

بهینه‌سازی چند هدفه به منظور تخصیص منابع محاسباتی و مخابراتی مبتنی بر دسترسی نامتعاد، مشارکت سرور ابری و سرور لبه در شبکه‌های نسل پنجم

نیما نوری^۱، دانشجوی کارشناسی ارشد؛ علی اکبر تدین^۲، دانشیار
۱- دانشگاه یزد - دانشکده مهندسی برق، nimanouri68@gmail.com
۲- دانشگاه یزد - دانشکده مهندسی برق، tadaion@yazd.ac.ir

چکیده - پردازش لبه موبایل یک فناوری نوظهور است که به کاربران اجازه می‌دهد، برای غلبه بر محدودیت‌های گوشی‌های موبایل، به جای ارسال بخشی از برنامه به سمت سرور ابری که در فاصله بسیار دور نسبت به کاربران شبکه قرار گرفته است، آن را به سرور لبه ارسال کنند تا بتوانند بر تأخیر بالای پردازش ابری که موجب افت کیفیت سرویس‌دهی می‌شود، غلبه کنند. در این مقاله، سیستمی شامل یک سلول و چندین کاربر مطرح شده است که کاربران شبکه برای انجام پردازش خود از سرورهای ابری و لبه که با هم در مشارکت هستند؛ تقاضای سرویس می‌کنند. نحوه دسترسی کاربران به طیف رادیویی، به صورت دسترسی چندگانه نامتعاد فرض می‌شود و برای مدل‌سازی مسئله در سمت کاربر و سرورها، از تئوری صف استفاده شده است. هدف اصلی این است که در شبکه، مجموع انرژی مصرف شده توسط کاربران، تأخیر دریافت سرویس کاربران و هزینه کل استفاده از سرورها را حداقل کنیم. مدل‌سازی ریاضی این مسئله، منجر به یک مسئله نامحدب مقید چند هدفه خواهد شد. از روش SCA برای به دست آوردن جواب بهینه سراسری، استفاده می‌شود. با استفاده از نتایج شبیه سازی نشان داده می‌شود که با فرضیات وارد شده در مدل پیشنهادی، میزان انرژی، تأخیر و هزینه کل شبکه در حدود ۵۰ درصد کاهش پیدا می‌کند.

واژه‌های کلیدی: رایانش لبه موبایل، دسترسی چندگانه نامتعاد، تابع نامحدب، تقریب محدب، تئوری صف.

Multi Objective Optimization for Computational and Communicational Resource Allocation Based on Non-orthogonal Multiple Access, Cloud and Edge Servers Participation in 5 Generation Networks

N. Nouri¹, MSc Student; A. A. Tadaion², Associate Professor
1-MSc Student of Electrical Engineering, Yazd University, Yazd, Iran, Email: nimanouri68@gmail.com
2- Faculty of Electrical Engineering, Yazd University, Yazd, Iran, Email: tadaion@yazd.ac.ir

Abstract: Mobile Edge computing is a new paradigm that allows users to combat the constraints of mobile devices. For overcoming the long delay of cloud computing which decreases the quality of service, the users send their data to the server which is located at the edge of the network instead of sending a part of the program to a cloud server which is located in place far from the users. In this paper, a system which consists of a cell and several users is considered in which the network users to accomplish their computations request service from cloud and edge servers that are participating with each other. It is assumed that the users access the radio spectrum by a Non-orthogonal Multiple Access (NOMA) method and the queue theory is used for modeling the problem from the users and server sides. The main goal is to minimize the total users' energy consumption, the delay in users' receiving service, and the whole cost of using servers. The mathematical model of this problem would result in a multi-objective constraint non-convex optimization problem. SCA method is used to achieve the globally optimum solution. Through the simulation results, it is shown that by the applied assumptions in the proposed model, the amount of energy, delay and total cost of network is decreased around 50%.

Keywords: Mobile Edge Computing, Non-orthogonal Multiple Access, Non convex Function, Convex Approximation, Queue Theory.

