

ИССЛЕДОВАНИЕ И ОБЗОР МЕХАНИЗМОВ БАЛАНСИРОВКИ НАГРУЗКИ НА ОСНОВЕ SDN В 5G/IMT-2020

Б. Данешманд, Л.А. Ту

Санкт-Петербургский государственный университет информационных технологий, механики
и оптики, г. Санкт-Петербург, Россия

Аннотация: ожидается, что рост числа мобильных устройств и потребность в пользовательских данных к 2030 году окажут беспрецедентное давление на текущую мобильную сеть. У будущих мобильных сетей должно быть несколько требований в отношении объема данных, задержки, качества обслуживания и опыта, мобильности, спектра и энергоэффективности. Поэтому в последнее время начались усилия по созданию более эффективных решений для мобильных сетей. С этой целью балансировка нагрузки привлекла большое внимание как многообещающее решение для более эффективного использования ресурсов, повышения производительности системы и снижения эксплуатационных расходов. Это эффективный способ сбалансировать трафик и уменьшить перегрузку в гетерогенных сетях в будущих сетях 5G / IMT-2020. Балансировка нагрузки - одна из наиболее важных задач, необходимых для максимального повышения производительности, масштабируемости и надежности сети. В настоящее время с появлением программно-конфигурируемых сетей (SDN) балансировка нагрузки для SDN стала важной проблемой в будущей сети 5G / IMT-2020. SDN позволяет использовать программируемые балансировщики нагрузки и обеспечивает гибкость для разработки и реализации стратегий балансировки нагрузки. В этом обзоре мы выделяем методы балансировки нагрузки на основе сетей SDN и предполагаемые требования к балансировке нагрузки в сетях 5G

Ключевые слова: балансировка нагрузки, программно-конфигурируемые сети, SDN, 5G / IMT-2020

Введение

Сеть пятого поколения (5G / IMT-2020) была запущена в 2018 году в Южной Корее [1]. Цель 5G - обеспечить высокую пропускную способность, уменьшить задержку, увеличить пропускную способность. Спрос на трафик данных со стороны пользователей мобильного широкополосного доступа рос за последние несколько лет. Сети 5G предназначены для поддержки мобильных широкополосных подключений и миллиардов устройств M2M, а также сверхнадежных устройств связи с малой задержкой в будущем [2]. Базовые станции 5G оптимизированы для поддержки малых задержек для таких устройств. Кроме того, согласно ежегодному прогнозу Cisco, количество пользователей Интернета в мире достигнет 5,3 миллиарда в 2023 году. Более того, по прогнозам Cisco, к 2023 году будет 14,7 миллиарда M2M-подключений. Следовательно, операторы должны разработать эффективную обработку данных в 5G, учитывая потенциал будущего роста использования данных и роста числа подписчиков. Безопасная инфраструктура существующих сетей и сервисы костей приводят к сложному, неэффективному распределению ресурсов и низкому использованию сетевых

ресурсов, особенно в беспроводных сетях. Различные методы балансировки нагрузки основаны на полезности поставщика услуг и удовлетворенности пользователей. Сеть 5G на основе SDN - еще одна область исследований для распределения ресурсов и подключения к ним в сети 5G. В этой статье представлен обзор балансировки нагрузки интеграции 5G с SDN. Программно-конфигурируемая сеть (SDN) - это развивающаяся архитектура, которая позволяет физически отделить панель управления сетью от уровня пересылки или уровня инфраструктуры, где панель управления управляет несколькими устройствами. Преимущества, предлагаемые SDN, включают автоматическую балансировку нагрузки, выделение ресурсов по запросу и возможность масштабирования сетевых ресурсов. Кроме того, SDN сокращает расходы на управление оборудованием и снижает затраты. Сети предоставляются без ручной настройки. Он предоставляет компаниям платформу для подготовки к новым технологиям, таким как облачные приложения, устройства Интернета вещей и приложения для работы с большими данными.

Балансировка нагрузки - это метод разделения рабочей нагрузки на несколько ресурсов во избежание перегрузки любых ресурсов [3]. Цели балансировки нагрузки - максимизировать пропускную способность, минимизировать

время отклика и оптимизировать трафик. Существенными и функциональными компонентами SDN являются протокол и контроллеры OpenFlow. Сетевой контроллер - это мозг архитектуры SDN. Он находится между сетевыми устройствами и приложениями. Он основан на операционных системах в вычислительной технике. В [4] контроллер определяется как программная абстракция, которая контролирует все функции любой сетевой системы. Он поддерживает контроль над сетью через интерфейсы, первый - южный интерфейс (например, OpenFlow), второй - северный интерфейс (например, API). Южный интерфейс абстрагирует функциональные возможности программируемых переключателей и подключает их к контроллеру. NOX - одна из первых общедоступных реализаций контроллера OpenFlow в Windows, Linux, Mac OS и других платформах. OpenFlow - первый из ведущих авторизованных коммуникационных интерфейсов, объединяющий уровни пересылки и управления в архитектуре SDN, обеспечивает прямой доступ и управление плоскостями пересылки на сетевом устройстве, таком как виртуальные или физические коммутаторы и маршрутизаторы. В этой статье мы сравниваем и рассматриваем алгоритмы балансировки нагрузки на основе SDN, выполняемые исследователями, и описываем преимущества и недостатки каждого метода. Балансировка нагрузки является важным компонентом сетевой инфраструктуры пятого поколения из-за растущего числа подключенных к Интернету устройств, ресурсы которых распределяются по широкому спектру систем и требуют подписки от конечных пользователей.

