مجله محاسبات نرم، جلد ۱۰، شماره ۱ (بهار و تابستان ۱۴۰۰)، صفحه ۳۲ـ۴۷
https://dx.doi.org/10.22052/scj.2022.243195.1005

مجله محاسبات نرم
SOFT COMPUTING JOURNAL
تارنمای مجله: scj.kashanu.ac.ir

دانشگاه کاشان
University of Kashan

ارائه روش حریصانه بهبود یافته برای افزایش تعداد کاربران سرویس داده شده در شبکه‌های ابر لبه✧

محمد شیرخانی[۱]، دانشجوی کارشناسی ارشد، کیهان خام فروش[۱*]، استادیار، مهسا ایزدبین[۱]، کارشناسی ارشد
[۱] گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی، واحد سنندج، سنندج، ایران.

اطلاعات مقاله

تاریخچه مقاله:
دریافت ۲۴ مهر ماه ۱۴۰۰
پذیرش ۲۸ بهمن ماه ۱۴۰۰

کلمات کلیدی:
تخصیص سرویس
زمان‌بندی کاربران
محاسبات لبه
شبکه ابری
الگوریتم حریصانه

چکیده

یکی از مهمترین چالش‌ها در محاسبات لبه، افزایش تعداد کاربران سرویس داده شده است بدون اینکه در پیچیدگی مساله تغییری حاصل شده و تاخیر بیشتری برای سرویس‌دهی به شبکه ابر لبه تحمیل شود. در ابر لبه، ابتدا تلاش می‌شود تا هر کاربر از سرور لبه خود، سرویس مورد نظرش را درخواست داده و این سرویس در اختیار وی قرار گیرد. در صورت عدم وجود سرویس در ابر لبه، از سرور لبه مجاور بر اساس فاصله، استعلام صورت می‌گیرد تا از نزدیک‌ترین ابرهای لبه که به صورت بی‌سیم می‌توانند با این ابر ارتباط داشته باشند، سرویس مورد نظر دریافت شود. اگر کاربر همچنان موفق به دریافت سرویس نشود، درخواست وارد پیوند بک‌هال شده و سرویس درخواستی در سایر سرورهای لبه جستجو می شود. در نهایت در صورتی که باز هم درخواست پاسخ داده نشود، سرویس از ابر دریافت می‌شود. استراتژی معرفی شده در مقاله، با توجه پیچیدگی بررسی شده، شرایطی را فراهم می‌کند که تعداد کاربران سرویس داده شده افزایش یابد بدون اینکه پیچیدگی تغییر کند. متوسط بهبود برای کاربران سرویس گرفته به کل کاربران در الگوریتم پیشنهادی نسبت به روش حریصانه ۰/۴ درصد و نسبت به روش بیشینه جریان ۱/۱ درصد است. همچنین این الگوریتم نسبت به الگوریتم بیشینه جریان و حریصانه در زمانبندی کاربران به ترتیب ۲۰٪ و ۲۲٪ بهبود را نشان می‌دهد.

© ۱۴۰۰ ـ مجلهٔ محاسبات نرم، کلیه حقوق محفوظ است.

۱. مقدمه

آنتن‌های مخابراتی به کاربرانی که تحت پوشش آن‌ها هستند خدمات و سرویس ارائه می‌دهند؛ بنابراین BTSها تمام تلاش خود را برای پوشش‌دهی کامل و با کیفیت در همه مناطق سلول انجام داده و راهکارهایی را برای سرویس‌دهی مناسب‌تر ارائه می‌دهند. به طور معمول در مناطق شلوغ مثل پایانه‌های اتوبوس، فرودگاه‌ها و راه‌آهن که ازدحام جمعیت وجود دارد به دلیل اینکه تعداد آنتن‌های برنامه‌ریزی شده به اندازه مورد نیاز نیست اختلالاتی مشاهده می‌شود که احتمالا می‌توان با اضافه کردن آنتن‌ها یا نزدیک‌تر کردن آن‌ها به هم، پوشش‌دهی مناسب‌تری انجام داد [۱].

شبکه زیرساخت از ارتباط BTS[1] ها با کاربران به صورت بی‌سیم و اتصال پرسرعت سیمی بین خودشان تشکیل می‌شود.

✧ نوع مقاله: پژوهشی
* نویسنده مسئول
پست(های) الکترونیک: hoba.114@gmail.com (شیرخانی)
k.khamforoosh@Iausdj.ac.ir (خام فروش)
m.izadbin@gmail.com (ایزدبین)

[1] Base Transceiver Station

نحوه ارجاع به مقاله: شیرخانی، محمد، خام فروش، کیهان، ایزدبین، مهسا، «ارائه روش حریصانه بهبود یافته برای افزایش تعداد کاربران سرویس داده شده در شبکه‌های ابر لبه»، مجله محاسبات نرم، جلد ۱۰، شماره ۱، ص۳۲ـ۴۷، بهار و تابستان ۱۴۰۰.

مجموعه BTS ها و کاربران تشکیل یک سلول را می‌دهند. در این سلول‌ها، BTS شامل سر رادیویی[1] و بخش پردازشی است و می‌تواند نقش یک سرور لبه را در ابر لبه ایفا کند. در هر سلول ابر لبه، سر رادیویی جهت اتصال با تجهیزاتی که دارای گیرنده رادیویی هستند، استفاده می‌شود و بخش پردازشی سلول متشکل از سرورهایی دارای واحد پردازش و داده برای پردازش درخواست‌های کاربران هر سلول است. از این سرورها به سرور لبه[2] تعبیر می‌شود و به سلولی که در آن سرور لبه قرار دارد، ابر لبه[3] می‌گویند. ابر لبه همانطور که ذکر شد برای کاهش بار پردازشی و ارتباطی ابر استفاده می‌شود. ابرهای لبه از طریق بک‌هال که نوعی اتصال زیرساخت سیمی است، ارتباط بین سلول‌ها، هماهنگ‌سازی و تبادل اطلاعات شبکه‌ای را برقرار می‌کنند. این اتصال، محدود به همین خدمات نیست و برای ارتباط فراسلولی کاربران نیز از آن استفاده می‌شود. البته ارتباط و تجزیه و تحلیل عملکردی از طریق بک‌هال، منجر به اختصاص پهنای باند بالا، تاخیر و ترافیک روی بک‌هال می‌شود [۱، ۲].

در حال حاضر در ابر لبه، هر دستگاه وظیفه پردازش اطلاعات خودش را دارد و تنها موارد مهم و اصلی را به سرور لبه ارسال می‌کند و این امکان را فراهم می‌نماید که داده‌ها به‌جای ارسال و پردازش در یک ابر بافاصله مکانی زیاد، در یک سرورلبه نزدیک پردازش شود [۱]. پردازش لبه در شبکه موبایل، مکمل پردازش ابری در ارائه سرویس به کاربران موبایل محسوب می‌شود. پردازش در لبه، نقاط ضعف پردازش ابری از جمله تاخیر زیاد، هزینه بالا و مسائل امنیتی و حریم شخصی را حل می‌کند [۲]. پیش از پیدایش ابر لبه، تمام دستگاه‌های هوشمند از فضای ابری جهت ذخیره‌سازی و پردازش اطلاعات خود استفاده می‌کردند و تمام اطلاعات را به ابرها ارسال و پس از پردازش توسط ابرها نتیجه نهایی را دریافت می‌نمودند. در این ساختار، حجم گسترده دستگاه‌های هوشمند و حجم زیاد اطلاعاتی که بین این دستگاه‌ها ردوبدل می‌شود باعث بروز مشکلات جایگذاری سرویس[4] و زمان‌بندی درخواست[5] در ابر لبه می‌شود که مساله بسیار مهمی است و در این مقاله به آن پرداخته می‌شود. بر اساس تقسیم‌بندی زیومی در مقاله [۳] تکنولوژی مورد بحث در این مقاله در شاخه Edge Computing و زیر شاخه Mobile edge Computing (MEC) قرار می‌گیرد. جایگذاری سرویس به معنی مشخص کردن نحوه چینش سرویس‌ها بر اساس ظرفیت هر ابر لبه است. جایگذاری به مفهوم قرارگیری داده‌های مربوطه و کدهای سرویس موردنظر در سرور لبه است. مساله از این جهت حائز اهمیت است که اولا ظرفیت ذخیره‌سازی سرورها محدود است و ثانیا سرویس گرفتن هر کاربر در هر سلول وابسته به نوع سرویسی است که جایگذاری می‌شود. به عبارت دیگر جایگذاری سرویس بر روی مساله زمان‌بندی درخواست‌ها تاثیرگذار است و می‌تواند نتایج آن را دستخوش تغییر کند.

از طرف دیگر، بر اساس آخرین دسته‌بندی توسط زیومی در مقاله [۳]، در مورد مطالعه اینترنت اشیا، مقاله پیش رو در دسته سرویس‌ها قرار می‌گیرد. مساله این مقاله، برنامه‌ریزی اتصال کاربران به هر لبه ابر و دریافت سرویس مورد درخواست آن‌ها است. ظرفیت اتصال به هر ابر لبه و همچنین پردازش درخواست‌ها دارای محدودیت ارتباطی و پردازشی است و وقتی پیچیده‌تر می‌شود که کاربران درخواست‌دهنده به نحوی زمان‌بندی شوند که علی‌رغم وجود تمام محدودیت‌ها، به سلولی وصل شوند که سرویس موردنظر آن‌ها در آن ارائه شده باشد. نوآوری مقاله نسبت به کارهای گذشته در ادامه بیان شده است.

