов может быть благодаря совершенствованию алгоритмов мониторинга и управления, а также модернизации самого программного обеспечения облачных платформ, обеспечивающих виртуализацию и контейнеризацию. В то же время многоуровневые облачные вычисления при всей доказанной эффективности являются статическим решением, заранее спроектированным и внедренным в сеть связи. Соответственно, каждый из уровней обладает жестко

закрепленным вычислительным и сетевым ресурсом тех серверов, которые были установлены оператором. Данная особенность имеет как положительные стороны (устойчивость, предсказуемость / детерминированность), так и недостатки (необходимость предварительного проектирования, внедрения и последующего обслуживания систем, ограниченность ресурсов и прочее) [1].

В то же время исследования в области интегрированных / гибридных сетей IMT-2030 и услуг телеприсутствия показали повышение потребности в вычислительных и сетевых ресурсах, поиске более оптимальных методов распределения задач, эффективных моделей интегрированных сетей, поиске решений в задачах управления ресурсами на основе инструментов искусственного интеллекта [2]. Например, в стратегии развития отрасли связи Российской Федерации (распоряжение Правительства РФ от 24 ноября 2023 года № 3339-р) приводится прогноз дальнейшего увеличения объема глобального Интернет трафика в среднем на 30 % в год, что может привести к росту в 30 раз в 2035 г., если сравнивать с 2022 г. Таким образом можно отметить возрастающую роль цифровых решений для обеспечения эффективности во всех отраслях экономики.

На данный момент, благодаря активному развитию микроэлектроники и повышению мощности вычислительных чипов, объему и скорости чтения записи компонентов памяти (ПЗУ, ОЗУ), появлению мощных графических процессоров, в том числе с нейромодулями / нейропроцессорами, появилась возможность перехода к следующему уровню декомпозиции архитектуры облачных вычислений. Использование сетевых устройств последней мили, а также конечных пользовательских устройств в качестве вычислительных узлов сформировали новый вид распределенных облачных вычислений – туманные вычисления. Туманные вычисления на концептуальном уровне утверждены в консорциуме OpenFog в качестве спецификации в феврале 2017 г. Соглашение о создании консорциума было подписано в 2015 г. следующими участниками: ARM холдинг, Cisco, Dell, Intel, Microsoft и университет Принстона. В июле 2018 г. данный документ был принят IEEE в качестве стандарта OpenFog Reference Architecture. Стоит отметить, что данный документ заложил только технологический базис в области распределенных облачных вычислений, при этом на данный момент все еще идет стадия активных научных изысканий. В большинстве случаев в научных публикациях уделяется внимание исследованиям и разработкам в области динамических туманных вычислений, базирующихся на множестве пользовательских устройств и устройств Интернета вещей. В данном типе туманных вычис-

лений существует целый пласт научных задач: архитектурные, алгоритмические, интеграционные, протокольные и другие. В статье приводятся результаты исследований в области разработки эффективных методов маршрутизации в сети динамических туманных вычислений, в том числе для миграции микросервисов в условиях D2D (*аббр. от англ. Device-to-Device*) архитектуры сети.

Маршрутизация в сети туманных вычислений

В сетях связи известны различные методы маршрутизации, в частности динамической маршрутизации, которые основаны в первую очередь на данных о наличии маршрутов. При этом каждый маршрут имеет свою характеристику. Соответственно эти данные хранятся в маршрутизаторах и распределены в сети. Например, на практике в основе протоколов можно встретить такие алгоритмы, как:

- Distance Vector (DV), вектор расстояний, каждый маршрутизатор в сети хранит вектор расстояний до всех других сетей в сети (ярким примером является протокол RIP);

- Link-State (LS), состояние соединения, здесь каждый маршрутизатор собирает данные о состоянии всех соединений в сети и после вычисляет кратчайший путь до других сетей (примерами могут служить достаточно известные и часто применимые протоколы IS-IS или OSPF);

- Path Vector (PV), вектор пути, данные алгоритмы подобны алгоритмам DV и при этом используют в дополнение информацию о маршрутах, а не только о существующих расстояниях (например, протокол маршрутизации BGP).

Выше в качестве примеров были приведены алгоритмы протоколов, относящиеся к опорным пакетным сетям связи. В то же время, с приходом концепции Интернета вещей [3] и предшествующей ей концепции M2M, были разработаны новые протоколы маршрутизации. Данные протоколы предназначены, в первую очередь, для беспроводных сетей связи, обладающих эффектом самоорганизации (предполагалась возможная динамика узлов сети, их гибридный режим работы с возможностью «сна» устройств и изменение архитектуры сети). Так, в качестве одного из критериев при разработке алгоритмов маршрутизации в сенсорных сетях является энергоэффективность. Вследствие данного требования и других особенностей сенсорных сетей протоколы маршрутизации стали иметь реактивный характер работы. То есть маршруты не хранятся в таблицах, а строятся заново при возникновении данной потребности. В качестве примера одного из таких алгоритмов можно привести алгоритм маршрутизации AODV (*аббр. от англ. Ad Hoc On-Demand Distance Vector* – протокол динамической маршрутизации для мобильных ad-hoc сетей),