Архитектура SDN и обзор балансировки нагрузки на основе SDN

SDN Архитектура

SDN - это полностью программно конфигурируемая компьютерная сеть, в которой уровни управления самой сетью и передачей данных отделены друг от друга путем передачи функций управления отдельному устройству - сетевому контроллеру [3]. SDN определяется способностью динамически управлять поведением сети с помощью программного обеспечения через открытые интерфейсы. Основное отличие SDN от обычных сетей заключается в централизованном интеллектуальном управлении и мониторинге сети, что позволяет прове-

рять, контролировать и изменять передаваемые потоки данных.

Технология SDN направлена на решение следующих задач:

- Повышение эффективности механизмов управления пропускной способностью сети.
- Снижение капитальных и эксплуатационных затрат.
- Упрощение управления сетью и повышение уровня ее автоматизации.
- Ускорение и автоматизация процесса создания новых сервисов и их запуска.
- Повышение безопасности всей информационно-коммуникационной системы.
- Повышение эффективности маршрутизации.

Сетевая инфраструктура SDN должна быть построена в соответствии с открытыми протоколами. OpenFlow должен быть унифицированным и обеспечивать возможность реализации мультивендорных технических решений. Архитектура SDN разделена на три основных уровня, а именно: уровень инфраструктуры (плоскость пересылки), уровень управления (плоскость управления) и уровень приложения, объединенные друг с другом, как показано на рис. 1.

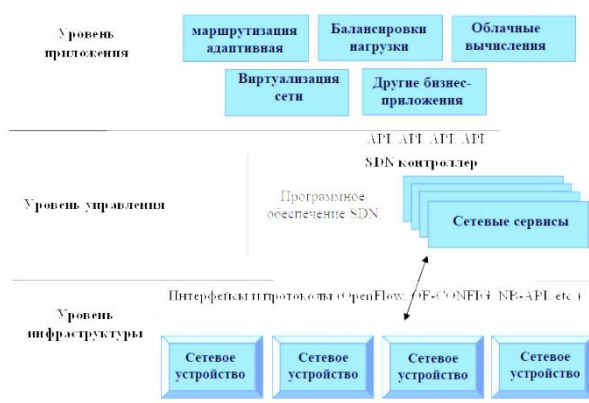


Рис. 1. Архитектура SDN

Балансировка нагрузки в SDN

Как правило, балансировка нагрузки обычно представляет собой отдельно развертываемую функцию в сети, которая распределяет нагрузку трафика между сетевыми путями (или элементами центра обработки данных) в соответствии с конкретными метриками оптимизации, такими как минимальная средняя нагрузка или стоимость канала.

Текущие решения для балансировки нагрузки эффективны, но предлагают ограниченную гибкость с точки зрения настройки. Благодаря технологии SDN балансировка нагрузки может быть интегрирована с любым элементом передачи в сети, например коммутатором OpenFlow, что устраняет необходимость в отдельных устройствах. Кроме того, SDN позволяет балансировать нагрузку для работы с любой степенью детализации потока.

Как работает балансировка нагрузки с помощью SDN? Балансировка нагрузки SDN удаляет протоколы на аппаратном уровне, чтобы улучшить управление сетью и диагностику. Балансировка нагрузки контроллера SDN принимает решения по управлению трафиком данных, не полагаясь на алгоритмы, определенные традиционным сетевым оборудованием. Балансировщик нагрузки на основе SDN экономит время работы, управляя всей сетью приложений и веб-серверами.

Методы балансировки нагрузки можно разделить на статические, динамические (распределенные и централизованные) или комбинированные (гибридные), как показано на рис. 2.

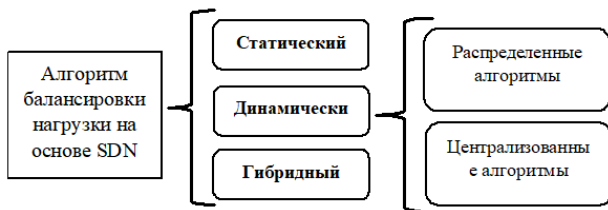


Рис. 2. Алгоритм балансировки нагрузки на основе SDN

А. Статическая балансировка нагрузки

В статическом алгоритме трафик равномерно распределяется между серверами. Статический алгоритм подходит для систем с небольшими изменениями нагрузки. Этот алгоритм должен иметь предварительную информацию из системных источников, чтобы гарантировать, что решение о переключении нагрузки не зависит от текущего состояния системы [6]. Однако алгоритмы статической балансировки нагрузки имеют недостаток, заключающийся в том, что задачи назначаются процессору или машине только после создания, а задачи не могут переноситься на любое другое устройство для балансировки нагрузки во время выполнения [7].