تاریخ ارسال مقاله: ۱۳۹۷/۰۴/۲۲

تاریخ اصلاح مقاله: ۱۳۹۷/۰۹/۱۷ و ۱۳۹۷/۱۱/۳۰

تاریخ پذیرش مقاله: ۱۳۹۸/۰۱/۰۳

نام نویسنده مسئول: علی اکبر تدین

نشانی نویسنده مسئول: ایران - یزد - صفائیه - بلوار دانشگاه - دانشگاه یزد - دانشکده مهندسی برق

۱- مقدمه

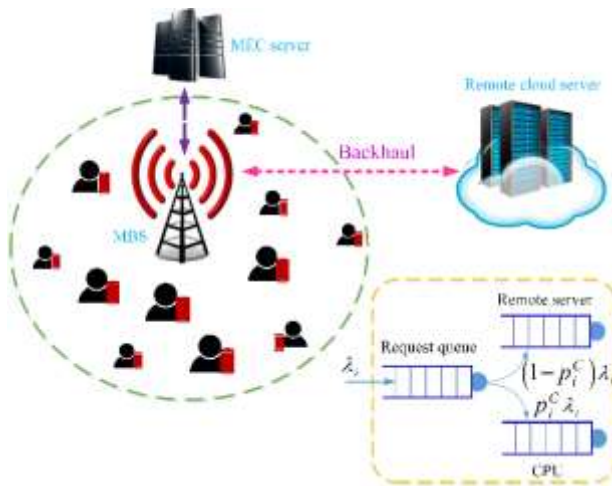
علاوه بر تعریف تأخیر به عنوان تابع هدف، الگوریتمی برای کاهش رد درخواست کاربر از سرور لبه را مطرح کرده است. در واقع از آنجا که درخواست کاربر در تایم اسلات‌های متوالی به سرور لبه ارسال می‌شود، در این مقاله هدف این است که تعداد رد درخواست کاربر از سرور نیز حداقل شود. لذا با توجه به ماهیت دینامیکی مسئله، پس از مدل‌سازی سعی شده از روش لیاپانوف برای حل مسئله استفاده شود. نویسنده با استفاده از روابط تحلیلی و نتایج شبیه‌سازی نشان داده است که الگوریتم مطرح شده می‌تواند بیش از ۶۴ درصد تأخیر در دریافت سرویس را کاهش دهد، بنحوی که تعداد رد درخواست حداقل شود. در [۱۷] و [۱۸] نیز نویسندگان با فرض تخلیه کامل در سیستمی تک کاربره به حداقل کردن انرژی مصرفی کل شبکه تحت قید تأخیر می‌پردازند که در [۱۷] به صورت یک مسئله تصمیم‌گیری مارکوف مدل شده است و برای حل این مسئله، دو روش آنالاین^۷ و آفلاین^۸ معرفی شده است. در نهایت نتایج شبیه‌سازی بیانگر این است که روش آفلاین، نسبت به روش آنالاین بیش از ۵۰ درصد عملکرد بهتری دارد و با الگوریتم‌های پیشنهادی حدود ۷۸ درصد انرژی مصرفی کاهش پیدا می‌کند. از آنجایی که در [۱۷] روش آفلاین عملکرد بهتری نسبت به روش آنالاین داشت، لذا در [۱۸] نویسندگان دو روش آفلاین دیگر برای بهبود روش قبلی ارائه کردند؛ روش آفلاین معین^۹ و تصادفی^{۱۰}. استفاده از این دو روش با توجه به نتیجه شبیه‌سازی منجر به کاهش قابل توجهی در مصرف انرژی شبکه می‌شود؛ به نحوی که استفاده از این روش نسبت به حالت پردازش محلی در حدود ۷۰ درصد و نسبت به پردازش سرور لبه حدود ۱۵ درصد منجر به کاهش مصرف انرژی شبکه می‌شود.