где каждый узел сети самостоятельно строит маршрут в условиях архитектуры ячеистой сети. Для примера также стоит рассмотреть протокол RPL (*аббр. от англ. Routing Protocol for Low-Power and Lossy Networks* – протокол маршрутизации для сетей с низким энергопотреблением и потерями), стандартизированный в IETF как RFC 6550. Как в названии отмечено, данный протокол в первую очередь направлен на сети с низким энергопотреблением и потерями (LLN, *аббр. от англ. Low-Power and Lossy Networks*) – устройства могут иметь достаточно ограниченную мощность и склонны к потерям пакетов ввиду различных факторов. В основе протокола RPL лежат направленные ациклические графы DODAG (*аббр. от англ. Destination Oriented Directed Acyclic Graph*). Соответственно, данный протокол образует древовидную топологию, называемую DAG (*аббр. от англ. Directed Acyclic Graph*). При этом каждый узел в сети RPL имеет предпочтительного родителя, действующего как шлюз. Для оценки маршрутов протокол использует две целевые функции: на основе оценки потери пакетов и оценки количества переходов [4]. Отличительной возможностью RPL является Energy-Aware Routing (энергосберегающая маршрутизация), то есть маршруты в сенсорной сети выбираются с учетом энергопотребления узлов. Алгоритмы маршрутизации позволяют учитывать характеристику узлов сети. Например, такие параметры как: остаточный заряд батареи, уровень сигнала и расстояние до соседних узлов. Соответственно данный протокол – одно из тех гибридных решений в маршрутизации, которое позволяет учитывать не только «стоимость» маршрута / соединения между узлами, но и их возможность.

Представленный выше достаточно краткий обзор существующих подходов к маршрутизации позволяет оценить объемы проведенных исследований и разработок в области задачи определения оптимальных маршрутов для передачи трафика. В дополнение стоит отметить концепцию программно-конфигурируемых сетей (SDN, *аббр. от англ. Software-Defined Networks*), которая в вопросах маршрутизации имеет свои отличительные возможности, основанные на принципе централизации управления сегментом сети и наличия карты сети и ее ресурсов у единого элемента мониторинга и управления – контроллера SDN.

Стоит отметить, что вышеприведенные протоколы в первую очередь направлены на выстраивание маршрута для передачи данных при различных архитектурах и сценариях работы сети: древовидная, ячеистая. Туманные вычисления, в частности рассматриваемые динамические туманные вычисления, имеют достаточно противоречивую сущность. С одной стороны, это сеть из вычислительных устройств, представляющих собой вычислительный

кластер для развертывания программного обеспечения услуг, с другой стороны – это сеть динамически передвигающихся в пространстве устройств с ограниченными физическими ресурсами (мощность и энерговооруженность), а также ограниченными вычислительными и сетевыми ресурсами, что больше напоминает самоорганизующиеся сенсорные сети. В дополнение стоит отметить, что кроме задачи в области обеспечения связности сети устройств для передачи данных, особо остро стоит вопрос служебной маршрутизации в сети динамических туманных вычислений. Для обеспечения работы fog кластера, с учетом требований качества обслуживания QoS (*аббр. от англ. Quality of Service* – уровень сервиса или обслуживания), необходимо предусмотреть реализацию процессов живой миграции контейнеров микросервисов. Теме миграции микросервисов на данный момент посвящено немало научных работ, в основном, где ставится задача определения узла в вычислительном кластере для последующей миграции контейнера услуги. Существуют также патенты, например [5].

Вопросу маршрутизации в туманных вычислениях, в том числе, посвящены ряд работ, где в большинстве своем рассматриваются частные сценарии реализации Fog и / или интеграции с другими технологиями, такими как SDN. В данной статье поднимается вопрос обеспечения эффективной маршрутизации в динамических туманных вычислениях для решения задачи живой миграции микросервиса в сценарии D2D-взаимодействия узлов без предварительного копирования контейнера с централизованного хранилища. При этом предлагаемый алгоритм учитывает многопараметрическую оптимизацию целевой функции, где, кроме оценки состояния соединения, производится оценка возможности самих узлов для передачи контейнера по fog-сети. Ранее в работе [6] была поднята проблематика маршрутизации в сетях туманных вычислений, где рассматривается вопрос учитываемых параметров сети и узла, а также проводится сравнение с протоколом AODV. В текущей статье представлены результаты работы, где приводится больше подробностей в части алгоритма маршрутизации, а также результаты более глубокого моделирования протокола, сравнение с аналогами и оценкой его эффективности.

Исследуемая модель и метод

На рисунке 1 модели сети туманных вычислений представлены следующие элементы:

- fog-устройства (FD); представлены следующим набором FD-1, FD-2, FD-3, FD-4, FD- n , FD- $(n+1)$;
- пользователи UD; представлены набором UD-1, UD-2, UD- m ;

- микросервис mS, располагающийся на устройствах FD-(n+1), FD-2 и мигрирующий между устройствами FD-n и FD-3;
- fog оркестратор, представлен набором серверов граничных вычислений Edge;
- NB-API (Northbound Application Programming Interface) – северный программный интерфейс, обеспечивающий связь между с и соответствующим fog оркестратором.