- هر کاربر این امکان را دارد که از ابر لبه خود، سرویس مورد درخواست را دریافت کند و اگر در این ابر لبه موجود نباشد سرویس مورد نظر را با استعلام از BTS های مجاور موجود در محدوده مجاز (محدودیت مسافت) جستجو می‌کند و اگر بازهم موفق به دریافت سرویس نشد، درنهایت وارد بک‌هال شده و به جستجوی سرویس درخواستی در همه ابرهای لبه می‌پردازد. در این مقاله با

[1] Radio Head
[2] Edge server
[3] Edge cloud

[4] Service placement
[5] Request Scheduling

این استدلال که اولین کاندید برای اختصاص سرویس به کاربر در صورتی که در ابر لبه خودش موجود نباشد سرور ابر لبه مجاور است، روش پیشنهادی بدون اینکه کل پهنای باند بک‌هال را برای یافتن بهترین سرور ابر لبه در شبکه برای اتصال به کاربر موردنظر اشغال کند، درخواست را از سرورهای ابر لبه مجاور دریافت می‌کند.

- برای درخواست‌های کاربران (که بر اساس اولویت (۱) سلول، (۲) سلول مجاور، (۳) بک‌هال و (۴) ابر انجام می‌شوند)، روش پیشنهادی سعی می‌کند تا بهبود مناسبی در تعداد زمان‌بندی‌های کاربران که هدف دیگر مقاله است، ایجاد کند.
- محدودیت مسافت به‌عنوان یکی دیگر از مفروضات مساله به مقاله اضافه شده است. حداکثر فاصله بین کاربر و BTS به معنی این است که کاربر برای اتصال به BTS نیاز است در فاصله تعیین شده‌ای برای اتصال به سلول قرار داشته باشد و اولویت اتصال با کاربرانی است که فاصله کمتری تا BTS دارند. این فرض کمک می‌کند تا امکان اتصال به سرورهای لبه مجاور بررسی شود.

بخش‌های مختلف این مقاله به این شکل ساماندهی شده‌اند: بخش ۲ به بررسی کارهای مرتبط می‌پردازد. بخش ۳ با ارائه روش پیشنهادی به بررسی بهینه روش‌های تخصیص کاربر و جایگذاری سرویس‌های مورد درخواست آن‌ها می‌پردازد. در بخش ۴ ارزیابی و در بخش ۵ نتیجه‌گیری ارائه می‌شود.

۲. کارهای مرتبط

تحقیقات روی شبکه‌های ابر لبه تا جایی ادامه یافته که کاربران با چندین مرحله پرش می‌توانند به سرویس‌ها دسترسی داشته باشند [۶-۴]. در این تحقیقات هدف، رسیدن به الگوریتمی است که بتواند سرورهای لبه را جایگذاری و کاربران را به طور متناسب به آنها متصل کند. بنابراین همه کاربران می‌توانند با یک تاخیر معقول به آنها متصل شوند. این روش‌ها براساس چگالی گره‌های موبایل، سلول‌ها را تقسیم‌بندی کرده و سعی می‌کند تعادل بار را در هر ابر لبه فراهم کنند. به این ترتیب چگالی دستگاه‌های موجود در سلول برای ایجاد اعتدال در بار ترافیکی تقریبا یکسان خواهد بود.

با اجازه یافتن کاربران مبنی بر دسترسی به چندین سرور لبه، کارهایی در خصوص کاهش هزینه [۷]، زمان مصرفی [۸] و همچنین مدل‌های مختلف بار کاری و معماری شبکه‌های لبه انجام شده است. این پژوهش‌ها عموما بر اساس منابع مختص هر شبکه انجام شده و این منابع، عموما منابع غیرقابل اشتراک هستند. در مرجع [۹] محل‌های مناسب برای جایگذاری سرویس و منابع در ساختار توزیع شده ابری استفاده شده است و با برآورده کردن تکرار منابع و داده‌ها بنا به درخواست کاربر هزینه کلی جابه‌جایی شبکه کمینه شده است.

مساله جایگذاری سرویس و زمان‌بندی کاربران جزء مسائل NP-hard است، اگرچه در بسیاری از مقالات با حذف یا اضافه کردن قیودی به مساله مانند امکان سرویس دادن بدون محدودیت به علت وجود توان محاسباتی و ذخیره‌سازی بی‌نهایت در مساله، پیچیدگی ساده‌تری برای مساله ایجاد می‌شود. به‌عنوان مثال در مقاله [۱۰] فقط از محدودیت‌های ذخیره‌سازی و محاسباتی به عنوان قیود محدود کننده نام برده شده و از محدودیت فاصله و شبکه بحثی به میان نیامده است. در این مقاله از یک برنامه خطی عددی مختلط (MILP) استفاده شده که به‌طور مشترک جایگذاری و مسیریابی محتوای اشیا با توجه به محدودیت‌های ذخیره‌سازی و پردازشی را بهینه می‌کند. مقاله [۱۱] با حل مساله بهینه‌سازی با استفاده از ابرهای ناهمگن (متشکل از ابر لبه و فضای ابری با فاصله دور)، به انجام وظایف[1] حساس به تاخیر کمک می‌کند. همچنین در زمان بار ترافیکی زیاد، منابع محاسباتی به کاربرانی که موجب تاخیر کوتاه‌تری هستند، تخصیص داده می‌شود. بالعکس در زمان کم باری شبکه، منابع ابتدا در اختیار کارهای با تاخیر بالاتر قرار می‌گیرد. از آنجایی که سیاست بارگیری[2] مبتنی بر ابرهای ناهمگن نسبت به سیاست بارگیری مبتنی بر ابر لبه برتری دارد، در نتیجه تاخیر بار کمتری حاصل می‌شود.

[1] Task
[2] Offload

در مقاله [۱۲] قیود پردازشی و پهنای باند شبکه دخیل شده ولی محدودیت‌های مسافت و ذخیره‌سازی مد نظر نبوده‌اند. تغییر هرکدام از این متغیرها پیچیدگی مساله را تغییر می‌دهد. این مقاله برای افزایش کیفیت سرویس (QoS) اقدام به معرفی چارچوبی برای بارگیری فرایندها جهت کاهش بار پردازشی و تخصیص بهینه منابع شبکه نموده است. برای این منظور برای هر کاربر به‌صورت مجزا مساله بارگیری محاسبات حل می‌شود که به استفاده از منابع شبکه و پردازشی بهینه منتج شده است. این امر باعث کاهش اشغال منابع شده و در عین حال محدودیت تاخیر را نیز برآورده می‌کند. با فرض بهینه بودن منابع ذکر شده می‌توان گفت فاکتور تاخیر، کاهش یافته و موجب افزایش کیفیت سرویس خواهد شد. روش‌های پیشنهادی این مقاله می‌توانند به عملکرد بهتر ابر لبه بر اساس درخواست‌های ایجاد شده در لحظه کمک کنند.

همان‌طور که ذکرشده موضوع اصلی این مقاله بر اساس تخصیص سرویس‌ها و زمان‌بندی کاربران است. در دسته‌ای از مقالات طرح مساله مشابه USSA بوده ولی هدف اصلی آن کاهش تاخیر است و بحث‌های سیگنالینگ و تخصیص کانال در آنها مطرح می‌شوند.

در مقاله [۱۳] ابتدا با استفاده از الگوریتم بیشینه جریان مساله زمان‌بندی کاربران حل شده که در قیاس با برنامه عددی خطی (LIP) نتیجه بهتری خواهد داشت. همچنین از محدودیت مسافت بحثی نشده و سه قید شبکه، محاسبات و ذخیره‌سازی در نظر گرفته شده و بررسی با دو الگوریتم متفاوت بیشینه جریان و حریصانه انجام شده است. روش پیشنهادی نیز با این مقاله مقایسه شده است.

مقاله [۱۴] در اصل توسعه مقاله [۱۳] بوده که توسط نویسندگان مشابه قبلی نوشته شده است. در مقاله قبل فرض بر این بود که یک بازه زمانی به عنوان یک فریم در نظر گرفته شده و این فریم چندین بار تکرار می‌گردد، ولی در مقاله [۱۴] این بازه به عنوان مثال یک بازه ۲۴ ساعته در نظر گرفته می‌شود. به این صورت هدر رفت منابع بین بازه فریم‌ها از بین رفته و به‌صورت بهینه‌تر به بررسی ابعاد مختلف چالش پرداخت خواهد شد.

در مقاله [۱۵] مساله بارگیری بار پردازشی مشترک برای شبکه‌های وسایل نقلیه که در آن ابر لبه متحرک[1] و محاسبات ابر[2] مشارکت دارند فرموله شده است. برای به حداکثر رساندن منفعت سیستم، بارگیری و تخصیص منابع محاسباتی بهینه‌سازی شده است. به این منظور طرحی پیشنهاد شده است که باعث می‌شود پیچیدگی سیستم بدون از دست دادن عملکرد آن کاهش یابد. این مزیت وقتی آشکار می‌شود که منابع محاسباتی استفاده شده سرورهای ابر لبه متحرک کاهش و تعداد وظایف انجام شده پرحجم افزایش و یا هزینه واحد برای منبع محاسبه کاهش می‌یابد.

در مقاله [۱۶] نیز به‌صورت همزمان (در طول مجموعه ابرهای لبه) به زمان‌بندی کاربران و جایگذاری سرویس پرداخته شده است. هر دو مساله ضرب‌الاجل کاربران و رفتار استراتژیک ابرهای لبه بررسی شده است. با در نظر گرفتن بازه زمانی مشخص شده از سوی کاربر حساس به تاخیر و با استفاده از تکنیک لیاپانوف[3] جهت بررسی پایداری شبکه اقدام شده است. سپس به بهره‌برداری از فضای ابرهای لبه که به‌صورت بهینه هر دو چالش جایگذاری و زمان‌بندی را حل نموده پرداخته شده است. در این مقاله همچنین کاهش هزینه و توازن بار مطرح بوده است.