در اغلب کارهای یادشده، مدل سیستم به صورت تک کاربره بوده است؛ لذا در [۱۹-۲۶] برخلاف کارهای قبلی، نویسندگان سعی کرده‌اند سیستم را به صورت چندکاربره مدل کنند. در [۱۹] نویسندگان با هدف حداقل کردن انرژی مصرفی شبکه تحت قید تأخیر، الگوریتمی تحت عنوان EECO^{۱۱} مطرح کرده‌اند که با این الگوریتم طبق نتایج شبیه‌سازی می‌توان در حدود ۱۵ درصد انرژی مصرفی شبکه را کاهش داد. همچنین در [۲۰] نویسندگان با هدف ایجاد یک مصالحه^{۱۲} بین انرژی و تأخیر در شبکه با مدل‌سازی این شبکه و استفاده از روش تئوری بازی توانستند تا حد قابل قبولی هزینه کل شبکه را کاهش دهند و برای اثبات آن از روابط تحلیلی و شبیه‌سازی در این مقاله استفاده شده است. در [۲۱] و [۲۲] نیز با هدف کاهش ترکیب وزن‌دار انرژی و تأخیر به ترتیب در مدلی تک کاربره و چندکاربره سیستم را مدل کرده و برای حل مسئله فوق از روش‌های ابتکاری^{۱۳} استفاده کرده‌اند و توانستند هزینه کل شبکه را به ترتیب به اندازه ۴۰ و ۷۰ درصد کاهش دهند؛ اما یکی از معایب روش دوم، پیچیدگی بالای الگوریتم مطرح شده است. در تمام مقالات فوق، فرض بر این بود که کاربران برای دریافت سرویس، از تخلیه کامل استفاده کنند؛ برخلاف کارهای قبلی در [۲۳-۲۶] فرض بر این است که کاربران شبکه برای دریافت سرویس از تخلیه جزئی^{۱۴} استفاده کنند. به عنوان مثال در [۲۳-۲۵] نویسندگان با هدف حداقل کردن انرژی شبکه با قید تأخیر

با رشد روزافزون فناوری گوشی‌های موبایل، برنامه‌های کاربردی با محبوبیت زیاد بین کاربران، همه‌روزه وارد بازار می‌شوند. این برنامه‌ها روزبه‌روز پیچیده‌تر و بالطبع دارای محاسبات سنگین‌تری می‌شوند؛ اما علیرغم افزایش فناوری در گوشی و برنامه‌های کاربردی، همچنان چالش‌های زیادی در منابع آن‌ها وجود دارد که نمونه‌ای از این چالش‌ها شامل طول عمر باتری، ظرفیت ذخیره‌سازی، ظرفیت محاسباتی و غیره هستند که استفاده از این برنامه‌ها را با محدودیت مواجه می‌کند. رایانش ابری موبایل^۱ (MCC)، به عنوان یک راه‌حل مؤثر برای غلبه بر این محدودیت در گوشی‌های موبایل، در سال‌های اخیر مطرح شده است با این هدف که بتوانیم از پتانسیل‌های موجود در رایانش ابری^۲ (CC) در پردازش موبایل بهره‌مند شویم [۱۱-۱]. در واقع می‌توان گفت پردازش ابری موبایل ترکیبی از پردازش ابری و پردازش موبایل است [۷]. با این روش می‌توان بخشی از برنامه موردنظر را که دارای پردازش پیچیده و محاسبات سنگین است، به سمت سرور ابری^۳ ارسال کرد [۸]. مزیت استفاده از این روش در این است که با این کار میزان مصرف انرژی در گوشی‌های همراه کاربران کاهش پیدا کرده که با افزایش طول عمر باتری و افزایش سرعت پردازش همراه خواهد بود [۱۲، ۱۳]. همچنین با استفاده از این نوع پردازش دیگر نیازی نیست که کاربران، گوشی خود را از نظر نرم‌افزاری و سخت‌افزاری ارتقا دهند. یکی از نقاط ضعف این نوع پردازش این است که سرور ابری معمولاً در فاصله بسیار دور از کاربر قرار می‌گیرد و باعث تأخیر در سرویس‌دهی می‌شود از این جهت برای بسیاری از برنامه‌هایی که به تأخیر حساس هستند منجر به افت کیفیت سرویس^۴ (QOS) می‌شود از این رو به تازگی مفهوم جدیدی به نام پردازش ابری لبه توسط موسسه استاندارد ارتباطات اروپا^۵ (ETSI) مطرح شده است؛ با این هدف که با قرار دادن این سرور نزدیکی کاربران شبکه، بر این ضعف غلبه کنند. البته لازم به ذکر است که این سرورها نسبت به سرورهای ابری دارای قدرت محاسباتی و ذخیره‌سازی کمتری هستند و تنها مزیت آن‌ها نزدیکی به کاربران شبکه است [۱۴].