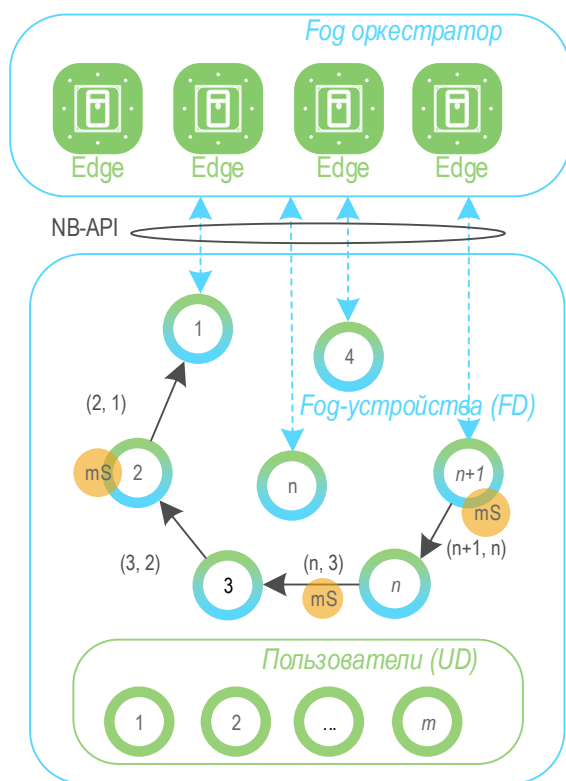


Рис. 1. Исследуемая модель
Fig. 1. The Model Under Research

На рисунке 1 приведены примеры связи между оркестратором и FD. Для полноценной работы сегмента туманных вычислений каждое устройство FD имеет соответствующую связь с оркестратором через NB-API для обмена в первую очередь служебной информацией.

В условиях динамических туманных вычислений, в первую очередь, рассматривается архитектура сети, где туманные устройства коммуницируют согласно принципу D2D, соответственно, важнейшей задачей здесь является выбор транзитных узлов для пересылки пакетов от отправителя к получателю.

Таким образом, главной характеристикой маршрута принимается стоимость миграции через него, и стоимость маршрута рассчитывается на основании группы важных параметров состояния сети. Ряд параметров относятся к определяющим QoS, на данный момент обозначенных как: использование ресурсов процессора узла, использование ресурсов

оперативной памяти узла, доля потерянных пакетов при передаче данных через канал связи, задержка при передаче данных через канал связи. Набор данных параметров отслеживается в реальном времени с помощью служебного северного API и пересылаются по сети по служебным каналам связи к оркестратору туманных вычислений. Таким образом, имея полную картину по нагрузке сети, а также анализ изменения спроса на тот или иной узел и / или микросервис, fog оркестратор контролирует процесс миграции. Вся получаемая статистика с устройства FD записывается в соответствующее хранилище данных и анализируется, и при возникновении необходимости в миграции микросервиса запускается алгоритм определения наиболее эффективного маршрута, с учетом как самих возможных путей, так и возможности самих транзитных узлов.

В работе [6] был определен алгоритм миграции микросервиса на базе анализа группы параметров. В частности, были исследованы вышеприведенные параметры и составлена фитнес-функция для расчета оценки пути. Отслеживаемые в данном случае характеристики состояния назовем метриками, учитываемыми предлагаемым протоколом. Некоторые из них также учитываются при расчете QoS:

- задержка на канале (D_{FD-n}) связи между узлами FD(n+1) и FD-n; измеряется с помощью специальных служб операционной системы и служебного ПО ведомой ноды оркестратора, запущенного на каждом узле, в миллисекундах (данные о задержке, при передаче данных через канал связи, являются достаточно важным параметром, так как миграция должна происходить достаточно быстро, чтобы остановка сервиса, перенос страниц его памяти и состояния, а также возобновление работы в новой среде, были незаметны для пользователей);
- доля потерянных пакетов (PL_{FD-n}) на канале связи между узлами FD(n+1) и FD-n; измеряется (в процентах) с помощью специальных служб, запущенных на каждом узле (важно знать долю потерянных пакетов при передаче данных через канал связи, так как доставка микросервиса на устройство назначения должна быть гарантированной, соответственно, недопустимо долгое ожидание завершения миграции в том числе потому, что сервер, на который происходит миграция сервиса, резервирует для него необходимое количество ресурсов, невозможность использования которых на протяжении продолжительного времени, при не останавливаемом процессе обработки данных может вызвать увеличение времени отклика на запросы и сбои в работе сети);
- нагрузка процессоров (CPU_{FD-n}) и оперативной памяти (RAM_{FD-n}) узла FD-n; измеряются (во флотах и килобайтах соответственно) посредством опроса каждого узла через оркестратор (важно

знать нагрузки процессора и оперативной памяти узла в активном состоянии, так как при предельном количестве используемых ресурсов у устройства может не хватить оставшихся для приема и пересылки пакетов при миграции микросервиса).

Выбор маршрута включает соответствующий процесс взаимодействия оркестратора сети туманных вычислений и n -го устройства FD, сам Алгоритм выбора оптимального маршрута миграции микросервиса на базе расчета фитнес-функции. Алгоритм осуществляет сбор статистики по использованию сетевых и вычислительных ресурсов, хранение данной статистики и последующее вычисление оптимального маршрута миграции микросервиса среди всех доступных маршрутов от узла $FD(n+1)$ до $FD-n$. Данный процесс отражен на рисунке 2.