پایین آوردن بار ترافیکی روی بک‌هال در مقالات مختلفی مورد توجه قرار گرفته است و راه‌کارهای متفاوتی به این منظور ارائه شده است. حتی در برخی پژوهش‌ها سعی شده تا با یک تعادل بین مصرف انرژی کاربران و بارگذاری بار از بک‌هال برای کاربر، بار ترافیک بک‌هال کاهش یابد [۱۷]. همچنین بسیاری از مقالات مانند [۱۸] و [۱۹] به بررسی تاثیر زمان‌بندی کاربران و وظایف به‌منظور کاهش تاخیر در ارسال اطلاعات پرداخته‌اند، در حالی که تعداد کاربران زمان‌بندی شده در آنها مورد بحث قرار نگرفته است. هرچند در این مقالات برای کاهش تاخیر تلاش‌های خوبی (نظیر کاهش بار بک‌هال و ارتقای عملکرد الگوریتم) انجام شده، اما به موضوع افزایش تعداد کاربران

[1] Mobile edge cloud
[2] Cloud computing
[3] Liapanov

زمان‌بندی شده پرداخته نشده است.

در مقاله [۲۰] تخصیص منابع و زمان‌بندی کاربران به وسیله الگوریتم‌های یادگیری ماشین TD3 و بهینه‌سازی مصرف انرژی و تاخیر با استفاده از الگوریتم مارکوف در شبکه‌های بی‌سیم انجام شده است.

وانگ و همکارانش در مقاله [۲۱] نیز سعی در کاهش تاخیر و بررسی توان انتقال کاربران در شبکه MEC داشتند که روش آنها زمان‌بندی کاربران و اختصاص سرویس‌ها از طریق سوابق قبلی (یادگیری ماشین) و انتخاب این سوابق بر پایه الگوریتم حریصانه را انجام می‌دهد. بنابراین بیشتر بحث و بررسی مقاله مربوط به کاهش تاخیر از طریق بهینه‌سازی بوده و در خصوص چینش سرویس‌ها صرفا به سوابق قبلی بسنده کرده‌اند.

مقاله [۲۲] در خصوص ارائه یک استراتژی برای اختصاص منابع براساس شبکه عصبی در شبکه تجهیزات سلامت می‌باشد. در این مقاله که برای کاهش تاخیر از معماری پردازش لبه استفاده شده است، هدف اصلی بهینه‌سازی تاخیر و مصرف انرژی در شبکه است که طبق محاسبات و نمودارها حاصل شده است. به هر حال افزایش تعداد سرورهای لبه موجب کاهش تاخیر در شبکه خواهد شد.

محاسبات لبه یک الگوی محاسباتی نوظهور است که منابع محاسباتی و ذخیره‌سازی را به لبه شبکه می‌آورد و از این رو تأخیر سرویس و ترافیک شبکه را به میزان قابل توجهی کاهش می‌دهد. در محاسبات لبه، بسیاری از برنامه‌ها از وظایف وابسته تشکیل شده‌اند که خروجی‌های برخی، ورودی‌های برخی دیگر است. نحوه بارگذاری این وظایف در لبه شبکه یک مشکل حیاتی و چالش برانگیز است که هدف آن تعیین مکان هر وظیفه در حال اجرا به منظور به حداکثر رساندن کیفیت خدمات است. این مقاله الگوریتم اکتشافی طراحی می‌کند که سازگاری قوی یا روش‌های مبتنی بر یادگیری دارد. اما بدون در نظر گرفتن وابستگی ذاتی کار، یک طرح تخلیه کار هوشمند پیشنهاد شده و توسط یک شبکه عصبی Sequence-to-Sequence (S2S) بهبود داده شده است [۲۵].

جدول (۱) خلاصه موارد ذکر شده را بیان می‌کند.

۳. روش پیشنهادی

با مشخص بودن مکان کاربران، آنتن‌های BTS که در این مقاله همان سرورهای ابر لبه هستند باید سرویس‌های مورد درخواست کاربران را با استفاده از الگوریتم حریصانه و بهره‌گیری از توزیع زیف به سرورهای ابر لبه موجود اختصاص دهند. اختصاص سرویس‌ها به هر BTS دارای محدودیت در تعداد است. لذا با در نظر گرفتن این محدودیت، به زمان‌بندی کاربران جهت دریافت خدمات از سرویس‌های موجود در سرورهای ابر لبه پرداخته می‌شود. باید در نظر داشت این محدودیت به صورت مستقیم با تعداد کاربران زمان‌بندی شده در ارتباط است. چرا که در صورت عدم وجود سرویس در سرور ابر لبه متناظر کاربر، در عمل ارتباط کاربر قطع شده و توانایی زمان‌بندی سرویس وجود نخواهد داشت.

ازجمله مزایای این مقاله که پیش‌تر به آن اشاره شد، اعمال محدودیت مسافت است که برای اتصال کاربران به سرور ابر لبه که همان BTS است لحاظ شده و به معنی حداکثر فاصله بین کاربر و سرور ابر لبه است. برای اتصال به سرور نیاز است در فاصله تعیین‌شده به سلول متصل شده و اولویت اتصال با کاربرانی است که فاصله کمتری تا سرور ابر لبه دارند. شکل (۱) معماری شبکه پیشنهادی را که به عنوان الگوریتم USSA نامگذاری شده است، با انجام مراحل یک تا چهار نشان می‌دهد. در مرحله ۱ در داخل سلول پاسخ سرویس جستجو می‌شود، در مرحله دوم از سلول‌های مجاوری که در بعد آن قرار دارند سوال می‌شود، در مرحله سوم یافتن سرویس از بک‌هال هدف قرار داده می‌شود و در گام چهارم درخواست به سراغ ابر خواهد رفت. تفاوت این معماری با کارهای انجام شده قبلی استفاده از متغیر مسافت و دادن این شانس به سرویس است که قبل از اینکه درخواست را به بک‌هال منتقل نماید آن را به سلول‌های مجاوری که در بعد دستگاه فرستنده قرار دارند، ارسال کند.