۲- کارهای پیشین

اخیراً، تحقیقات بسیار زیادی در خصوص پردازش ابری و لبه صورت گرفته است [۱۵-۱۸]. در [۱۵] مدلی تک کاربره شامل سرور لبه، با فرض تخلیه کامل^۶ سعی در حداقل کردن تأخیر دریافت سرویس کاربر در شبکه را دارد. پس از مدل‌سازی مسئله برای حل آن یک الگوریتم جستجوی یک‌بعدی برای به دست آوردن جواب بهینه مسئله مطرح شده که در نهایت نتایج حاصل از شبیه‌سازی نشان‌دهنده این است که الگوریتم پیشنهادی قادر است که تأخیر دریافت سرویس را در حدود بیش از ۸۰ درصد در مقایسه با پردازش محلی و حدود ۴۰ درصد نسبت به پردازش ابری کاهش دهد. در [۱۶] همانند حالت قبل نویسنده با در نظر گرفتن مدل دینامیکی برای سیستمی تک کاربره به همراه سرور لبه



شکل ۱: مدل سیستم پیشنهادی

۳- مدل سیستم

شکل ۱ مدل سیستم پیشنهادی را نشان می‌دهد. همان‌طور که در شکل دیده می‌شود، سیستمی شامل N کاربر در یک سلول، یک ایستگاه پایه مجهز به سرور لبه و یک سرور ابری در فاصله دور از آن در نظر گرفته شده است. این سرورها منابع ذخیره‌سازی و محاسباتی را برای کاربران فراهم می‌کنند. کاربران درون سلول می‌توانند از طریق ایستگاه پایه، به سرور ابری یا سرور لبه دسترسی پیدا کنند. البته لازم به ذکر است که سرور ابری و ایستگاه پایه، از طریق لینک بک‌هال^{۲۰}، باهم در ارتباط هستند. کل کاربران درون شبکه را می‌توان با مجموعه $\mathcal{N} = \{1, 2, \dots, N\}$ نمایش داد. هر یک از کاربران شبکه، قصد اجرای برنامه‌ای را دارند که منجر به تولید یک سری از درخواست‌ها می‌شود. در این مقاله، مدل صف در سمت کاربر، سرور لبه و سرور ابری به ترتیب به صورت $M/M/1$ ، $M/M/c$ ، $M/M/\infty$ در نظر گرفته شده است [۳۲، ۳۳] و فرض می‌شود که کاربران و ایستگاه پایه تک آنتن هستند. همچنین نحوه دسترسی کاربران به طیف رادیویی به صورت دسترسی چندگانه نامتعامل است و به دلیل اینکه هم مسیر فراسو و هم فرسو در این مقاله مورد بررسی قرار می‌گیرد، فرض می‌کنیم که کاربران شبکه و ایستگاه پایه برای ارسال اطلاعات خود از کدگذاری برهم‌نهی^{۲۱} (SC) استفاده می‌کنند و مجهز به تکنیک^{۲۲} (SIC) برای کاهش اثر تداخل شبکه هستند. هر یک از این کاربران می‌توانند، قسمتی (و یا کل) از درخواست خود را به سرور لبه تعبیه‌شده در ایستگاه پایه، از طریق لینک رادیویی ارسال کنند. با توجه به ظرفیت سرور لبه، اگر مجموع میزان نرخ درخواستی، از سمت کاربران، از نرخ سرویس‌دهی کل، در سرور لبه کمتر باشد، در این صورت کل کاربران می‌توانند بدون از دست دادن کیفیت سرویس، سرویس مطلوب خود را از سرور لبه دریافت کنند؛ اما در غیر این صورت، بخشی از درخواست کاربران به سرور ابری، منتقل خواهد شد. فرض می‌شود، سرویس‌های درخواستی هر کاربر $i \in \mathcal{N}$ ، دارای توزیع پواسن با متوسط نرخ ورود λ_i باشد [۳۳]. از طرفی، این درخواست‌ها دارای حجم محاسباتی بالا و متقابلاً مستقل از هم هستند و می‌توانند هم به صورت