В. Динамическая балансировка нагрузки

Методы динамической балансировки нагрузки более эффективны, чем статические аналоги из-за динамического распределения предварительно запрограммированных шаблонов балансировки нагрузки [8], [9]. Правильная балансировка нагрузки имеет решающее значение для оптимизации минимального времени отклика, максимальной пропускной способности, минимального потребления ресурсов, масштабируемости и предотвращения перегрузки ресурсов. Метод динамической балансировки нагрузки может быть получен из двух методов.

1. В нераспределенном методе один узел (централизованный) получает все запросы и распределяет их по серверам. Централизованные контроллеры реализуют всю логику уровня управления в одном месте. В таком контроллере один сервер берет на себя все действия на уровне управления. Основными преимуществами этих контроллеров являются простота и управляемость, поскольку они обеспечивают единую точку управления. Однако они страдают от проблем с масштабируемостью, поскольку каждый сервер имеет ограниченную емкость для обработки устройств плоскости данных. Несколько распределенных алгоритмов упоминаются в следующих разделах.

- QoS-Aware алгоритм;
- Эвристические подходы;
- Балансировка нагрузки Wardrop.

2. В распределенном методе все узлы используются совместно с распределением запросов. Распределенный контроллер не имеет проблем с масштабируемостью и обладает преимуществом высокой производительности при больших нагрузках трафика. Существует несколько централизованных алгоритмов, упомянутых в следующих подразделах:

- Платформа управления маршрутизацией (RCP).
- Алгоритм балансировки нагрузки на базе сервера.
- Алгоритм DUTE.

С. Гибридная балансировка нагрузки

Эти методы используются для преодоления недостатков методов динамической и статической балансировки нагрузки, а также для сбора преимуществ и недостатков статических и динамических алгоритмов для разработки

нового метода [8]. Это означает, что объединение преимуществ двух или более существующих алгоритмов, динамического или статического алгоритма, может создать новый алгоритм.

Алгоритмы динамической балансировки нагрузки могут показаться более эффективными, чем статические, но они более сложны, поскольку учитывают текущее состояние системы. Эта сложность часто приводит к накладным расходам и неправильным решениям по балансировке нагрузки. Вот почему так важно получить полную картину механизма балансировки нагрузки, прежде чем выбирать балансировщик нагрузки.

Таблица 1

Сравните статические и динамические характеристики

Особенности балансировки нагрузки	Статическая балансировка нагрузки	Динамическая балансировка нагрузки
Распределение нагрузки	Во время компиляции	Во время выполнения
Стабильность	Более	Меньше
Сложность и стоимость	Меньше	Более
Надежность и время отклика	Меньше	Более
Предсказуемость	Легкая	Трудная
Перегрузка процессора и колебание состояния	-	Значительные
Утилизация ресурсов	Меньше	Более
Задержка связи	Меньше	Более

Важной и фундаментальной особенностью статических методов является предварительное знание системы. Правило программируется непосредственно в балансировщике нагрузки статическими методами, поскольку поведение пользователя непредсказуемо. Методы статической балансировки нагрузки в сети могут быть неэффективными. Динамические методы более эффективны, поскольку нагрузка распределяется динамически в соответствии с некоторыми запрограммированными шаблонами в балансировке нагрузки [23]. Правильная балансировка нагрузки помогает максимизировать масштабируемость, минимизировать время отклика, максимизировать энергопотребление, минимизировать потребление ресурсов, предотвратить перегрузку любого отдельного ресурса и многое другое. Основные качественные параметры для балансировки нагрузки в SDN представлены в табл. 2.

Таблица 2

Параметры балансировки нагрузки

№	Параметры балансировки нагрузки	Описание
1	Среднее количество синхронизаций в минуту	Среднее количество синхронизаций состояния контроллера в минуту в SDN.
2	Накопленная частота	Индекс производительности, позволяющий оценить точность алгоритма.
3	Степень балансировки нагрузки	Показатель равномерности распределения нагрузки между объектами.
4	Потребление энергии	Количество потребляемой энергии в сети.
5	Время исполнения	Продолжительность выполнения программы.
6	Пересылка записей	Маршрутизаторы используют таблицу пересылки, чтобы принять решение об отправке пакета.
7	Гарантированная скорость передачи данных (GBR)	Один из параметров качества обслуживания (QoS) в сетях для обеспечения пропускной способности канала-носителя.
8	Задержка	Время, необходимое для пересылки пакета по сети.
9	Стоимость миграции	Она состоит из двух основных затрат: стоимости обмена сообщениями и стоимости загрузки. Некоторые сообщения должны передаваться между контроллерами для переключения миграции, например запросы на миграцию, запросы ролей и асинхронные сообщения.
10	Накладные расходы	Любое сочетание чрезмерного или косвенного времени вычислений, памяти, полосы пропускания или других ресурсов, необходимых для выполнения конкретной задачи, является накладным.
11	Скорость потери пакетов	Потеря пакетов происходит, когда один или несколько пакетов не достигают места назначения. Обычно это происходит из-за перегрузки сети. Это процент потерянных пакетов по сравнению с отправленными.