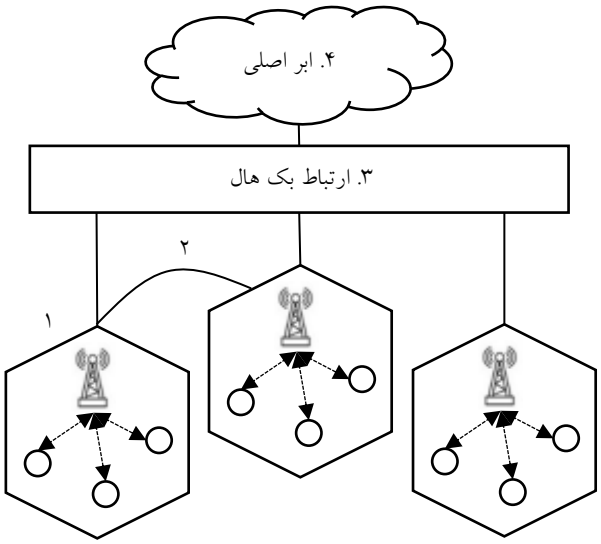
جدول (۱) – مقایسه روش‌های قبلی

مرجع	سال	ویژگی‌ها	تکنیک مورد استفاده برای حل مساله	شبکه مورد استفاده	مزایا	معایب
[۴]	۲۰۱۵	استفاده از هر دو نوع منابع قابل اشتراک و غیرقابل اشتراک در مسئله توزیع‌شده سرویس/ محتوا	استفاده از برنامه خطی عددی برای جایگذاری سرویس و به‌کارگیری منابع ذخیره‌سازی، محاسباتی و ارتباطی در یک شبکه ابری توزیع‌شده فرموله شده است.	شبکه ابری توزیع‌شده	به حداقل رساندن هزینه کلی شبکه. تبدیل شدن رابطه بهینه‌سازی به یک معادله خطی قابل حل.	به علت ساختار شبکه برای برنامه‌های حساس به تاخیر کاربردی نیست. عدم در نظر گرفتن هزینه‌های وابسته به بار برای راحتی مساله. عدم بررسی فرمول در شبکه‌های ناهمگن
[۱۰]	۲۰۱۷	استفاده از هر دو نوع منابع قابل اشتراک و غیرقابل اشتراک در مسئله توزیع‌شده سرویس یا محتوا	استفاده از یک MILP مشابه که به‌طور اشتراکی شامل منابع ذخیره‌سازی و محاسباتی شده است.	شبکه ابری توزیع‌شده	بهینه شدن جایگذاری و مسیریابی محتوا در شبکه. در نظر گرفتن هزینه‌های فعال‌سازی و عملیاتی بارهای کاربران. تحویل یکپارچه یا چند بخشی پاسخ به کاربران.	در خصوص زمان‌بندی کاربران بحثی نشده است. در خصوص محدودیت شبکه در محاسبات لحاظ نشده است.
[۱۱]	۲۰۱۷	استفاده از شبکه‌های ابری ناهمگن متشکل از EC و Remote Cloud. زمان‌بندی EC ها برای کاهش تاخیر برنامه‌های حساس به تاخیر.	با استفاده از مساله بهینه‌سازی به بهبود QoS، تاخیر، بارگذاری به روش مشتق‌گیری (بهینه‌سازی مقعر) و شبیه‌سازی عددی حل شده است.	شبکه ابری لبه متحرک	حل مساله با استفاده از روش بهینه‌سازی مقعر. سیاست متغیر زمان‌بندی در زمان ترافیک بالا و ترافیک پایین کاربران. بررسی ۳ نوع سیاست زمان‌بندی و نتیجه‌گیری نهایی آن‌ها.	بررسی مساله فقط در شبکه ناهمگن. عدم در نظر گرفتن پارامتر مسافت
[۱۲]	۲۰۱۸	معرفی Framework وسیع شامل مکانیزم بارگذاری بهینه برای تخصیص منابع ارتباطی و پردازشی	بارگذاری بهینه منابع بررسی می‌شود که بر اساس آن با اختصاص منابع ارتباطی و بارگذاری بار پردازشی موجب افزایش Qos می‌گردد. این مساله به‌وسیله یک راه‌حل تقریبی بهینه مورد بررسی قرار می‌گیرد	شبکه ابری لبه متحرک	استفاده از روشی منطبق بر اصل لانه کبوتری در زمان‌بندی درخواست کاربران. رتبه‌بندی کاربران برای اولویت‌بندی تخصیص منابع	عدم دخیل شدن قیود ذخیره‌سازی و مسافت در محاسبات
[۱۳]	۲۰۱۸	منابع قابل اشتراک و غیرقابل اشتراک با در نظر گرفتن محدودیت‌های شبکه، محاسباتی و ذخیره‌سازی	با استفاده از یک برنامه خطی عددی (ILP) و یکی از روش‌های بهینه‌سازی (روش بیشینه جریان) حل شده است.	شبکه ابری و شبکه ابری لبه متحرک مشترکا	در نظر گرفتن سه قید شبکه، محاسباتی و ذخیره‌سازی. بررسی از دو الگوریتم متفاوت بیشینه جریان و حریصانه.	افزایش بار کاری روی یک‌هال. عدم در نظر گرفتن قید مسافت.
[۱۴]	۲۰۱۹	در نظر گرفتن بازه زمانی بزرگتر از یک فریم	جایگذاری سرویس با استفاده از الگوریتم بیشینه جریان و زمان‌بندی کاربران با استفاده از الگوریتم حریصانه انجام شده است.	شبکه ابری و شبکه ابر لبه	بهینه شدن استفاده از منابع به دلیل بررسی بیشتر از یک فریم زمانی	برای ساده‌سازی مساله NP سخت فرض شده استفاده کاربران از منابع یکسان است. عدم در نظر گرفتن محدودکننده مسافت

ادامه جدول (۱) – مقایسه روش‌های قبلی

مرجع	سال	ویژگی‌ها	تکنیک مورد استفاده برای حل مساله	شبکه مورد استفاده	مزایا	معایب
[۱۵]	۲۰۱۹	با بارگذاری محاسبات روی کلود و تخصیص منابع (CCORAO) سعی در بهینه‌سازی شبکه مرزی خودروها دارد.	بر اساس سه فاکتور تاخیر پردازش، هزینه منابع محاسباتی و عامل نرمال‌سازی، در مرحله اول الگوریتم‌های استراتژی بارگذاری و تخصیص منابع محاسباتی و سپس با استفاده از یک الگوریتم توزیع‌شده پاسخ نهایی مساله حاصل می‌شود.	شبکه ابری و شبکه ابری لبه متحرک مشترکا	این الگوریتم مناسب منابع محاسباتی کافی ندارند؛ می‌باشد.	عدم در نظر گرفتن قید ذخیره‌سازی و شبکه.
[۱۶]	۲۰۲۱	جهت زمان‌بندی کاربران و جایگذاری سرویس از ابرهای لبه استفاده‌شده و این مسائل بر اساس تابع هزینه و توازن بار زمان‌بندی شده است. تابع هزینه بیشتر بر اساس زمان اتمام کار بوده و عدم زمان‌بندی یک کاربر در ابر لبه به دلیل زمان بالای پاسخگویی آن می‌باشد.	کاربران را با دو فرض حساس به تاخیر و غیر حساس دسته‌بندی نموده است. کاربران حساس به تاخیر با تکنیک لیاپانوف و کاربران غیر حساس به تاخیر با الگوریتم ابداعی DO-JCPS (الگوریتم تقریبی) در چند مرحله زمانی تا زمان پایان کار زمان‌بندی شده است.	شبکه ابری و شبکه ابر لبه مشترکا	لحاظ نمودن تابع هزینه و توازن بار. بررسی پایداری شبکه ابر لبه با استفاده از تکنیک لیاپانوف	مسائل مربوط به کاربران حساس به تاخیر در یک مرحله زمانی بررسی شده است. از هیچ قید و محدودیتی استفاده‌نشده است و فقط هزینه زیاد موجب عدم زمان‌بندی است.
[۲۰]	۲۰۲۰	در این مقاله از الگوریتم‌های یادگیری ماشین برای زمان‌بندی کاربران و اختصاص سرویس استفاده شده و در آن به بحث تاخیر و مصرف انرژی پرداخته شده است.	استفاده از الگوریتم TD3 برای حل مساله و مارکوف برای بهینه‌سازی	شبکه‌های بیسیم	بدست آمدن تاخیر پایین‌تر با استفاده از سیاست اتخاذ شده	حل مساله در بستر شبکه‌های وایرلس
[۲۱]	۲۰۲۰	استفاده از فن‌های بهینه‌سازی برای کاهش تاخیر و استفاده از سوابق قبلی برای تخصیص کاربران و سرویس‌ها استفاده شده است.	استفاده از الگوریتم بهینه‌سازی غیرخطی برای کاهش تاخیر و الگوریتم یادگیری ماشین تقویت‌شده برای تخصیص کاربران و سرویس‌ها	شبکه ابری لبه متحرک	سعی کمتر در مراحل بهینه‌سازی با کمک یادگیری ماشین	عدم بررسی محدودیت‌های تخصی منابع و کاربران و صرفا استفاده از آرشیو سوابق و تاریخچه اطلاعات قبلی.
[۲۲]	۲۰۲۱	بررسی شبکه اینترنت اشیای سلامت IOHT و بهینه‌سازی تاخیر و مصرف انرژی با شبکه عصی	استفاده از ترکیب الگوریتم‌های شبکه عصبی و یادگیری ماشین	شبکه لبه سلامت محور	قابلیت بسط پژوهش و کارامد بودن روش ارائه شده در کاربران بالاتر	عدم اشاره به نحوه تخصیص سرویس‌ها و زمان‌بندی کاربران

همان‌طور که مشاهده می‌شود در روش پیشنهادی تلاش می‌شود لینک بک‌هال تا حد ممکن اشغال نشود زیرا اشغال بیش از حد آن باعث پایین آمدن کارایی سیستم و تاخیر در عملکردهای حیاتی آن می‌شود [۲۳]. هر کاربر این امکان را دارد که از ابر لبه خود سرویس مورد درخواست را بگیرد و اگر در آن موجود نباشد سرویس مورد نظر را با استعلام از سرورهای ابر لبه مجاور موجود در محدوده مجاز (محدودیت مسافت) جستجو می‌کند و اگر بازهم موفق به دریافت سرویس نشد، درنهایت وارد بک‌هال شده و به جستجوی سرویس درخواستی در سایر ابرهای لبه می‌پردازد که این دسترسی باید کمینه باشد زیرا که افزایش بار روی بک‌هال از آن جهت حائز اهمیت است که می‌تواند عملکرد سرویس‌دهی و ارتباطات شبکه را دچار آسیب کند.



شکل (۱). ابرهای لبه تشکیل شده از مجموعه BTS ها در سلول‌های مختلف [۱۳] در شکل درخواست از همسایه با خط سبز مشخص شده است

در الگوریتم پیشنهادی، برای زمان‌بندی کاربران چند مرحله طی می‌شود تا طی این مراحل یک‌به‌یک کاربران متصل شوند. برای این منظور کاربران به دو دسته تقسیم می‌شوند: (۱) کاربرانی که درخواست سرویس ندارند و فقط به سلول متصل می‌شوند تا به محض درخواست سرویس، درخواست آن‌ها در اسنپ‌شات بعدی مورد بررسی قرار گیرد و (۲) کاربران متقاضی سرویس که بر اساس نوع سرویس به سه زیر مجموعه دیگر تقسیم می‌شوند. (۲-۱) کاربرانی که سرویس آن‌ها در نزدیک‌ترین سرور ابر لبه قرار دارد، (۲-۲) کاربرانی که سرویس آن‌ها در سرورهای ابر لبه مجاور قرار دارد و (۲-۳) کاربرانی که سرویس آن‌ها در هیچ یک از سرویس‌های ابر لبه قرار ندارد. هر یک از این کاربران در ابتدا برای اتصال به هر سرور ابر لبه کاندید می‌شوند. سپس بر اساس نزدیکی به آن و با قید محدودیت‌های ارتباطی و پردازشی به سرور مورد نظر متصل می‌شوند.

کاربرانی که درخواستی برای سرویس ندارند، بررسی دیگری نداشته و صرفا بر اساس بعد مسافت برای اتصال به سرور ابر لبه نزدیک‌تر کاندید می‌شوند. اما کاربرانی که نیازمند دریافت سرویس هستند وضعیت متفاوتی دارند. کاربرانی که سرویس مورد نظر آن‌ها بر اساس جایگذاری سرویس در نزدیک‌ترین سرور ابر لبه وجود دارد، در اولین مرحله کاندید اتصال به نزدیک‌ترین سرور می‌شوند. هدف از این بخش نسبت دادن بیشترین تعداد کاربران به هر سلول برای جلوگیری از اشغال شدن بک‌هال است که لازمه این کار، قرارگیری هر سرویس در بهترین مکان ممکن است. بنابراین با اعمال محدودیت‌هایی، جایگذاری سرویس‌ها و پاسخ‌دهی بهینه به درخواست کاربران انجام می‌شود.