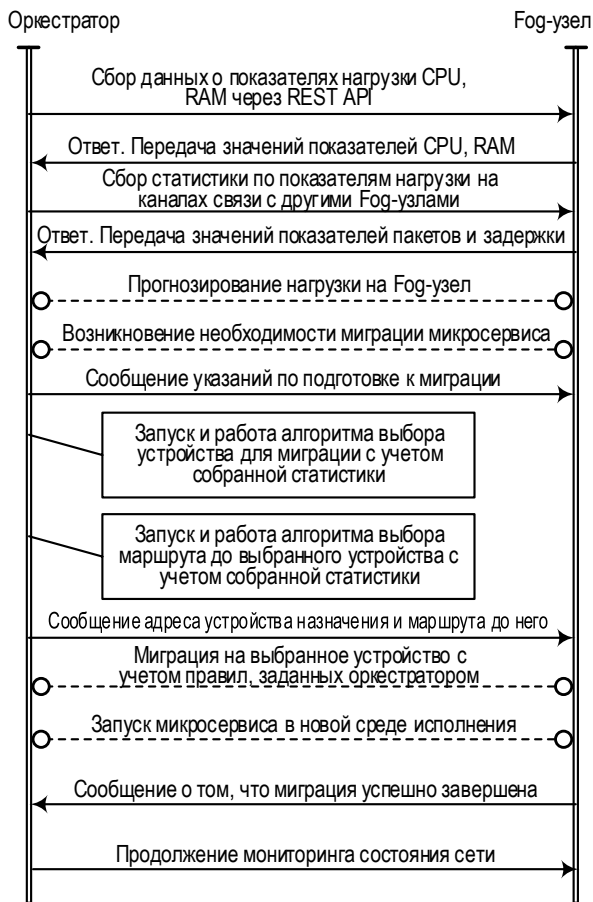


Рис. 2. Диаграмма взаимодействия оркестратора и FD- n
Fig. 2. Diagram of Interaction Between the Orchestrator and FD- n

Используется следующая фитнес-функция для нахождения стоимости одного ребра:

$$f_{FD(n+1),FD-n} = CPU_{FD-n}w_{CPU} + RAM_{FD-n}w_{RAM} + D_{FD-n}w_D + PL_{FD-n}w_{PL}, \quad (1)$$

где определены следующие веса:

$$w_{CPU} = w_D = w_{PL} = 0,275, w_{RAM} = 0,175. \quad (2)$$

Исследование влияния данных весов на конечный результат, а также обоснование их значений представлено в работе [3]. Веса должны определяться с помощью метода сравнения совокупности объектов друг с другом или метода опроса экспертов, исходя из условий поставленной задачи. Существующие веса были рассчитаны на основе метода парных сравнений. Сам метод парных сравнений предполагает сравнение всех исследуемых объектов во всех возможных парах среди этих объектов по заданному соответствующему критерию. В контексте рассматриваемой задачи были выдвинуты следующие условия:

1) нагрузка процессора узла имеет больший приоритет, чем нагрузка оперативной памяти узла (в текущей задаче мощность процессора узла туманных вычислений необходима для обработки процесса получения микросервиса и его пересылки к следующему участнику маршрута);

2) нагрузка процессора узла приоритетом эквивалентна задержке канала связи и оценке потерянных пакетов на этом канале связи (данные параметры в задаче выбора маршрута имеют одинаковую значимость и соответственно влияние на конечный результат);

3) оперативная память узла туманных вычислений имеет чуть меньший приоритет по отношению к задержке на канале связи и оценке потерянных пакетов на этом же канале связи;

4) задержка канала связи и доля потерянных пакетов на этом канале связи имеют идентичную значимость.

Функция расчета стоимости маршрута в общем виде [6] представлена следующей формулой:

$$F_{route}(1, n+1) = \sum f_{FD(n+1),FD-n}, \quad (3)$$

а искомый путь миграции микросервиса [6] в сети динамических туманных вычислений определяется как:

$$route_{migration} = \min_{CPU, RAM, D, PL} \{F_{route}(1, n+1)\}. \quad (4)$$

Полученный маршрут (4) принимается за оптимальный и сообщается вместе с остальной служебной информацией (способ миграции, узел-отправитель, узел-получатель и т. д.) соответствующим узлам туманных вычислений в сети по служебным каналам связи. Узлы принимают от оркестратора указания, после чего происходит миграция микросервиса с узла $(n+1)$ на узел 1 (см. рисунок 1). Таким образом, можно утверждать, что миграция микросервиса будет происходить по наиболее эффективному маршруту среди всех доступных в сети между узлами $FD(n+1)$ до $FD-n$.

Диаграмма взаимодействия между узлом туманных вычислений и соответствующим управляющим оркестратором отражена на рисунке 2.

Предлагаемый протокол маршрутизации микросервисов сети устройств туманных вычислений имеет также функционал обнаружения новых устройств для их последующего подключения к вычислительному кластеру. Процесс обнаружения новых устройств в сети является важной частью стабильного функционирования динамической сети туманных вычислений. Предполагается, что подключение нового устройства в сеть будет инициироваться самим новым устройством путем регистрации на специальном сервере своих учетных данных для дальнейшего добровольного предоставления свободных ресурсов устройства под задачи обработки данных. Отдельное внимание стоит уделить вопросу безопасности: необходимо обеспечить защиту от несанкционированного доступа в сеть туманных вычислений или подключение с целью нанесения ущерба, или хищения данных. Эту проблему должна решить технология цифровых сертификатов, позволяющая однозначно верифицировать и идентифицировать конкретное подключаемое устройство. Данный вопрос активно обсуждается в работе [11] Исследовательской комиссии МСЭ-Т. Однако предположим, что вопрос безопасности решен и для подключения требуется пройти аутентификацию, не допускающую злонамеренных вторжений в сеть, к примеру на основе технологии токенов с ограниченным временем жизни.

Рассмотрим возможные состояния устройства на протяжении жизненного цикла этого устройства в качестве fog-узла. На рисунке 3 представлена диаграмма состояний устройства.