در شبکه‌ای تک کاربره مسئله خود را مدل کرده و برای تخصیص بهینه منابع تعریف شده در مسئله خود الگوریتمی را ارائه داده‌اند که نتایج حاصل از شبیه‌سازی نشان‌دهنده این است که الگوریتم‌های مطرح‌شده به ترتیب قادرند در حدود ۴۷، ۴۰ و ۲۵ درصد انرژی مصرفی را با رعایت قید تأخیر کاهش دهند؛ اما قابل ذکر است که در بین روش‌های فوق، الگوریتم‌های مطرح‌شده در [۲۳] و [۲۴] پیچیدگی کمی داشته در صورتی که الگوریتم مطرح‌شده در [۲۵] نسبت به دو الگوریتم دیگر پیچیدگی کمتری دارد. در [۲۶] نیز برای ایجاد یک مصالحه بین توان مصرفی کاربران شبکه و تأخیر دریافت سرویس با قید پایداری صف بافر در سمت کاربران در شبکه‌ای چندکاربره سیستم را به صورت دینامیک مدل کرده و از آنجاکه مسئله فوق دارای ماهیت دینامیکی است، برای حل مسئله از الگوریتمی آنلاین مبتنی بر روش لیاپانوف استفاده کرده است؛ همچنین این مقاله بحث تخصیص پهنای باند بهینه به کاربران نیز پرداخته است. الگوریتم مطرح‌شده در این مقاله قادر است به مقدار قابل توجهی میزان توان و تأخیر در شبکه را کاهش دهد. تعدادی از کارهای جدید نیز بر روی تخصیص منابع در شبکه‌های ناهمگن متمرکز شده‌اند [۲۷-۳۱]. به عنوان مثال در [۲۹-۲۷] نویسندگان به مسئله افزایش بهره‌وری انرژی و طیف رادیویی در یک شبکه چندکاربره ناهمگن پرداخته‌اند. همچنین [۳۰] به بحث تخصیص بهینه توان ارسالی در یک شبکه فوق مترکم ناهمگن مبتنی بر امواج موج‌میلیمتری پرداخته است و الگوریتمی برای انتخاب کاربران سرویس گیرنده از سرور مطرح کرده است. در [۳۱] به بحث تخصیص بهینه توان در یک شبکه ناهمگن دولایه پرداخته شده و الگوریتمی برای کاهش اثر تداخل شبکه ارائه شده است. در این مقاله، مدلی معرفی می‌شود که در آن سرور ابری و سرور لبه^{۱۵} با همکاری هم به درخواست کاربران شبکه پاسخ می‌دهند. در این مدل نحوه دسترسی کاربران به منابع رادیویی را به صورت دسترسی چندگانه نامتعامل^{۱۶} (NOMA) در نظر می‌شود. هر یک از این سرورها در قبال خدماتی که به کاربر ارائه می‌دهند، هزینه‌ای را دریافت می‌کنند. با در نظر گرفتن توان ارسالی هر کاربر برای ارسال داده به ایستگاه پایه^{۱۷} (BS) و تعیین محل پردازش (به صورت محلی^{۱۸} و یا از سرورها) به عنوان متغیرهای بهینه‌سازی، تابع هدف مسئله موردنظر را به صورت مینی‌م‌سازی متوسط انرژی مصرفی، تأخیر در دریافت سرویس و هزینه دریافت سرویس مطلوب برای کاربران شبکه تعریف می‌شود. پس از مدل‌سازی سیستم به فرم یک مسئله بهینه‌سازی، دیده می‌شود این مسئله دارای فرم نامحدب است؛ سپس برای به دست آوردن جواب بهینه مسئله از روشی موسوم به SCA^{۱۹} استفاده می‌شود.

در ادامه به موضوعات زیر پرداخته می‌شود. در بخش سوم مدل سیستم موردنظر شرح داده می‌شود و مسئله تخصیص بهینه منابع به صورت ریاضی فرمول‌بندی می‌شود. در بخش چهارم به معرفی روش حل پیشنهادی و حل مسئله فوق و در بخش پنجم به نتایج حاصل از شبیه‌سازی پرداخته می‌شود. در بخش آخر نتیجه‌گیری و کارهای آتی ارائه می‌شود.