Продолжение табл. 2

12	Коэффициент пиковой нагрузки	Для измерения характеристик маршрута.
13	Доля согласованных потоков сроков	Этот параметр представляет процент потоков, удовлетворяющих крайнему сроку.
14	Использование ресурсов	Степень использования сетевых ресурсов, таких как использование канала, полосы пропускания, процессора и памяти.
15	Время отклика	Он определяется интервалом, который начинается от принятия запроса или задания до ответа на запрос или задачу для сервера.
16	Среднеквадратичная ошибка (RMSE)	Метрика для оценки производительности балансировки нагрузки. Лучшая производительность имеет меньшее RMSE.
17	Пропускная способность	Количество данных, правильно перемещенных из одного места в другое за определенный период.
18	Нагрузка	Объем работы, которую должен выполнить контролер. Чтобы сбалансировать рабочие нагрузки между контроллерами, были введены подходы к балансировке нагрузки.

Балансировка нагрузки может быть реализована в программном или физическом оборудовании. Выбранный метод будет зависеть от типа обслуживаемой службы или приложения, а также от состояния сети и серверов на момент запроса. Эти методы будут использоваться в комбинации для определения лучшего сервера для обслуживания новых запросов. Текущий уровень запросов к балансировщикам нагрузки часто определяет, какой метод используется. Когда нагрузка невелика, то достаточно одного из простых методов балансировки нагрузки. Во время высокой нагрузки более сложные методы обеспечивают равномерное распределение запросов [3, 10, 11].

Балансировка нагрузки в SDN приводит к обнаружению наилучшего пути и сервера для самой быстрой доставки запросов.

Балансировка нагрузки SDN имеет следующие преимущества:

- Более низкая стоимость;

- Большая масштабируемость;
- Более высокая надежность;
- Гибкость в настройке;
- Уменьшено время на развертывание;
- Автоматизация;
- Возможность построить сеть без какого-либо программного/аппаратного обеспечения конкретного производителя.

Для оптимизации сетевого потока и достижения интеллектуальной балансировки нагрузки использовались различные методы и алгоритмы. Мы упомянули, что типы текущих методов балансировки нагрузки на основе SDN, которые используются в сетях пятого поколения, описаны ниже:

❖ Балансировка нагрузки мультиконтроллеров SDN

Единственный контроллер представляет собой проблему узкого места. Использование централизованного контроллера может ограничить масштабируемость и надежность, в то время как децентрализованные контроллеры в сетях SDN работают лучше. Таким образом, использование нескольких контроллеров станет жизненно важным решением для улучшения масштабируемости, надежности и возможностей плоскости управления. Чтобы повысить производительность балансировки нагрузки, важно использовать контроллеры с несколькими SDN вместо централизованных контроллеров [12, 13, 14].

❖ Балансировка нагрузки сервера (SLB)

Серверная балансировка нагрузки (SLB) развертывает один балансировщик нагрузки перед несколькими серверами. Таким образом, можно разумно распределить нагрузку на несколько серверов, чтобы полностью использовать ресурсы сервера. В [11] авторы предлагают стратегию балансировки нагрузки для распределения запросов клиентов по нескольким серверам. Балансировка нагрузки сервера обеспечивает сетевые службы и доставку контента с использованием ряда алгоритмов балансировки нагрузки.

ИТ-команды все чаще полагаются на балансировщики нагрузки на серверы, здесь мы упомянули некоторые преимущества балансировки нагрузки сервера (SLB):

- Повышение масштабируемости;
- Резервирование;

- Обслуживание и производительность.

❖ **Выбор различных ссылок. Балансировка нагрузки**

Система активно контролирует качество с точки зрения задержки, потери пакетов и джиттера, а также пропускную способность с точки зрения пропускной способности каналов и направляет трафик на наиболее подходящий канал на основе бизнес-намерений, определенных как политика. Этот метод увеличивает масштабируемость и эффективное использование сетевых ресурсов. В [15] авторы предложили алгоритм оптимизации пути динамической балансировки нагрузки (DLPO), основанный на SDN, полезный для сетевых топологий центров обработки данных. Алгоритм с несколькими каналами может быстро сбалансировать нагрузку на каналы в сети, чтобы исключить перегруженные пути [15, 16, 17].

❖ **Балансировка нагрузки на основе искусственных нейронных сетей**

Искусственные нейронные сети (ИНС) прогнозируют спрос и, таким образом, распределяют ресурсы в соответствии с этим спросом. Таким образом, ИНС всегда поддерживают активные серверы в соответствии с текущим спросом, что приводит к более низкому энергопотреблению, чем консервативный подход с избыточным выделением ресурсов. Кроме того, высокое использование серверов приводит к большему энергопотреблению, сервер, работающий с более высоким коэффициентом использования, может обрабатывать больше рабочих нагрузок с аналогичным энергопотреблением. Использование метода искусственной нейронной сети (ANN), основанного на SDN в балансировке нагрузки, связано с тем, что этот метод способен категоризировать и выбирать входные данные в указанные группы или заранее определенные пути с основной ролью в балансировке нагрузки [18, 19, 20, 21].