На рисунке 3 состояние 0 соответствует начальному состоянию узла, именно с него начинается жизненный цикл устройства в роли fog-узла. Далее единственным возможным действием, выполняемым для перехода в состояние 1, является включение режима «поделиться ресурсами». После аутентификации устройства и его подключения к сети, как это было уточнено выше, устройство начинает считаться fog-узлом и может выполнять пересылку пакетов, то есть узел переходит в состояние 1. Перед миграцией микросервиса на устройство анализируется количество предоставленных им ресурсов, и если их достаточно, то решение о миграции и последующем выполнении задач будет принято. Если же ресурсов не хватает для решения задач обработки данных, узел будет оставаться в режиме пересылки сообщений, либо может принять решение о выходе из сети. Предположим, что ресурсов достаточно и микросервис был перенесен на новый fog-узел, после чего введен в активный режим, тогда узел переходит в состояние 2. Стоит отметить, что хоть и принцип распределения ресурсов не должен допускать перегрузки узлов, все же такие ситуации исключать нельзя. Если узел, выполняющий

задачу, теряет свою способность далее обеспечивать работу микросервиса, то может быть принято решение о выходе узла из сети, либо о переходе устройства в состояние 1, то есть в режим пересылки. Тогда устройство выполняет передачу задач иным узлам с помощью миграции микросервисов. Если же было принято решение о выходе из сети, тогда устройство также передает задачи иным узлам, имеющим для этого достаточное количество ресурсов, все микросервисы с устройства перемещаются на другие fog-узлы, после чего исходный узел покидает сеть и возвращается в состояние 0. Состояние 3 соответствует завершению жизненного цикла устройства в роли fog-узла.

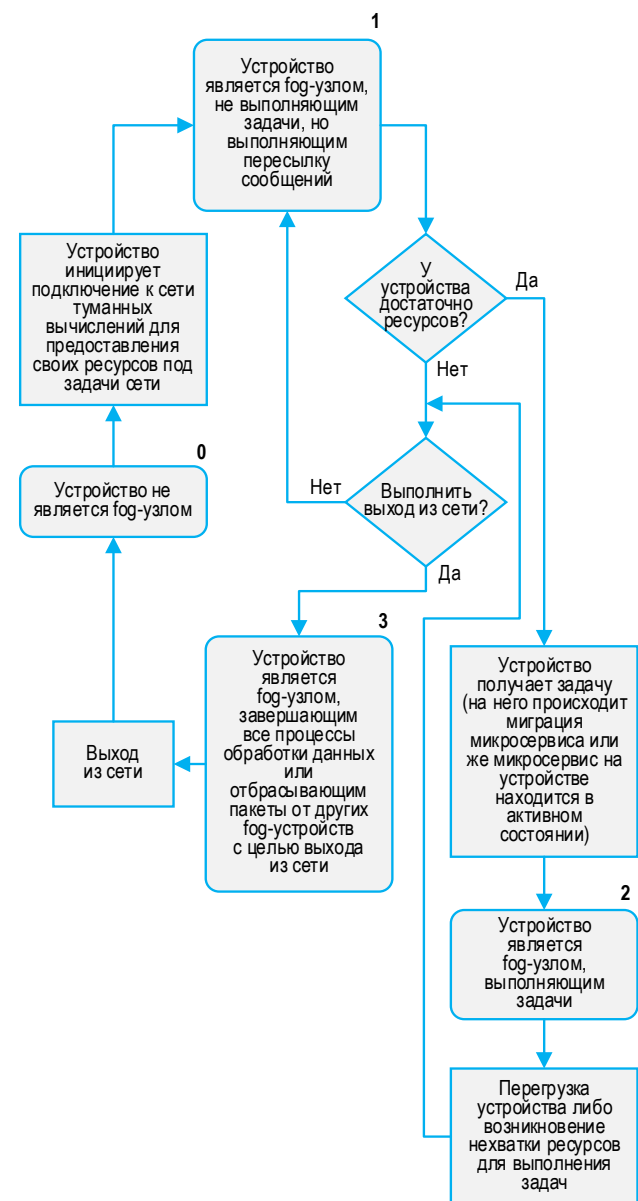


Рис. 3. Диаграмма состояния FD-n

Fig. 3. FD-n States Diagram

Рассмотрим подробнее переход узла из состояния 0 в состояние 1. При подтверждении статуса «сдачи в аренду» своих вычислительных ресурсов устройство проходит авторизацию для подключения к сети. Предлагается предоставлять владельцу устройства выбор из двух вариантов подключения: прямое подключение на базовую станцию (далее – БС) и прямое подключение на БС с поддержкой D2D-коммуникаций по отношению к узлам-соседям. В соответствии с выбранным вариантом подключения осуществляется тот или иной сценарий поведения исходного устройства как нового узла в сети туманных вычислений. Рассмотрим упомянутые сценарии. Стоит также уточнить, что согласно модели, представленной на рисунке 1, оркестратор узла туманных вычислений соответствующего сегмента находится на граничном облаке Edge, которое располагается на самой БС или рядом с ней.

1. Прямое подключение на БС

При прямом подключении на БС, сразу после верификации устройства и одобрения заявки на подключение, исходным узлом формируется сообщение Hello-E, направляемое на БС. Адрес БС сообщается узлу до отправки сообщения после верификации. Так как пользователь сам инициировал подключение к сети, сервисы, осуществляющие это подключение, в том числе отвечают за первичную установку связи между БС и новым узлом сети, они и формируют сообщение. Сообщение Hello-E необходимо для передачи своего сетевого адреса только что подключенного к сети устройства. БС, при получении от устройства этого сообщения, формирует ответ, подтверждающий получение адреса узла. После получения ответа от БС, устройство воспринимает последнюю как шлюз для обмена данными с другими узлами Fog-сети.