جدول ۱: پارامترهای کلیدی مورد استفاده

مفهوم	نماد
تعداد کاربران شبکه	N
متوسط نرخ ورود درخواست کاربر شماره i	λ_i
احتمال تخلیه کاربر شماره i	p_i^c
توانایی محاسباتی دستگاه کاربر شماره i	u_i^M
حجم بار نرمالیزه شده در کاربر شماره i	l_i^M
توان محاسباتی در حالت محلی در کاربر شماره i	κ_i
تعداد بیت داده ورودی در هر درخواست کاربر شماره i	θ_i
پهنای باند کانال	W
میزان توان مصرفی برای ارسال داده در کاربر شماره i	P_i
نویز گوسی سفید جمع‌شونده	n_i
حداکثر توان موجود کاربر شماره i	P_i^{\max}
ضریب کانال بین کاربر i و ایستگاه پایه	h_i
توان نویز محیط	N_0
تعداد سرورهای موجود در سرور لبه	c
نرخ سرویس‌دهی در هر یک از سرورهای لبه	u^F
نرخ ارسال داده در هر یک از سرورهای لبه	u_b^F
میزان تأخیر کل ایجادشده در لینک بک‌هال	T^o
نرخ سرویس‌دهی در سرور ابری	u^{cc}
نرخ ارسال داده در سرور ابری	u_b^{cc}

همچنین نسبت سیگنال به اغتشاش (مجموع نویز و تداخل) برای کاربر i نیز به صورت زیر قابل تعریف است [۳۴، ۳۵].

$$SINR_i = \frac{p_i^c H_i}{1 + \sum_{j=1, j \neq i}^N p_j^c H_j} \quad (۴)$$

که در رابطه فوق [۳۴، ۳۵]:

$$H_i = \frac{h_i}{N_0} \quad (۵)$$

$$N_0 = \mathbf{E}[|n_i|^2] \quad (۶)$$

همچنین فرض کنید که کاربران شبکه به گونه‌ای شماره‌گذاری شده باشند که [۳۴، ۳۵]:

$$|H_1| \geq |H_2| \geq \dots \geq |H_N| \quad (۷)$$

به این معنی که کاربران شبکه برحسب شرایط کانال به صورت نزولی مرتب شده باشند؛ به عنوان مثال کاربر شماره یک و N به ترتیب دارای قوی‌ترین و ضعیف‌ترین کانال در بین کاربران در شبکه می‌باشند. با توجه به اینکه در گیرنده برای کاهش اثر تداخل شبکه از تکنیک SIC استفاده شده است، تداخلی که در سیگنال هر کاربر ایجاد می‌شود ناشی از کاربرانی از شبکه است که دارای کانال قوی‌تری نسبت به آن کاربر هستند؛ به این معنی که اثر تداخل کاربران ضعیف‌تر در گیرنده توسط این تکنیک حذف می‌شود. در نتیجه، در مسیر فراسو، نرخ ارسال اطلاعات کاربر i برحسب Bits/seconds را می‌توان به فرم زیر نوشت.

$$R_i = W \log_2 \left(1 + \frac{P_i H_i}{1 + \sum_{j=1}^{i-1} p_j^c H_j} \right) \quad (۸)$$