مساله جایگذاری سرویس و زمان‌بندی کاربران، یک مساله بهینه‌سازی NP-Hard است. یکی از الگوریتم‌های حل آن، بیشینه جریان می‌باشد که با مدل کردن اتصال کاربران به ابرهای لبه و سرویس‌گیری آنها به شکل یک گراف، با استفاده از روش فورد فالکرسون[1] حل می‌شود. در این روش، ابتدا جایگذاری سرویس‌ها با استفاده از الگوریتم حریصانه انجام شده و سپس با استفاده از الگوریتم بیشینه جریان، زمان‌بندی کاربران انجام می‌شود. در اولین مرحله جایگذاری سرویس، از الگوریتم حریصانه استفاده می‌شود که شبه کد آن در الگوریتم (۱) مشاهده می‌شود. برای جایگذاری سرویس، مقاله [۱۲] به خاطر پیچیدگی کمتر از بین روش‌های موجود حل، الگوریتم حریصانه را پیشنهاد می‌دهد و با توجه به اینکه الگوریتم پیشنهادی در این

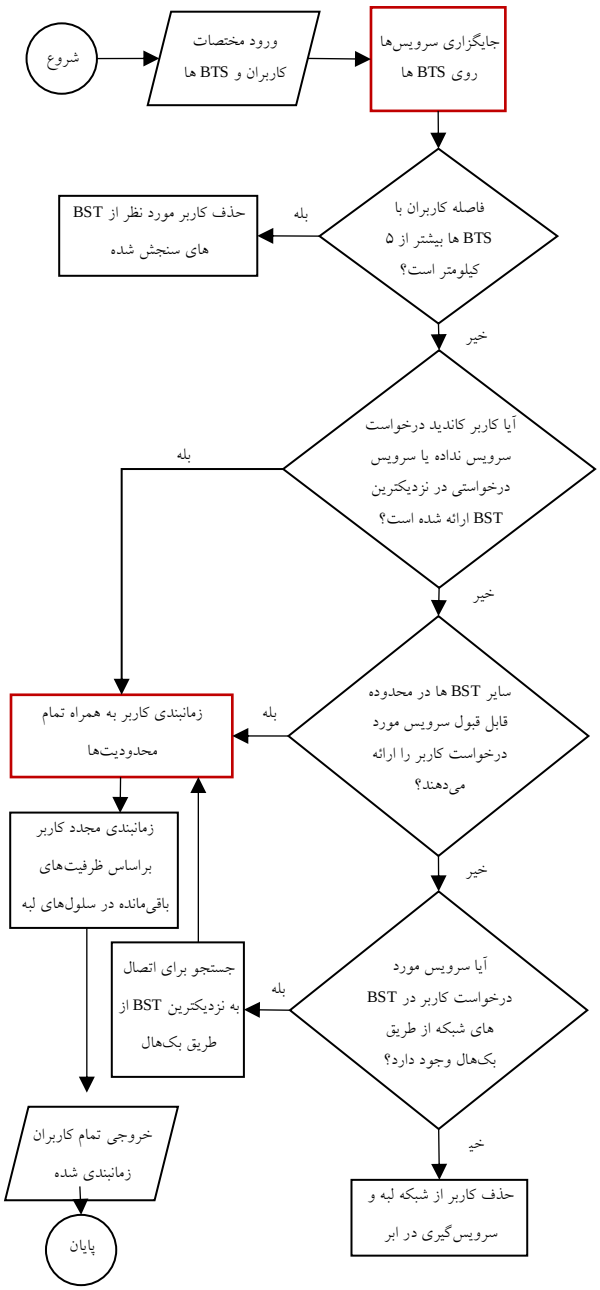
[1] Ford Folkerson

مقاله، پیش از مراجعه به بک‌هال از همسایه‌های مجاور ابر لبه برای سرویس‌دهی استفاده می‌کند، تعداد کاربران سرویس داده شده را افزایش داده و مطابق شبه کد الگوریتم (۱) تنها مقدار ثابتی به پیچیدگی اضافه می‌کند که در تغییر مقدار بزرگی زمانی نقشی ندارد. خروجی این کد ماتریس جایگذاری شده سرویس‌ها است که مشخص می‌کند در هر ابر لبه کدام سرویس جایگذاری شده است. بنابراین با داشتن وضعیت هر ابر لبه و مشخص بودن کاربران کاندید برای آنها می‌توان باقیمانده عملیات را با استفاده از مساله بیشینه جریان حل نمود.

الگوریتم (۱)- جایگذاری سرویس با استفاده از الگوریتم حریصانه

```
/*The algorithm has input is “n” Edge Clouds and “s” Services This Services will placed in each edge cloud Servers.*/
00 Input: service_n , EC
/*The “SP” Matrix is n×s and Each matrix element is determined that each Service assign to which EC*/
01 SP_n← Ø
02 SE ← {service_1, service_2,...}
03 Foreach EC_t
04   For i=1:count(SE)
/*Check for Storage Constraint, if violation occurred the loop would break*/
05     If count(SP_n)<S
/*Check for repetitive Data*/
06       If service_i exist in SP_n
07         Continue
08       end
09       SP_n== SPFunction(service_i,)
10     Else
11       Break
12     end
13   end
14 end
15 Output: SP matrix t×S
/*The Placement of Services in each EC was apply with Greedy Algorithm*/
```

برای حل مساله بیشینه جریان از الگوریتم فورد-فالکرسون استفاده می‌شود. به این ترتیب برای حل مساله زمان‌بندی کاربران نیاز است اتصال کاربران به سلول‌های ابر لبه و جایگذاری سرویس‌ها، مجموعا روی یک گراف ترسیم شود. این گراف با استفاده از الگوریتم فورد-فالکرسون حل شده و حاصل آن بیشینه کاربران زمان‌بندی شده در هر ابر لبه را نشان می‌دهد. استفاده از روش بیشینه جریان برای حل این مساله با تقریب ½ به درخواست‌ها پاسخ می‌دهد که به معنی زمان‌بندی شدن نصف کاربران موجود در تمام ابرهای لبه است. راه‌حل جایگزین در این مقاله، استفاده از الگوریتم حریصانه است [۲۴]. نمودار شکل (۲) و الگوریتم (۲) به جزئیات روش ابداعی مقاله پرداخته است. به این ترتیب که پس از جایگذاری سرویس‌ها، با استفاده مجدد از الگوریتم حریصانه می‌توان کاربران را در هر سلول جایگذاری نمود. به دلیل انعطاف این روش می‌توان پس از زمان‌بندی کاربران، باقی‌مانده کاربران را با توجه به ظرفیت‌های باقی‌مانده هر ابر لبه مجددا زمان‌بندی نمود.



شکل (۲)- نمودار روش پیشنهادی

الگوریتم (۲) با گرفتن ورودی‌ها که موقعیت کاربران، BTS ها و ماتریس جایگذاری سرویس‌هایی است که از الگوریتم (۱) به دست آمده است، آغاز می‌شود. برای هر کاربری که درخواست سرویس دارد به‌وسیله توزیع زیف، سرویس درخواست شده اختصاص داده می‌شود. محدودیت مسافت کاربر تا BTS با عبارت Range مشخص شده است. اگر فاصله هر کاربر کمتر از Range باشد و همچنین سرویس مورد درخواست کاربر در ابر لبه مربوطه موجود باشد، آن کاربر به ابر لبه متصل می‌شود، در غیر این صورت کاربران باقی‌مانده که سرویس مورد درخواست آنها در ابر لبه موجود نیست، به‌عنوان کاربران رزرو شده طبقه‌بندی می‌شوند.

الگوریتم (۲) –شبه کد الگوریتم پیشنهادی با در نظر گرفتن محدودیت‌های شبکه، پردازشی و مسافت با استفاده از الگوریتم حریصانه

```
00 Input Users Place, BTS Place, SP Matrix ;
01 Foreach User_i do
02   If Request_{User_i} == 1
03     User_i <-Assign- service_i ;
04   End
05 End
06 Dis = Distance_Fun(User Place,BTS Place) ;
07 Foreach User_i , EC_t do
08   If Dis < Range
09     If Service_{User_i} == Service_{EC_t}
10       EC_t <-Assign- User_i ;
11     Else
12       Reserved_User ← User_i ;
13     End
14   End
15 End
16 Foreach Reserved_{User} EC_j
     that Service_{Reserved_User} == Service_{EC_j} do
17   If Dis_{Reserve_User} <Range
18     EC_j <-Assign- Reserved_User ;
19   End
20 End
21 Sort EC_t by Dis ;
22 Scheduled_Users ← Select Top N rows From EC;
23 If Service_{Scheduled_Users} > P
24   Remove Extra Scheduled_Users;
25 End
```

کاربران رزرو شده مجددا نزدیک‌ترین سرور ابر لبه موجود که در آن سرویس مورد نظر ارائه شده است، را پیدا و به آن متصل می‌شوند. در صورتی که در سرورهای ابر لبه نزدیک، سرویس مورد نظر وجود نداشته باشد، با اتصال به نزدیک‌ترین آنها، از طریق بک‌هال برای یافتن سرویس مورد نظر اقدام می‌نمایند.