2. Прямое подключение на БС с поддержкой D2D

Этот процесс несколько отличается от первого описанного варианта. Помимо установки связи с БС, устройство отправляет широковещательное сообщение Hello-N, направленное на передачу своего адреса ближайшим узлам сети туманных вычислений – возможным соседям. При этом на полученное сообщение будут реагировать только те из узлов, которые также поддерживают D2D и находятся в ближайшей зоне взаимодействия с исходным узлом – в пределах одного хопа (перехода). Это необходимо не только для того, чтобы поддерживать актуальную топологию сети при подключении к ней новых устройств, но и для возможности быстро устанавливать связь с соседними узлами, если они также поддерживают D2D. Fog-узлы в данном случае будут «общаться» напрямую, поэтому пакеты, отправляемые с устройства, не будут проходить через БС. Предполагается, чтобы через БС осуществлялась связь fog-узлов с другими узлами сети, причем такими, которые находятся на расстоянии

большем, чем 1 прыжок (переход). Устройства, получившие пакет Hello, не пересылают его далее, а записывают адрес нового узла в свои таблицы маршрутизации, после чего каждый из них, в случае, если также поддерживает связь D2D, формирует ответ на сообщение Hello-N с информацией о своем адресе. Новое же устройство, получив ответ, записывает адреса из пришедших пакетов в свои таблицы маршрутизации, устанавливается дуплексная связь между узлами-соседями.

Если ни один узел не ответил на отправленное сообщение Hello-N, устройство в любой ситуации считает шлюзом для обмена данными с остальными узлами сети БС.

После подключения к сети через один из сценариев устройство полноценно становится частью сети туманных вычислений, то есть переходит в состояние 1. Также другие узлы сети получают возможность производить миграцию различных микросервисов на новый узел, в зависимости от его доступных вычислительных ресурсов, а также использовать его как транзитный при передаче пакетов, если это необходимо.

Моделирование и результаты

Для апробации предложенного протокола и его анализа была разработана программная модель и реализованы 2 сценария. Предложенный протокол для динамических туманных вычислений был реализован в среде моделирования NS-3.

Были рассмотрены два основных сценария.

Сценарий 1. Взаимодействие «устройство-Fog» (производилась оценка эффективности предлагаемого протокола, когда устройства выгружают / распределяют свои задачи на FD-n; для этого была рассмотрена топология из пяти устройств UD, подключенных к узлу туманных вычислений).

Сценарий 2. Взаимодействие «Fog-Fog» (рассматривалось десять распределенных узлов туманных вычислений с различной подвижностью, движущихся по квадрату со стороной 10 км).

В таблице 1 представлены параметры, рассматриваемые для настройки моделирования. При выполнении моделирования рассматривались разнообразные данные различных приложений, в частности: датчиков Интернета вещей, обработки изображений, обработки видео и управления дорожным движением и др. Сформированные для моделирования наборы данных были собраны из открытых источников, представленных на ресурсах [7–11].

Задачи были разделены на пять основных категорий по сложности вычислений:

- категория (I), очень простые задачи, включая данные датчиков;
- категория (II), простые задачи, эквивалентные обработке простых изображений и веб-сайтов;

- категория (III), задачи средней сложности, включая 3D-изображения и простые видеоролики;
- категория (IV), сложные задачи, в том числе обработку видео высокого качества;
- категория (V), высокосложные задачи, включая работу с 3D-видео.

ТАБЛИЦА 1. Параметры моделирования

TABLE 1. Simulation Parameters

Параметр	Значение
Сценарий 1: UD-FD	1×1км ²
Сценарий 2: FD-FD	10×10 км ²
Количество устройств UD (N)	5
Количество устройств FD	10
Размещение МЕС	равноудаленное
Скорость перемещения fog-устройств	$V_i \in \{10, 20, 30, 40, 50\}$ км/ч
Дальность передачи	500 м
Максимальный уровень передачи	20 дБм
Память / RAM	2048 Мб
Память / HDD	5 Гб
Процессор / CPU	$\epsilon \in [0.7, 2.4]$ ГГц
Максимальная рабочая нагрузка	50 событий/с
Скорость передачи данных	18 Мб/с

В рамках моделирования задачи случайным образом распределялись между рассматриваемыми устройствами, при этом каждая последующая задача назначалась после выполнения предыдущей на узле.

Для оценки предлагаемого протокола рассматривались три основных показателя производительности:

- 1) надежность предложенного протокола оценивалась путем измерения количества успешно полученных пакетов; коэффициент доставки пакетов (PDR, аббр. от англ. Packet Delivery Ratio) был рассчитан для предложенного метода и также сравнен с традиционными протоколами;
- 2) энергоэффективность с точки зрения затрат энергии на задачу;
- 3) задержка оценивалась через расчет улучшения значения задержки обработки назначенных задач.

Выражение (5) использовалось для расчета процентного улучшения задержки по сравнению с традиционными протоколами.

На рисунке 4 представлены PDR пяти рассматриваемых устройств на разных расстояниях от устройства туманных вычислений. На рисунке

представлены результаты предлагаемого метода и существующего традиционного подхода. PDR отличается от одного устройства к другому из-за их разнородности. Однако все устройства достигли более высокого PDR при использовании предлагаемого протокола. По мере удаления устройства UD от FD fog-узла PDR уменьшается; однако это снижение меньше при использовании предлагаемого протокола.

В результате анализа полученных результатов можно утверждать, что предложенный протокол улучшил PDR в среднем на 34 % по сравнению с существующими традиционными подходами.