محلی و هم به صورت ریموت، توسط سرور ابری و یا سرور لبه، اجرا شوند. تعداد بیت داده تولیدشده توسط کاربر i ، با θ_i نمایش داده می‌شود. پارامتر p_i^c بیانگر این است که کاربر i با چه احتمالی درخواست خود را از سرور دریافت می‌کند؛ بدیهی است که $0 \leq p_i^c \leq 1$. طبیعتاً، سرویس درخواستی کاربر موردنظر با احتمال $(1 - p_i^c)$ به صورت محلی، انجام می‌شود. از طرفی، سرویس درخواستی توسط هر کاربر که به سرور ارسال می‌شود نیز دارای توزیع پواسن با نرخ متوسط $p_i^c \lambda_i$ است که به عنوان نرخ تخلیه^{۳۳} برای کاربر i در نظر گرفته می‌شود. بدیهی است که قسمتی از پردازش که به صورت محلی انجام می‌شود، دارای توزیع پواسن با نرخ متوسط $\lambda_i (1 - p_i^c)$ هست که به عنوان نرخ پردازش محلی در نظر گرفته می‌شود. پارامتر u_i^M به عنوان توان محاسباتی کاربر i در نظر گرفته می‌شود. از طرفی، l_i^M بیانگر این است که چند درصد از ظرفیت واحد پردازنده مرکزی^{۳۴} (CPU) کاربر i اشغال شده است. در واقع می‌توان گفت، $l_i^M = 1$ نشان‌دهنده این است که کل ظرفیت CPU کاربر i اشغال شده است و بالعکس، $l_i^M = 0$ بیان می‌کند که ظرفیت CPU کاربر i کاملاً خالی است. پارامترهای کلیدی مورد استفاده در این مقاله، به صورت خلاصه در جدول ۱ معرفی شدند.

۳-۱- پردازش محلی

با توجه اینکه مدل صف در سمت کاربران شبکه به فرم $M/M/1$ در نظر گرفته شده است، میزان زمان موردنیاز برای پردازش محلی سرویس مطلوب برای کاربر i از رابطه زیر قابل محاسبه است [۳۴]:

$$T_i^M(p_i^c) = \frac{1}{u_i^M (1 - l_i^M) - (1 - p_i^c) \lambda_i} \quad (۱)$$

انرژی موردنیاز برای پردازش محلی سرویس مطلوب هر کاربر نیز با رابطه زیر قابل بیان است:

$$E_i^M(p_i^c) = \kappa_i T_i^M(p_i^c) = \kappa_i \frac{1}{u_i^M (1 - l_i^M) - (1 - p_i^c) \lambda_i} \quad (۲)$$

که در رابطه فوق، ضریب κ_i بیانگر توانی است که هر کاربر در حالت پردازش محلی مصرف می‌کند و به ماهیت طبیعی دستگاه کاربران بستگی دارد. برای سهولت، فرض می‌شود که این پارامتر در طول پردازش و دریافت سرویس مطلوب، ثابت باشد.

۳-۲- انتقال در مسیر فراسو

اگر سیگنال ارسالی توسط هر کاربر با نماد χ_i نمایش داده شود به طوری که $\mathbf{E}[|\chi_i|^2] = 1$. کل کاربران شبکه برای ارسال داده خود به ایستگاه پایه از منبع طیفی مشترک استفاده می‌کنند. کاربران برای ارسال داده خود از کدگذاری برهم‌نهی استفاده می‌کنند؛ همچنین ایستگاه پایه از تکنیک SIC برای کاهش اثر تداخل سیگنال دریافتی استفاده می‌کند. سیگنالی که ایستگاه پایه توسط کاربران شبکه دریافت می‌کند را به صورت زیر می‌توان نشان داد [۳۴، ۳۵].

$$r^{BS} = \sum_{i=1}^N \sqrt{P_i h_i} \chi_i + n_i \quad (۳)$$

$$C(c, \rho^F) = \frac{\left(\frac{c\rho^F}{c!}\right)\left(\frac{1}{1-\rho^F}\right)}{\sum_{k=0}^{c-1} \frac{(c\rho^F)^k}{k!} + \left(\frac{c\rho^F}{c!}\right)\left(\frac{1}{1-\rho^F}\right)} \quad (16)$$

بعد از اتمام پردازش در سمت سرور لبه، میزان زمان موردنیاز برای ارسال نتایج به سمت کاربران شبکه به صورت زیر قابل بیان است [۳۴]:

$$T_b^F(\lambda_p^F) = \frac{1}{u_b^F - \lambda_p^F} \quad (17)$$

که در آن u_b^F بیانگر نرخ ارسال اطلاعات در سرور لبه است.