در مرحله بعدی کاربران اختصاص یافته به هر ابر لبه، بر اساس محدودیت مسافت و شبکه، مرتب شده و برای اتصال انتخاب می‌شوند. در این مرحله ممکن است تعدادی از کاربران بیشتر از ظرفیت پردازشی، درخواست سرویس داشته باشند که لازم است کاربران مازاد بر ظرفیت پردازش از کاربران زمان‌بندی شده حذف شوند.

در این مقاله، با اعمال محدودیت مسافت دقت در پاسخ الگوریتم بهتر می‌شود. اگرچه اضافه کردن این متغیر ممکن است در ابتدا موجب شود کاربران کمتری زمان‌بندی شوند؛ ولی با به‌کارگیری زمان‌بندی‌های متمادی این محدودیت برطرف می‌شود. لازم به ذکر است که ظرفیت‌های خالی موجود در ابرهای لبه ناشی از عدم جایگذاری مناسب سرورهای ابر لبه است و به عبارت دیگر کاربر دیگری که نیازمند سرویس باشد نمی‌تواند به آن متصل شود.

برای مشاهده کدهای این مقاله می توانید به پیوند https://github.com/tanoo90/USSAPaper مراجعه کنید.

۴. ارزیابی کارایی

۴,۱. تنظیمات شبیه سازی

دیتاست انتخاب شده در این مقاله تاکسی‌های شهر سانفرانسیسکو در سال ۲۰۰۸ است که موقعیت مکانی آنها در مختصات دسیمال جغرافیایی به همراه زمان و وجود مسافر ثبت شده است. موقعیت هر تاکسی به طور متوسط هر ۱ دقیقه به‌روزرسانی می‌شود، بنابراین تاکسی‌هایی که بیش از ۵ دقیقه موقعیت خود را به‌روز نکرده باشند، در بازه زمانی مورد بررسی، از مجموعه کاربران حذف می‌شوند. محدودیت دامنه ۵ کیلومتری در مقاله لحاظ شده است [۲۱] و برد آنتن‌های BTS در نسل 4G بین ۷۰۰ تا ۹۰۰ متر در مناطق شهری متراکم است که در کنار آن ظرفیت BTS ها و تعداد کاربران نیز افزایش پیدا کرده است. برای مکان‌یابی BTS ها نیز از موقعیت آنتن‌ها در

سایت www.antennasearch.com استفاده شده است. برای پیاده‌سازی روش پیشنهادی از نرم‌افزار Matlab 2016b استفاده شده و متغیرهای آن در جدول (۲) آورده شده است.

جدول (۲)- متغیرهای استفاده‌شده در پیاده‌سازی مساله

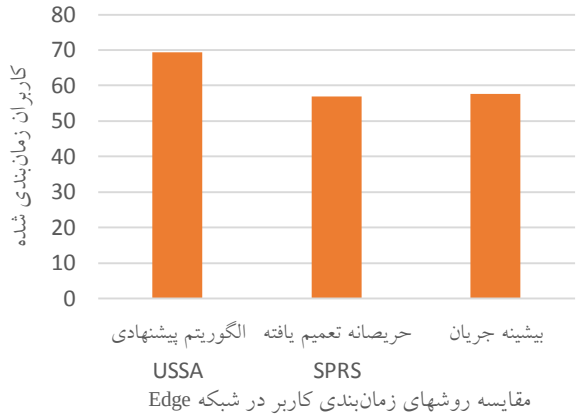
متغیر	توضیح	مقدار
N	تعداد کل کاربران کاندید شده برای اتصال	280
|EC|	تعداد ابر لبه	6
|L|	تعداد سرویس	1000
D	محدودیت فاصله	5km
N	محدودیت شبکه	15
P	محدودیت پردازش	10
S	محدودیت ذخیره‌سازی	5
α	توان توزیع زیف برای درخواست کاربران	0.6

برای تقریب بهتر و اطمینان از پاسخ‌های حاصل شده، برای به دست آوردن پاسخ زمان‌بندی کاربران، ۱۰۰ مرتبه مساله در موقعیت‌های متفاوت کاربران حل شده و میانگین پاسخ‌ها به‌عنوان خروجی اعلام شده است. به‌عنوان مثال، وقتی گفته می‌شود در شرایط جدول (۲) تعداد ۶۹/۳۷ کاربر زمان‌بندی شده‌اند به این معنی است که با متغیرهای ذکرشده در جدول(۲) برای موقعیت‌های مختلف کاربران در مکان‌های مختلف و آنتن‌های BTS با موقعیت‌های اتفاقی، تعداد ۱۰۰ بار مساله حل شده و میانگین آن مقدار ۶۹/۳۷ کاربر است.

۴،۲. تعریف معیارهای کارایی

۴،۲،۱. مقایسه تعداد کاربران زمان‌بندی شده

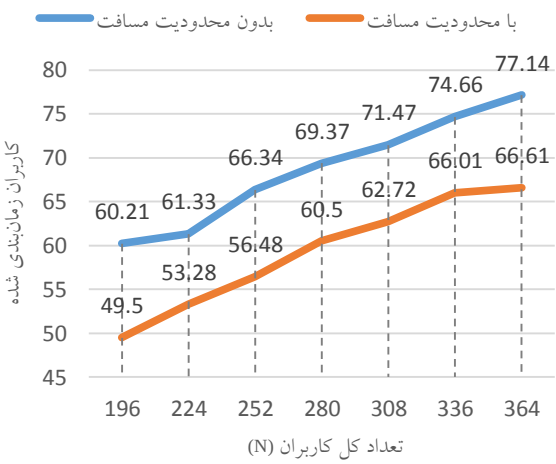
روش زمان‌بندی مقاله که در ابتدا برای سرویس دادن به کاربر سلول را بررسی می‌کند سپس بر اساس فاصله به سلول مجاور درخواست می‌دهد و در گام سوم بک‌هال و در نهایت ابر را بررسی می‌کند، مزایای زیادی دارد. در مقاله، محاسبات انجام شده در دو حالت مختلف، با احتساب محدودیت مسافت و بدون آن برای مقایسه و ارزیابی بهتر صورت گرفته است. همان‌طور که در شکل (۳) مشاهده می‌شود، در الگوریتم ارائه شده حتی با در نظر گرفتن محدودیت مسافت نتایج بهتری حاصل شده است. در مقاله [۱۳] در ۱۰۰ بازه زمانی تعداد کاربران زمان‌بندی شده در الگوریتم بیشینه جریان ۵۷/۶ و در الگوریتم حریصانه تعمیم یافته ۵۶/۸۵ است. در روش پیشنهادی بدون در نظر گرفتن محدودیت مسافت تعداد ۶۹/۳۷ کاربر زمان‌بندی شده‌اند که نسبت به الگوریتم بیشینه جریان و حریصانه به ترتیب ۲۰٪ و ۲۲٪ بهبود را نشان می‌دهد. همچنین پس از در نظر گرفتن محدودیت مسافت تعداد ۶۰/۵ کاربر زمان‌بندی شده‌اند که باز هم نسبت به روش مقاله [۱۳] مقدار بهتری را نشان می‌دهد.



شکل (۳)- مقایسه تعداد کاربران زمان‌بندی شده در بیشینه جریان، حریصانه تعمیم‌یافته SPRS و حریصانه تعمیم‌یافته USSA

۴،۲،۲. مقایسه کاربران زمان‌بندی شده به کل کاربران

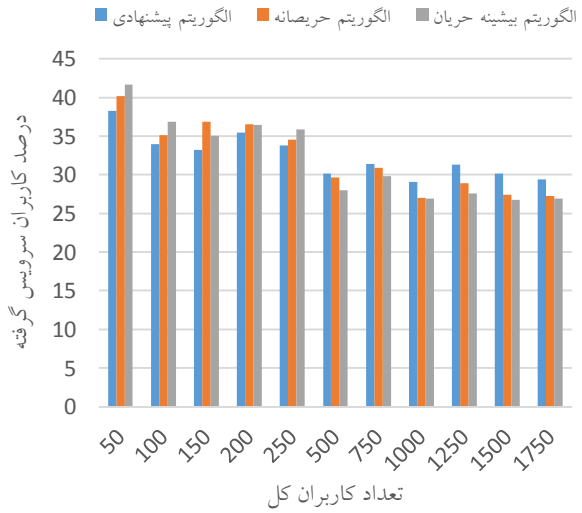
در شکل (۴) مقایسه‌ای برای تاثیر تغییر تعداد کل کاربران بر روی تعداد کاربران زمان‌بندی شده انجام شده است. همان‌طور که در این شکل مشخص شده، محدودیت مسافت در نتایج حاصل در حدود ۱۵٪ تاثیرگذار بوده و موجب کاهش نتایج حاصل شده است. همچنین افزایش تعداد کاربران کاندید برای اتصال به معنی افزایش تطابق شرایط کاربران با آنتن‌های BTS است. بنابراین بدیهی است با افزایش تعداد کاربران کاندید اتصال، تعداد کاربران زمان‌بندی شده نیز افزایش خواهد یافت. این تغییرات تا سقف مجموع محدودیت شبکه (۹۰ کاربر) افزایش خواهد یافت و بعد از آن به علت عدم توانایی در اتصال پیشرفتی نخواهد داشت. قابل ذکر است شیب این خط به مرور کم شده و به ندرت تمام ظرفیت شبکه پر می‌شود.