❖ **Балансировка нагрузки в стандарте IEEE 802.11 (беспроводные каналы связи)**

Стандарт IEEE 802.11 указывает, что клиентское устройство решает, к какой точке доступа подключаться. В средах с высокой плотностью пользователей выбор клиентского устройства для подключения к одной точке до-

ступа может привести к перегрузке точки доступа [22]. Это также может привести к колебаниям в ассоциации AP, поскольку клиентское устройство имеет ограниченные данные о производительности сети. Алгоритмы, развернутые на контроллере SDN, могут динамически балансировать нагрузку точек доступа (AP), выбирая менее загруженные из области совместного использования для ассоциации клиента.

Исследование методов балансировки нагрузки

Несколько методов и алгоритмов могут применяться с использованием нескольких методов организации очередей в методах балансировки нагрузки в сети. Выбранный метод будет зависеть от типа обслуживаемой службы или приложения, а также от состояния сети и серверов на момент запроса. Описанные ниже методы будут использоваться в комбинации для определения лучшего сервера для обслуживания новых запросов. Текущий уровень запросов к балансировщикам нагрузки часто определяет, какой метод используется. Когда нагрузка невелика, то достаточно одного из простых методов балансировки нагрузки. Во время высокой нагрузки используются более сложные методы для обеспечения равномерного распределения запросов. Программная балансировка нагрузки (SLB) обычно предлагается как контроллер доставки приложений (ADC), который работает на стандартном сервере или виртуальной машине. Аппаратное устройство балансировки нагрузки (HLD) - это автономное оборудование, на котором выполняется программное обеспечение балансировки нагрузки.

Обычно он развертывается как часть пары на случай отказа одного устройства балансировки нагрузки. Программная балансировка нагрузки предлагает те же функции, что и HLD, но не требует специального устройства балансировки нагрузки. Программное обеспечение для балансировки нагрузки может работать на обычном сервере или даже на виртуальном сервере. В табл. 3 мы упомянули преимущества и недостатки каждого метода.

Таблица 3

Характеристики SLB и HLB	
SLB (БАЛАНСИРОВКА НАГРУЗКИ НА ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ) ¹	
ПРЕИМУЩЕСТВА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ	НЕДОСТАТКИ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ
<ul style="list-style-type: none"> Гибкое внедрение. Большая масштабируемость для балансировки нагрузки программного обеспечения. Уменьшение затрат на балансировку нагрузки через облако. 	<ul style="list-style-type: none"> При масштабировании сверх начальной емкости может возникнуть некоторая задержка при настройке программного обеспечения балансировки нагрузки. Текущие расходы на обновления.
HLB (БАЛАНСИРОВКА НАГРУЗКИ НА АППАРАТНОЕ ОБЕСПЕЧЕНИЕ) ²	
АППАРАТНЫЕ ПРЕИМУЩЕСТВА	АППАРАТНЫЕ НЕДОСТАТКИ
<ul style="list-style-type: none"> Высокая пропускная способность благодаря программному обеспечению, работающему на специализированных процессорах. Повышенная безопасность, поскольку только организация может получить физический доступ к серверам. Фиксированная стоимость после покупки. 	<ul style="list-style-type: none"> Требуется больше персонала и опыта для настройки и программирования физических машин. Невозможность масштабирования при установленном ограничении количества подключений. В подключении отказано или обслуживание ухудшается до тех пор, пока не будут приобретены и установлены дополнительные машины. Более высокая стоимость покупки и обслуживания физического балансировщика сетевой нагрузки. Владение аппаратным балансировщиком нагрузки также может потребовать платных консультантов для управления им

Балансировка нагрузки имеет жизненно важное значение для решения проблемы чрезмерного трафика в сети. Это было одно из первых привлекательных приложений в сетях SDN. Обычно следующие методы используются в комбинации для поиска лучшего сервера для ответа на новые запросы. В этом разделе мы рассмотрим некоторые из наиболее часто используемых подходов к балансировке нагрузки, которые кратко изложены в табл. 4.

Таблица 4

Метод балансировки нагрузки			
№	Методы балансировки нагрузки	Описание	С ³ / Д ⁴
1	Round Robin (RR)	Простой способ распределения клиентских запросов по группе серверов. Клиентский запрос пересылается каждому серверу по очереди.	С
2	Weighted Round Robin (WRR)	Основан на простом алгоритме балансировки нагрузки Round Robin для учета различных характеристик сервера приложений. Администратор назначает вес каждому серверу приложений на основе выбранных ими критериев, чтобы продемонстрировать способность серверов приложений обрабатывать трафик.	С
3	Least Connection	Алгоритм динамической балансировки нагрузки, при котором клиентские запросы распределяются на сервер приложений с наименьшим количеством активных соединений на момент получения клиентского запроса. В случаях, когда серверы приложений имеют аналогичные характеристики, сервер приложений может быть перегружен из-за более длительных подключений; этот алгоритм учитывает активную нагрузку на соединение.	Д
4	Weighted Least Connection	Основан на алгоритме балансировки нагрузки по наименьшему количеству подключений для учета различных характеристик сервера приложений. Администратор назначает вес каждому серверу приложений на основе выбранных ими критериев, чтобы продемонстрировать способность серверов приложений обрабатывать трафик.	Д
5	Software Defined Networking (SDN) Adaptive	SDN Adaptive - это алгоритм балансировки нагрузки, который объединяет знания уровней 2, 3, 4 и 7 и ввод от контроллера SDN для принятия более оптимизированных решений по распределению трафика. Это позволяет информации о статусе серверов, статусе запущенных на них приложений, работоспособности сетевой инфраструктуры и уровне перегрузки в сети играть роль в принятии решения о балансировке нагрузки.	Д