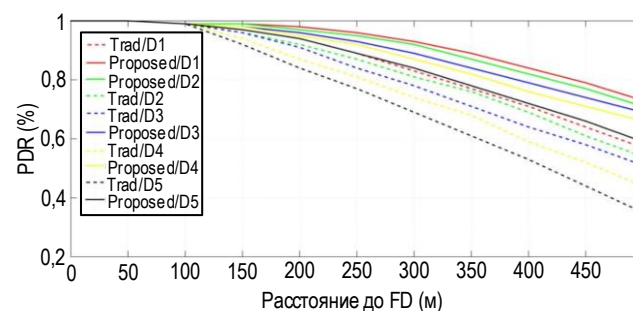


Рис. 4. Значение PDR

Fig. 4. PDR Value

На рисунке 5 представлен средний процент потребляемой энергии конечных устройств (по сравнению с полной батареей) в пяти случаях для трех методов. Три из них являются рассматриваемая топология сети с традиционным протоколом (метод 1), рассматриваемая топология с протоколом, введенным в [12] (метод 2), и топология с предлагаемым протоколом (метод 3).

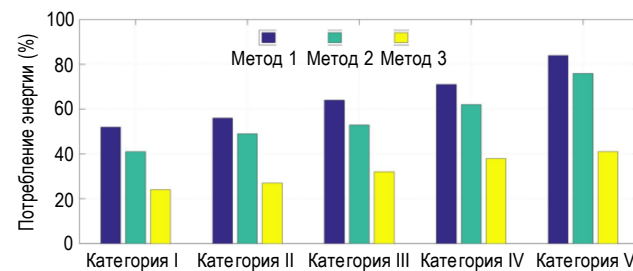


Рис. 5. Потребление энергии

Fig. 5. Energy Consumption

Пять рассмотренных случаев относятся к пяти категориям задач, особенности которых были приведены выше. Во всех пяти случаях предложенный метод позволил потреблять меньше энергии, чем в двух других.

$$Latency\ improvement\ (\%) = \frac{Average\ latency_{Traditional} - Average\ latency_{System(i)}}{Average\ latency_{Traditional}} \quad \forall i \in \mathbb{R}, \quad (5)$$

Стоит отметить, что предложенный протокол значительно снизил потребление энергии при больших рабочих нагрузках. В основном это связано с внедрением микросервисов и их грамотным распределением в туманных вычислениях

В результате анализа полученных данных моделирования можно сделать вывод, что предложенный протокол позволил добиться среднего снижения потребляемой энергии на 41 % по сравнению с методом 1 (традиционный протокол) и на 29 % по сравнению с методом 2 (протокол представлен в работе [12]).

На рисунке 6 представлены значения процента улучшения задержки предлагаемой модели по сравнению с традиционными протоколами для передаваемых данных пяти устройств. Результаты были получены для пяти категорий сервисов (определены выше).

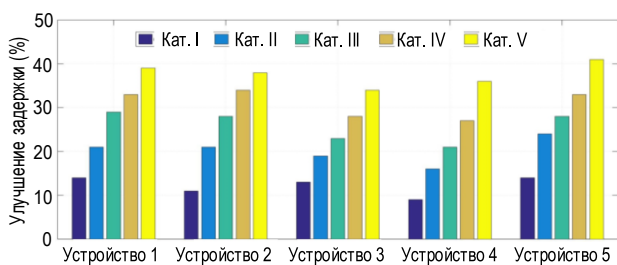


Рис. 6. Процент улучшения задержки

Fig. 6. Percentage of Latency Improvement

Таким образом, предложенный протокол уменьшил задержку для всех типов данных, при этом для

Список источников

1. Кучерявый А.Е., Маколкина М.А., Парамонов А.И., Выборнова А.И., Мутханна А.С., Матюхин А.Ю. и др. Модельная сеть для исследований и обучения в области услуг телеприсутствия // Электросвязь. 2022. № 1. С. 14–20. DOI:10.34832/ELSV.2022.26.1.001. EDN:GBQWCV
2. Волков А.Н., Мутханна А.С.А., Кучерявый А.Е., Бородин А.С., Парамонов А.И., Владимиров С.С. и др. Перспективные исследования сетей и услуг 2030 в лаборатории 6G Meganetlab СПбГУТ // Электросвязь. 2023. № 6. С. 5–14. DOI:10.34832/ELSV.2023.43.6.001. EDN:CJSYLS
3. Volkov A., Muthanna A., Koucheryavy A. AI/machine learning for ultra-reliable low-latency communication // ITU News Magazine. 2020. Iss. 5. PP. 65–68.
4. Лавшук О.А., Листопад Н.И. Метод маршрутизации в сетях ПоТ с использованием кластеризации для протокола RPL // Проблемы физики, математики и техники. 2023. № 4(57). С. 74–80. DOI:10.54341/20778708_2023_4_57_74. EDN:RVSCGG
5. Mariocco C., Doshi K., Guim Bernat F.G., Smith N.M., Spoczynski M., Verrall T., et al. Methods and apparatus to coordinate edge platforms. Patent US, no. US 2021/0014133 A1, 2021.
6. Тефилова М.Р., Кузьмина Е.А., Волков А.Н. Определение маршрута миграции микросервиса в сети туманных вычислений // Информационные технологии и телекоммуникации. 2023. Т. 11. № 1. С. 50–60. DOI:10.31854/2307-1303-2023-11-1-50-60. EDN:ETUDJB
7. Cruz S., Aguiar A. Cooperative Localization in Vehicular Networks Dataset. IEEEDataPort. 2020. URL:https://iee-dataport.org/open-access/cooperative-localization-vehicular-networks-dataset (Accessed 28 June 2024)
8. From Images to 3D Shapes (FI3S). Kaggle. URL: https://www.kaggle.com/datasets/lehomme/from-images-to-3d-shapes/fi3s (Accessed 20 April 2024)
9. TikTok Dataset // Kaggle. URL: https://www.kaggle.com/datasets/yasaminjafarian/tiktokdataset (Accessed 20 April 2024).
10. YouTube Faces With Facial Keypoints. Kaggle. URL: https://www.kaggle.com/datasets/selfishgene/youtube-faces-with-facial-keypoints (Accessed 20 April 2024)
11. Google Scraped Image Dataset. Kaggle. URL: https://www.kaggle.com/datasets/duttadebadri/image-classification (Accessed 20 April 2024)
12. Ateya A.A., Muthanna A., Gudkova I., Gaidamaka Y., Algarni A.D. Latency and energy-efficient multi-hop routing protocol for unmanned aerial vehicle networks // International Journal of Distributed Sensor Networks. 2019. Vol. 15 Iss. 8. DOI:10.1177/1550147719866392