شکل (۴) – نمودار مقایسه‌ای برای درک تاثیر محدودیت مسافت در کاربران زمان‌بندی شده

۴،۲،۳. مقایسه تعداد کاربران زمان‌بندی شده براساس تعداد سرویس

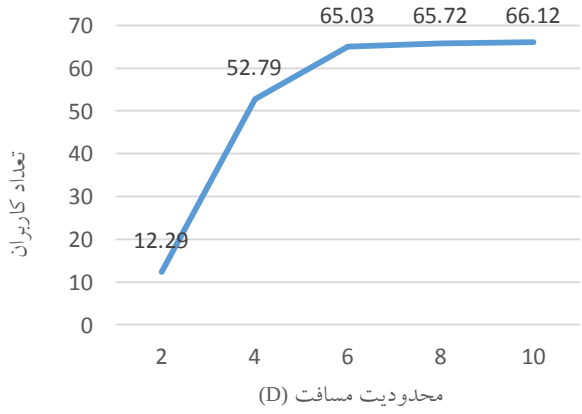
کاهش تعداد سرویس هر چند در بازه وسیعی بررسی شده است ولی به دلیل انتخاب توزیع مناسب زیف در اختصاص سرویس‌ها برای کاربران و آنتن‌های BTS، تاثیر کمتری بر روی کاربران زمان‌بندی شده گذاشته است که در شکل (۵) مشخص می‌شود. تغییرات تعداد کاربران دارای سرویس متناسب با تعداد کل کاربران زمان‌بندی شده بوده و حدودا کمتر از ۵۰٪ آنها را در هر سه الگوریتم تشکیل می‌دهد. بنابراین مشاهده می‌شود که تغییر تعداد سرویس موجب زمان‌بندی بیشتر کاربران بدون سرویس نمی‌شود. البته همان‌طور که از نمودار مشخص است در تعداد بیشتر از ۲۵۰، الگوریتم پیشنهادی بهتر عمل نموده و در مقادیر کمتر از آن هم تنها اندکی ضعیف‌تر است. علت این امر اعمال محدودیت مسافت است که در تعداد کاربران بیشتر فواصل کوتاه‌تر گره را داریم و در صورت عدم دریافت سرویس در سلول خود، ابتدا به سلول‌های همسایه وصل می‌شود و وارد بک‌هال نمی‌شود. اما اگر تعداد کمتر و مسافت دورتر باشد وارد بک‌هال می‌شود و تعداد کمتری را زمان‌بندی می‌کند. متوسط بهبود برای کاربران سرویس گرفته به کل کاربران در الگوریتم پیشنهادی نسبت به حریصانه ۰/۴ درصد و نسبت به بیشینه جریان ۱/۱ درصد است.



شکل (۵)– نمودار کاربران زمان‌بندی شده و سرویس گرفته

۴،۲،۴. مقایسه تعداد کاربران زمان‌بندی شده براساس محدودیت فاصله

شکل (۶) به تغییرات تعداد کاربران زمان‌بندی شده با توجه به تغییر محدودیت مسافت پرداخته شده است. افزایش محدودیت مسافت بیش از مقدار ۶ کیلومتر تاثیر چندانی در افزایش تعداد کاربران زمان‌بندی شده نداشته و به نظر می‌رسد انتخاب دامنه بین ۴ تا ۶ کیلومتر و به‌طور متوسط ۵ کیلومتر بهترین نتیجه را برای دامنه آنتن‌های BTS در این مقاله در بر خواهد داشت. همچنین از شکل (۶) مشخص است که در الگوریتم پیشنهادی با افزایش مسافت، تعداد کاربران سرویس داده شده ثابت می‌شود.



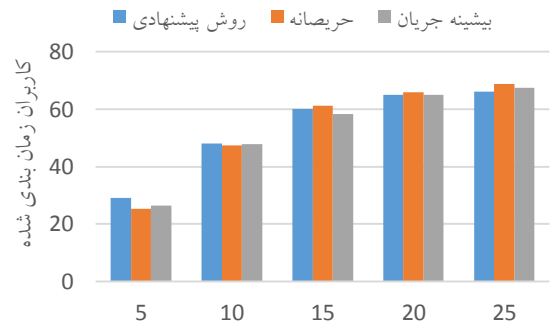
شکل (۶)– نمودار تغییرات کاربران زمان‌بندی شده با تغییر محدودیت مسافت

دلیل این موضوع می‌تواند همپوشانی دامنه BTS ها برای سرویس‌دهی و اتصال کاربران باشد. قابل ذکر است در دیتاست‌های انتخاب شده حداکثر فضای ممکن یک فضای ۵۰×۵۰ کیلومتری است که در هر سری محاسبات، فضای ممکن کوچک‌تر و مساوی این فضا می‌باشند. برخی از کاربران خارج از دامنه موثر آنتن‌ها هستند و دورترین کاربر با BTS در حدود ۴۳ کیلومتر و نزدیک‌ترین کاربر در حد چند متر با BTS فاصله دارند.

در روش پیشنهادی اگر کاربر موجود در ابر لبه امکان سرویس گرفتن نداشته باشد (برخلاف روش مقایسه شده که درخواست را در بک‌هال منتشر می‌کند) ابتدا سعی می‌نماید در همسایه‌های ابر لبه به جستجوی سرویس بپردازد و در صورتی که پیدا نشد دنبال بک‌هال برود. علت اینکه نتیجه ابتدا بدتر شده است این است که چون تعداد کاربران در ابتدا کم است همه آنها بیشتر در ابر لبه مختص خود سرویس می‌گیرند و کمتر سراغ بک‌هال می‌روند و ابتدا سراغ لبه‌های مجاور می‌روند در حالی که با افزایش تعداد آنها، عدم پاسخگویی بیشتر شده و این مساله خود را بهتر نشان می‌دهد.

۴,۲,۵. مقایسه تعداد کاربران زمان‌بندی شده براساس محدودیت شبکه

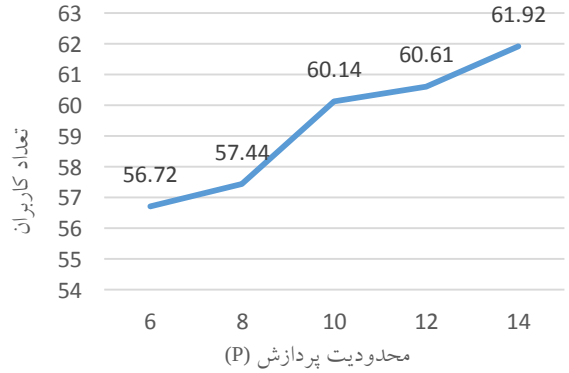
شکل (۷) به بررسی تغییرات در کاربران زمان‌بندی شده برحسب تغییر محدودیت شبکه پرداخته است. همان‌طور که در نمودار ستونی آبی در مقادیر مختلف مشخص است، تغییرات محدودیت شبکه از حدود ۲۵ کاربر به بعد با تغییرات بسیار جزئی انجام می‌شود و تاثیر چندانی در نتیجه نهایی نخواهد داشت. تعداد کاربران کاندید به اندازه ۲۸۰ کاربر برای محدودیت شبکه بیش از ۱۵ عدد به مرور اشباع شده و در اصل کاربری با شرایط مورد نظر جهت اتصال به BTS وجود نخواهد داشت. چنانچه تعداد کل کاربران کاندید به همراه محدودیت شبکه افزایش یابد، تعداد کاربران زمان‌بندی شده نیز به همراه آنها افزایش خواهند یافت.همچنین از مقایسه نمودارها مشخص است که تعداد کاربران زمان‌بندی شده برای محدودیت شبکه کمتر از ۱۵ در روش پیشنهادی بهتر و برای بیشتر از آن اندکی کمتر است که ناشی از همین تاثیرگذاری مسافت و اتصال به سلول‌های همسایه نسبت به بک‌هال در محدودیت شبکه کمتر و برعکس شدن شرایط یعنی انتخاب بک‌هال در محدودیت شبکه بیشتر در الگوریتم‌های حریصانه و بیشینه جریان است.



شکل (۷)- تغییر کاربران زمان‌بندی شده نسبت تغییرات محدودیت شبکه

۴,۲,۶. مقایسه تعداد کاربران زمان‌بندی شده براساس محدودیت پردازش

با توجه به ارتباط محدودیت پردازش با سرویس کاربران و تاثیر مستقیم کاربران دارای سرویس در این محدودیت، همان‌طور که در شکل (۸) آمده است تغییر زیاد نخواهد بود. تغییرات موجود بین هر پله معدود کاربرانی هستند که به دلیل محدودیت هر سلول برای پردازش و پاسخ به درخواست سرویس کاربر از زمان‌بندی بازمانده بودند. با افزایش محدودیت پردازش، در کل ۶ سلول برای هر مرحله افزایش حداکثر ۲ کاربر بیشتر زمان‌بندی شده است.



شکل (۸) - تغییرات محدودیت پردازش در کاربران زمان‌بندی شده

یکی دیگر از ارزیابی‌هایی انجام شده، بررسی و مقایسه این الگوریتم با جایگذاری سرویس ارائه شده در مقاله [۲۵] است که در آن با استفاده از یادگیری تقویتی جایگذاری سرویس‌ها در ابر لبه انجام می‌شود. اگر چه، در جایگذاری با یادگیری تقویتی، ارتباط کارها با یک گراف بدون‌دور جهت‌دار (DAG) با وابستگی بین آنها نشان داده شده است و در عمل الگوریتم ارائه شده بهبود را در این شرایط ایجاد می‌کند، اما برای سنجش کارایی روش پیشنهادی، داده‌های واقعی تاکسی‌های سانفرانسیکو (که کارهای آن وابستگی به هم ندارند) به عنوان ورودی به هر دو الگور یتم داده شده و نتایج در جدول (۳) گزارش شده است.