¹Software Load Balancing

²Hardware Load Balancing

³Статический

⁴Динамический

Продолжение табл. 4

6	Fixed Weighting	Алгоритм балансировки нагрузки, при котором администратор назначает вес каждому серверу приложений на основе критериев по их выбору, чтобы продемонстрировать способность серверов приложений обрабатывать трафик. Сервер приложений с наибольшим весом получит весь трафик. Если сервер приложений с наибольшим весом выходит из строя, весь трафик будет направлен на следующий сервер приложений с наибольшим весом.	Д
7	Weighted Response Time	Алгоритм балансировки нагрузки, при котором время ответа серверов приложений определяет, какой сервер приложений получит следующий запрос. Время ответа сервера приложений на проверку работоспособности используется для расчета весовых коэффициентов сервера приложений. Сервер приложений, отвечающий быстрее всех, получает следующий запрос.	Д
8	Source IP Hash	Алгоритм, который объединяет исходный и целевой IP-адреса клиента и сервера для создания уникального хеш-ключа. Ключ используется для привязки клиента к определенному серверу. Поскольку ключ может быть повторно сгенерирован, если сеанс прерван, запрос клиента направляется на тот же сервер, который он использовал ранее. Это полезно, если важно, чтобы клиент подключился к сеансу, который все еще активен после отключения.	Д
9	Random	Как следует из названия, этот алгоритм сопоставляет клиентов и серверы случайным образом, то есть с использованием базового генератора случайных чисел. В случаях, когда балансировщик нагрузки получает большое количество запросов, случайный алгоритм сможет равномерно распределить запросы по узлам. Как и Round Robin, случайного алгоритма достаточно для кластеров, состоящих из узлов с аналогичной конфигурацией (ЦП, ОЗУ и т. д.).	Д

Выводы

В статье представлен обзор механизма балансировки нагрузки в SDN. Решения на основе SDN используются для обеспечения балансировки нагрузки в сетях 5G / IMT-2020. Были исследованы архитектура SDN и методы балан-

сировки нагрузки, которые подразделяются на три группы: динамические, статические и гибридные. В статье упоминаются существенные качественные параметры балансировки нагрузки в SDN.

Мы попытались рассмотреть типы методов балансировки нагрузки на основе SDN, которые используются в балансировке нагрузки. Кроме того, мы сравнили два метода SLB и HSL и исследовали преимущества и недостатки обоих. Основываясь на результатах наших предыдущих разделов, механизмы балансировки нагрузки на основе SDN обеспечивают обзор сети, поэтому эти методы обычно улучшают производительность системы по сравнению с традиционными подходами балансировки нагрузки.

В следующих работах мы сосредоточимся на практической реализации и сравнении различных методов балансировки в области сети 5G, чтобы мы могли достичь лучшего метода балансировки нагрузки в сетях следующего поколения (5G) путем сравнения и комбинирования различных алгоритмов.

Литература

1. Toward the network of the future: From enabling technologies to 5G concepts / D. Bega, M. Gramaglia, C.J. Bernardos Cano, A. Banchs, X. Costa-Perez // Trans Emerging Tel Tech. 2017.
2. Subburayalu N., Natarajan S. and Das D. Dynamic Load Balancing across Multi-radio Access Bearers in 5G // 11th International Conference on Communication Systems Networks (COMSNETS). Bengaluru, India, 2019, Pp. 306-311.
3. Load Balancing Mechanisms in the Software Defined Networks: A Systematic and Comprehensive Review of the Literature / A.A. Neghabi, N. Jafari Navimipour, M. Hosseinzadeh and A. Rezaee // IEEE Access. 2018. Vol. 6. Pp. 14159-14178.
4. Survey of Promising Technologies for 5G Networks / Nam Tuan Le, Mohammad Arif Hossain, Amirul Islam, Doyun Kim, Young-June Choi, Yeong Min Jang // Mobile Information Systems. 2016. vol. 2016, Article ID 2676589. 25 p.
5. Open Networking Foundation (ONF). URL: <https://opennetworking.org>.
6. Nadeem Shah and Farik M. Static Load Balancing Algorithms In Cloud Computing: Challenges Solutions // International Journal of Scientific Technology Research. 2015. 4. Pp. 365-367.
7. Yadav A.V. and Mohammad F. Different Strategies for Load Balancing in Cloud Computing Environment: A Critical Study. 2014. Vol. 3. Is. 1. Pp. 85-90.
8. Milani A.S. and Jafari N. Load balancing mechanisms and techniques in the cloud environments Systematic literature review and future trends // NetwComput. Appl., 2016. Vol. 71. Pp. 86-98.
9. Chen S., Chen Y., and Kuo S. CLB: A novel load balancing architecture and algorithm for cloud services // Comput. Electr. Eng. 2017. Vol. 58. Pp. 154-160.