данных микросервиса высококачественного видео, задержка была значительно снижена по сравнению с традиционными протоколами.

В результате проведенного моделирования была доказана эффективность предложенного протокола как с точки зрения уменьшения задержек, так и для сохранения энергоресурсов устройств туманных вычислений.

Заключение

В статье представлены исследования в области динамических туманных вычислений, актуальность которых растет каждый год с учетом требований и характеристик услуг телеприсутствия. Сети IMT-2030 призваны более эффективно работать с сетевыми и вычислительными ресурсами, следовать повестке бережливого производства. Данные цели сформировали целый пласт научных задач в области динамических туманных вычислений.

В работе представлен разработанный метод маршрутизации в сети динамических туманных вычислений, в том числе для задачи миграции микросервисов, позволяющий учитывать особенности и ограничения устройств туманных вычислений. Результаты моделирования нового протокола маршрутизации в среде динамических туманных вычислений показали его эффективность, а именно: снижение потребляемой энергии устройствами туманных вычислений на 41 % и уменьшение доли потерянных пакетов в среднем до 34 %.

References

1. Koucheryav A.E., Makolkina M.A., Paramonov A.I., Vybornova A.I., Muthanna A.S., Matyuhin A.Yu., et. al. Model network for research and training in the field of telepresence services. *Electrosvyaz*. 2022;1:14–20. (in Russ.) DOI:10.34832/ELSV.2022.26.1.001. EDN:GBQWCV
2. Volkov A.N., Muthanna A.S.A., Koucheryavy A.E., Borodin A.S., Paramonov A.I., Vladimirov S.S., et. al. Advanced research on networks and services 2030 in the 6G Meganetlab SPbSUT. *Electrosvyaz*. 2023;6:5–14. (in Russ.) DOI:10.34832/ELSV.2023.43.6.001. EDN:CJSYLS
3. Volkov A., Muthanna A., Koucheryavy A. AI/machine learning for ultra-reliable low-latency communication. *ITU News Magazine*. 2020;(5):65–68.
4. Lavshuk O.A., Listopad N.I. Routing method in IIoT networks using clustering for the RPL protocol. *Problems of physics, mathematics and technology*. 2023;4(57):74–80. (in Russ.) DOI:10.54341/20778708_2023_4_57_74. EDN:RVSCGG
5. Mariocco C., Doshi K., Guim Bernat F.G., Smith N.M., Spoczynski M., Verrall T., et al. *Methods and apparatus to coordinate edge platforms*. Patent US, no. US 2021/0014133 A1, 2021.
6. Tefikova M.R., Kuzmina E.A., Volkov A.N. Determination of the route migration for a microservice in a fog computing. *Telecom IT*. 2023;11(1):50–60. (in Russ.) DOI:10.31854/2307-1303-2023-11-1-50-60. EDN:ETUDJB
7. *IEEEDataPort*. Cruz S., Aguiar A. Cooperative Localization in Vehicular Networks Dataset. 2020. URL:<https://iee-dataport.org/open-access/cooperative-localization-vehicular-networks-dataset>. [Accessed 28 June 2024].
8. From Images to 3D Shapes (FI3S). *Kaggle*. URL: <https://www.kaggle.com/datasets/lehomme/from-images-to-3d-shapesfi3s> [Accessed 20 April 2024].
9. *Kaggle*. TikTok Dataset. URL: <https://www.kaggle.com/datasets/yasaminjafarian/tiktokdataset> [Accessed 20 April 2024].
10. *Kaggle*. YouTube Faces With Facial Keypoints. URL: <https://www.kaggle.com/datasets/selfishgene/youtube-faces-with-facial-keypoints> [Accessed 20 April 2024].
11. *Kaggle*. Google Scraped Image Dataset. URL: <https://www.kaggle.com/datasets/duttadebadri/image-classification> [Accessed 20 April 2024].
12. Ateya A.A., Muthanna A., Gudkova I., Gaidamaka Y., Algarni A.D. Latency and energy-efficient multi-hop routing protocol for unmanned aerial vehicle networks. *International Journal of Distributed Sensor Networks*. 2019;15(8). (in Russ.) DOI:10.1177/1550147719866392


Статья поступила в редакцию 10.07.2024; одобрена после рецензирования 09.08.2024; принята к публикации 12.08.2024.

The article was submitted 10.07.2024; approved after reviewing 09.08.2024; accepted for publication 12.08.2024.

Информация об авторе:

ВОЛКОВ
Артем Николаевич

кандидат технических наук, доцент, кафедры сетей связи и передачи данных
Санкт-Петербургского государственного университета телекоммуникаций
им. проф. М.А. Бонч-Бруевича

 <https://orcid.org/0009-0002-4296-1822>

Автор сообщает об отсутствии конфликтов интересов.

The author declares no conflicts of interests.