جدول (۳).مقایسه درصد کاربران سرویس گرفته بر اساس تعداد کاربران

تعداد کاربران زمان‌بندی شده	درصد کاربران سرویس گرفته	
	روش پیشنهادی	یادکیری تقویتی
50	40.20	38.56
100	33.94	32.18
150	33.19	32.73
200	35.45	33.27
250	33.37	33.15
500	29.69	27.89
750	31.40	31.28
1000	29.05	30.31
1250	31.32	33.82
1500	27.43	30.10
1750	29.42	31.44

همان طور که از نتایج جدول (۳) مشخص است، در ابتدا الگوریتم پیشنهادی نسبت به این الگوریتم عملکرد بهتری را از خود نشان می‌دهد اما با افزایش تعداد کاربران زمان‌بندی شده دقت آموزش بالا رفته و نتایج بهتری برای درصد کاربران سرویس گرفته نمایش داده می‌شود. هر چند به لحاظ زمان اجرا روش یادگیری تقویتی نسبت به روش پیشنهادی این مقاله کندتر خواهد بود چون مرحله آموزش را خواهد داشت.

۴،۳. پیچیدگی

با توجه به این که در این مقاله از الگوریتم فورد–فولکرسون (که در مقاله [۱۲] ارائه شده است) استفاده می‌شود پیچیدگی آن $O(|N||U|^2)$ است که در آن N تعداد ابرهای لبه و U تعداد درخواست کاربران است. همچنین از الگوریتم‌های حریصانه برای جایگذاری سرویس و زمان‌بندی کاربران استفاده شده است که بخش‌های پیشنهادی اضافه شده به مقاله [۱۲] برای اضافه کردن محدودیت مسافت در الگوریتم‌های (۱) و (۲) در پیچیدگی الگوریتم تغییری ایجاد نمی‌کند. بنابراین پیچیدگی حل مساله جایگذاری سرویس و زمان‌بندی با روش حریصانه دارای پیچیدگی $O(N^2|U||L|S)$ است که در آن N تعداد ابرهای لبه، U تعداد درخواست‌های کاربران، L تعداد سرویس‌ها و S محدودیت ذخیره‌سازی است. همان طور که از الگوریتم (۲) مشخص است، اضافه کردن متغیر جدید D تغییری در پیچیدگی راه‌حل حریصانه ایجاد نمی‌کند در حالی که تعداد کاربران سرویس داده شده بیشتر می‌شود.

۵. نتیجه‌گیری

مساله جایگذاری سرویس و زمان‌بندی درخواست کاربران در ابرهای لبه بسیار مهم است. در مقاله موجود با اعمال روش‌های مناسب و کارا، زمان‌بندی و جایگذاری سرویس بهبود داده شده است. همچنین محدودیت‌های واقعی‌تر مطابق وضعیت و نیاز روز به مساله اضافه شده و در نهایت نتایج بهبود قابل ملاحظه‌ای را نشان می‌دهند. مهم‌ترین نوآوری موجود در این مقاله، استفاده از سرورهای لبه مجاور (BTS) قبل از استفاده از بک‌هال با اضافه کردن متغیر فاصله در زمان عدم وجود سرویس در سرور لبه مربوط به کاربر است که باعث می‌شود از بک‌هال تا جای ممکن کمتر استفاده شود. اگرچه در این مقاله برای پیدا کردن سرورهای نزدیک مجاور از یک هیوریستیک استفاده می‌شود، پیشنهاد می‌شود محققان در کارهای آینده بر پیدا کردن سرور مناسب با الگوریتم‌های متاهیوریستیک تمرکز نمایند زیرا هر الگوریتم و روشی که بتواند ارتباطات محلی را جایگزین استفاده از بک‌هال نماید می‌تواند زمان‌بندی درخواست کاربران و جایگذاری سرویس را بهتر و سریع‌تر انجام دهد. همچنین می‌توان برای انتخاب سرورهای لبه کاندید از میان آنهایی که سرویس مورد نظر را دارند با استفاده از روش‌های تصمیم‌گیری

چند شاخصه مانند تاپسیس و تعریف متغیرهای مناسب برای آن انتخاب‌ها، کاندیدهای بهتری را معرفی کرد.

تعارض منافع: نویسندگان اعلام می‌کنند که هیچ تعارض منـافعی ندارند.

مراجع

[1] Borcoci E., Vochin M., and Obreja S., "Mobile Edge Computing versus Fog Computing in Internet of Vehicles," in research gate, Bucharest, Romania, 2018.

[2] Chunlin L., Sun H., Yi C., and Luo Y., "Edge cloud resource expansion and shrinkage based on workload for minimizing the cost," Future Generation Computer Systems, 101:327-340, 2019.

[3] Li X. and Xu L. D., "A Review of Internet of Things-Resource Allocation," IEEE Internet of Things Journal, 8(11), 2021.

[4] Jia M., Cao J., and Liang W., "Optimal cloudlet placement and user to cloudlet allocation in wireless metropolitan area networks," IEEE Transactions on Cloud Computing, 5(4):725-737, 2017.

[5] Xu Z., Liang W., Xu W., Jia M., and Guo S., "Efficient algorithms for capacitated cloudlet placements," IEEE Transactions on Parallel and Distributed Systems, 27: 2866–2880, 2016.

[6] Ceselli A., Premoli M., and Secci S., "Mobile edge cloud network design optimization," IEEE/ACM Transactions on Networking, 25:1818–1831, 2017.

[7] Wang L., Jiao L., Li J., and Muhlhauser M., "Online resource allocation for arbitrary user mobility in distributed edge clouds," in 37th International Conference on Distributed Computing Systems (ICDCS), 2017.

[8] Tan H., Han Z., Li X.-Y., and Lau F. C. M., "Online job dispatching and scheduling in edge-clouds," in IEEE INFOCOM, 2017.

[9] Sabaghian K., Khamforoosh K., and Ghaderzadeh A., "Presentation of a new method based on modern multivariate approaches for big data replication in distributed environments," Plos one 16, no. 7 (2021): e0254210.

[10] Llorca J., Tulino A. M., Sforza A., and Sterle C., "Optimal content distribution and multi-resource allocation in software defined virtual," in AIRO ODS, 2017.

[11] Zhao T., Zhou S., Guo X., and Niu Z., "Tasks Scheduling and Resource Allocation in Heterogeneous Cloud for Delay-bounded Mobile Edge Computing," in IEEE ICC 2017 SAC Symposium Cloud Communications and Networking Track, 2017.

[12] Chen X., Li W., Lu S., Zhou Z., and Fu X., "Efficient Resource Allocation for On-Demand Mobile-Edge Cloud Computing," IEEE Transactions on Vehicular Technology , 67(9): 8769 - 8780, 2018.

[13] He T., Khamfroush H., Wang S., La Porta T., and Stein S., "Joint Service Placement and Request Scheduling in Edge Clouds with Sharable and Non-Sharable Resources," in 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS), Vienna, Austria, 2018.

[14] Farhadi V., Mehmeti F., He T., La Porta T., Khamfroush H., Wang S., and Chan K. S., "Service Placement and Request Scheduling for Data-Intensive Applications in Edge Clouds," in IEEE INFOCOM 2019 - IEEE Conference on Computer Communications, Paris, 2019.

[15] Zhao J., Li Q., Gong Y., and Zhang K., "Computation Offloading and Resource Allocation For Cloud Assisted Mobile Edge Computing in Vehicular Networks," IEEE Transactions on Vehicular Technology, 68: 7944 - 7956, 2019.

[16] Li Y., Dai W., Gan X., Jin H., Fu L., Ma H., and Wang X., "Cooperative Service Placement and Scheduling in Edge Clouds: A Deadline-Driven Approach," IEEE Transactions on Mobile Computing, p. 1, 2021.

[17] Han T. and Ansari N., "Network Utility Aware Traffic Load Balancing in Backhaul-Constrained Cache-Enabled Small Cell Networks with Hybrid Power Supplies," IEEE Transactions on Mobile Computing, 16(10): 2819 - 2832, 2017.

[18] Wei Y., Zhang Z., Yu F. R., and Han Z., "Joint User Scheduling and Content Caching Strategy for Mobile Edge Networks Using Deep Reinforcement Learning," in IEEE International Conference on Communications Workshops, Kansas City, MO, USA, 2018.

[19] Li C., Bai J., and Tang J. H., "Joint optimization of data placement and scheduling for improving user experience in edge computing," Journal of Parallel and Distributed Computing, 125:93-105, 2020.

[20] Yin B., Chen Z., and Tao M., "Joint User Scheduling and Resource Allocation for Federated Learning over Wireless Networks," in IEEE Global Communications Conference, 2020.

[21] Wang S., Chen M., Liu X., and Yin C., "A Machine Learning Approach for Task and Resource Allocation

in Mobile Edge Computing Based Networks," IEEE Internet of Things Journal, 8(3): 1358-1372, 2020.

[22] Wang J. and Wang L., "A Computing Resource Allocation Optimization Strategy for Massive Internet of Health Things Devices Considering Privacy Protection in Cloud Edge Computing Environment," Journal of Grid Computing, vol. 19, 2021.

[23] Molner N., de la Oliva A., Stavrakakis I., and Azcorra A., "Optimization of an integrated fronthaul/backhaul network under path and delay constraints," Ad Hoc Networks, 83: 41-54, 2019.

[24] Korte B. and Vygen J., "Network flows," Combinatorial Optimization, 8: 153-184, 2000.

[25] Wang J., Hu J., Min G., Zhan W., Zomaya A., and Georgalas N., "Dependent Task Offloading for Edge Computing based on Deep Reinforcement Learning," in IEEE Transactions on Computers, 2021, doi: 10.1109/TC.2021.3131040.