10. Mustafa Hasan Al.B., Nurul A.Z., Z. Zainal Abidin Load balancing algorithms in software defined network // International Journal of Recent Technology and Engineering (IJRTE), 2019. Vol. 7. Is. 6S5. Pp. 686-692.
11. Intelligent Load Balancing Techniques / S. Thabo, M. Thabiso, A. Stephen, K. Kefalotse, D. Setso, B. Gabanthone, S. Seth // Software Defined Networks: A Survey" Electronics 9. 2020. № 7: 1091.
12. Reliable and load balance-aware multi-controller deployment in SDN / T. Hu, P. Yi, J. Zhang and J. Lan // China Communications. 2018. vol. 15. № 11. Pp. 184-198.
13. Sridev K., Saifulla M.A. Multi Controller Load Balancing in Software Defined Networks: A Survey // Advances in Decision Sciences, Image Processing, Security and Computer Vision. Learning and Analytics in Intelligent Systems. Springer. Cham. Vol 3. Pp. 417-425.
14. Load-Balancing Multiple Controllers Mechanism for Software-Defined Networking / YW. Ma, JL. Chen, YH. Tsai, et al. Wireless // Pers Commun 94, 3549-3574(2017).
15. Lan Y., Wang K. and Hsu Y. Dynamic load-balanced path optimization in SDN-based data center networks // 10th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP), Prague, Czech Republic. 2016. Pp. 1-6.
16. Hopps C. RFC2992: Analysis of an Equal-Cost Multi-Path Algorithm. RFC Editor. USA. 2000.
17. An Effective Path Load Balancing Mechanism Based on SDN / J. Li, X. Chang, Y. Ren, Z. Zhang and G. Wang // IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications. Beijing. China, 2014. Pp. 527-533.
18. Cui Chen-xiao, and Xu Ya-bin Research on Load Balance Method in SDN // International Journal of Grid and Distributed Computing. 2016. Vol. 9. № 1. Pp. 25-36.
19. Pradana A.M., Purboyo T.W. and Latuconsina R. A simulation of load balancing in software defined network (SDN) based on Artificial Neural Networks method // ARPJ Journal of Engineering and Applied Sciences. 2020. Vol. 15. № 6.
20. Alex M.R. Ruelas and Christian Esteve Rothenberg. 2018. A Load Balancing Method based on Artificial Neural Networks for Knowledge-defined DataCenter Networking // Proceedings of the 10th Latin America Networking Conference (LANC '18). Association for Computing Machinery, New York, NY, USA, P. 106-109.
21. SDN Controller Load Balancing Based on Reinforcement Learning / Z. Li, X. Zhou, J. Gao and Y. Qin // IEEE 9th International Conference on Software Engineering and Service Science (ICSESS). Beijing. China. 2018. Pp. 1120-1126.
22. Yen L.H., Yeh T.T., Chi K.H. Load balancing // IEEE 802.11 networks. IEEE Internet Comput. 2009. № 13. Pp. 56-64.
23. Ghosh, Anish, Mrs. T. Manoranjitham. A study on load balancing techniques in SDN // International Journal of Engineering Technology [Online], 7.2.42018:174-177.

Поступила 10.12.2021; принята к публикации 15.02.2022

Информация об авторах

Бехруз Данешманд – аспирант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики (197101, Россия, г. Санкт-Петербург, Кронверкский пр., 49), e-mail: daneshmandbehrooz@gmail.com

Ту Ле Ань – аспирант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики (197101, Россия, г. Санкт-Петербург, Кронверкский пр., 49), e-mail: leanhtutcdt@gmail.com

STUDY AND REVIEW OF SDN-BASED LOAD BALANCING MECHANISMS IN 5G / IMT-2020

B. Daneshmand, L.A. Tu

Saint Petersburg State University of Information Technologies, Mechanics and Optics, St. Petersburg, Russia

Abstract: the growing number of mobile devices and the demand for user data by 2030 are expected to put pressure on the current mobile network in an unprecedented way. Future mobile networks must have several requirements regarding data amount, latency, quality of service and experience, mobility, spectrum, and energy efficiency. Therefore, efforts have recently begun for more efficient mobile network solutions. To this end, load balancing has attracted much attention as a promising solution for greater resource utilization, improved system performance, and reduced operating costs. This is an effective way to balance traffic and reduce congestion in heterogeneous networks in future 5G/IMT-2020 networks. Load Balancing is one of the most critical tasks required to maximize network performance, scalability, and robustness. Nowadays, with the emergence of Software-Defined Networking (SDN), Load Balancing for SDN has become a significant issue in future network 5G/IMT-2020. SDN allows for programmable load balancers and provides the flexibility to design and implement load balancing strategies. In this survey, we highlight the methods of load balancing based on SDN networks and prospective load balancing requirements on 5G networks

Key words: load balancing, software defined networking (SDN), 5G/IMT-2020

References

1. Bega D., Gramaglia M., Bernardos Cano C.J., Banchs A., Costa-Perez X. "Toward the network of the future: From enabling technologies to 5G concepts", *Trans Emerging Tel Tech.*, 2017.
2. Subburayalu N., Natarajan S., Das D. "Dynamic load balancing across multi-radio access bearers in 5G," *11th International Conference on Communication Systems Networks (COMSNETS)*, Bengaluru, India, 2019, pp. 306-311.
3. Neghabi A.A., Jafari Navimipour N., Hosseinzadeh M., Rezaee A. "Load balancing mechanisms in the software Defined networks: a systematic and comprehensive review of the literature," *IEEE Access*, 2018, vol. 6, pp. 14159-14178
4. Nam Tuan Le, Mohammad Arif Hossain, Amirul Islam, Do-yun Kim, Young-June Choi, Yeong Min Jang "Survey of promising technologies for 5G networks", *Mobile Information Systems*, 2016, vol. 2016, Article ID 2676589, pp. 25
5. Open Networking Foundation (ONF), available at: <https://opennetworking.org>.
6. Shah N., Farik M. "Static load balancing algorithms in cloud computing: challenges solutions", *International Journal of Scientific Technology Research*, 2015, vol. 4 (2015), pp. 365-367
7. Yadav A.V., Mohammad F. "Different strategies for load balancing in cloud computing environment: a critical study," 2014, vol. 3, issue 1, pp. 85-90
8. Milani A.S., Jafari N. "Load balancing mechanisms and techniques in the cloud environments Systematic literature review and future trends", *J. NetwComput. Appl.*, 2016, vol. 71, pp. 86-98
9. Chen S., Chen Y., Kuo S. "CLB: a novel load balancing architecture and algorithm for cloud services", *Comput. Electr. Eng.*, 2017, vol. 58, pp. 154-160.
10. Mustafa Hasan Al.B, Nurul A.Z, Zainal Abidin Z. "Load balancing algorithms in software defined network," *International Journal of Recent Technology and Engineering (IJRTE)*, April 2019, vol. 7, issue 6S5, pp.686-692.
11. Semong T., Maupong T., Anokye S., Kehulakae K., Dimakatso S., Boipelo G. Sarefo S. "Intelligent load balancing techniques in software Defined networks: a survey", *Electronics* 9, 2020, no. 7, pp. 1091.
12. Hu T., Yi P., Zhang J., Lan J. "Reliable and load balance-aware multi-controller deployment in SDN," *China Communications*, Nov. 2018, vol. 15, no. 11, pp. 184-198
13. Sridev K., Saifulla M.A. "Multi controller load balancing in software defined networks: a survey", *Advances in Decision Sciences, Image Processing, Security and Computer Vision. Learning and Analytics in Intelligent Systems*, vol 3. Springer, Cham, pp. 417-425
14. Ma Y.W., Chen J.L., Tsai Y.H. et al. "Load-balancing multiple controllers mechanism for software-Defined networking", *Wireless Pers Commun*, 2017, vol. 94, pp. 3549-3574.
15. Lan Y., Wang K., Hsu Y. "Dynamic load-balanced path optimization in SDN-based data center networks," *2016 10th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, Prague, Czech Republic, 2016, pp. 1-6
16. Hopps C. "RFC2992: Analysis of an Equal-Cost Multi-Path Algorithm", RFC Editor, USA, 2000.
17. Li J., Chang X., Ren Y., Zhang Z., Wang G. "An Effective path load balancing mechanism based on SDN", *2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications*, Beijing, China, 2014, pp. 527-533.
18. Cui Chen-xiao, Xu Ya-bin "Research on load balance method in SDN," *International Journal of Grid and Distributed Computing*, 2016, vol. 9, no. 1, pp. 25-36
19. Andika Malraherawan Pradana, Tito Waluyo Purboyo, Roswan Latuconsina "A simulation of load balancing in software defined network (SDN) based on Artificial Neural Networks method," *ARPN Journal of Engineering and Applied Sciences*, March 2020, vol. 15, no. 6
20. Ruelas A.M.R., Rothenberg C.E. "A load balancing method based on Artificial neural networks for Knowledge-defined data center networking", *Proc. of the 10th Latin America Networking Conference (LANC '18)*, Association for Computing Machinery, New York, USA, 2018, pp. 106-109.
21. Li Z., Zhou X., Gao J., Qin Y. "SDN controller load balancing based on reinforcement learning", *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China, 2018, pp. 1120-1126.
22. Yen L.H., Yeh T.T., Chi K.H. "Load balancing in IEEE 802.11 networks", *IEEE Internet Comput.* 2009, vol. 13, pp. 56-64
23. Anish G., Manoranjitham T. "A study on load balancing techniques in SDN", *International Journal of Engineering Technology*, 2018, no. 7.2.4, pp. 174-177.

Submitted 10.12.2021; revised 15.02.2022

Information about the authors

Behrooz Daneshmand, graduate student, St. Petersburg National Research University of Information Technologies, Mechanics and Optics (49 Kronverkskiy ave., St. Petersburg 197101, Russia), e-mail: daneshmandbehrooz@gmail.com
Le Anh Tu, graduate student, St. Petersburg National Research University of Information Technologies, Mechanics and Optics (49 Kronverkskiy ave., St. Petersburg 197101, Russia), e-mail: leanhtutcdt@gmail.com