۳-۴- پردازش در سرور ابری

همان‌طور که قبلاً هم بیان شد، در صورتی که سرور لبه دچار اضافه‌بار شود، برای جلوگیری از تأخیر نامطلوب، درصدی از پردازش‌های ارسالی توسط کاربران، به سمت سرور ابری که در فاصله دورتری نسبت به کاربران قرار گرفته است؛ ارسال می‌کند. ارتباط بین این دو سرور از طریق لینک بک‌هاال است که میزان تأخیر ایجادشده در این مسیر را با T^O نشان داده و در این مقاله مقداری ثابت فرض می‌شود. همچنین همان‌طور که قبلاً هم بیان شد در این مقاله برای سرور ابری مدل $M/M/\infty$ در نظر گرفته شده است؛ از طرفی در عمل میزان ظرفیت محاسباتی و ذخیره‌سازی در سمت سرور ابری بسیار بیشتر از سرور لبه است و با توجه به نوع صف در نظر گرفته شده در این سرور میزان تأخیر ناشی از انتظار درخواست کاربران در صف بافر بسیار ناچیز و کاملاً قابل‌اغماض است. پارامتر u^{CC} بیانگر نرخ سرویس‌دهی در سرور ابری را نشان می‌دهد. با توجه به این توضیحات و مدل در نظر گرفته شده، میزان تأخیر ایجادشده شامل زمان موردنیاز برای پردازش در سرور ابری و تأخیر ایجادشده در لینک بک‌هاال با رابطه زیر قابل محاسبه است [۳۶]:

$$T_{wait}^{CC} = T^O + \frac{1}{u^{CC}} \quad (18)$$

بعد از اتمام پردازش در سرور ابری، نتایج باید دوباره از طریق ایستگاه پایه به سمت کاربران شبکه منتقل شوند. میزان تأخیر ناشی از انتظار نتایج خروجی در صف بافر برای ارسال به سمت کاربر موردنظر نیز به فرم زیر قابل بیان است [۳۶]:

$$T_b^{CC}(p_i^C) = \frac{1}{u_b^{CC} - (\lambda_{total}^M - \lambda_p^F)} \quad (19)$$

۳-۵- انتقال در مسیر فروسو

در این مقاله از محاسبه میزان انرژی و تأخیر ایجادشده در هنگام دریافت نتایج خروجی در مسیر فروسو صرف‌نظر شده است. چون در بسیاری از برنامه‌های کاربردی میزان داده حاصل از خروجی پردازش در سرور بسیار کمتر از داده ارسالی است و قابل صرف‌نظر کردن است [۳۷-۳۵]. طبق (۲) و (۱۱) می‌توان میزان انرژی مصرفی برای کاربر i را به صورت زیر تعریف کرد:

$$E_i(p_i^C, P_i) = (1 - p_i^C)E_i^M(p_i^C, P_i) + p_i^C E_i^S(p_i^C, P_i) \quad (20)$$

همچنین، میزان تأخیر ایجادشده برای دریافت سرویس هر کاربر را

که در رابطه فوق، W پهنای باند کل شبکه است. همچنین P_i نیز میزان توانی است که کاربر i صرف ارسال اطلاعات به ایستگاه پایه می‌کند به طوری که $0 \leq P_i \leq P_i^{\max}$ هم حداکثر میزان توان موجود در هر دستگاه است. h_i نیز بیانگر ضریب کانال بین کاربر i و ایستگاه پایه و N_0 نیز توان نویز موجود در شبکه است. با توجه به (۹)، زمان لازم برای ارسال اطلاعات به سمت ایستگاه پایه به صورت زیر است [۳۶]:

$$T_i^t(p_i^C, P_i) = \frac{P_i^C \lambda_i \theta_i}{R_i} \quad (10)$$

همچنین، میزان انرژی لازم که صرف انتقال این اطلاعات به سمت ایستگاه پایه می‌شود عبارت است از [۳۶]:

$$E_i^S(p_i^C, P_i) = P_i T_i^t(p_i^C, P_i) = P_i \frac{P_i^C \lambda_i \theta_i}{R_i} \quad (11)$$

۳-۳- پردازش در سرور لبه

سرور ابری تعبیه شده در ایستگاه پایه قادر است که به تعدادی از کاربران شبکه با حفظ کیفیت سرویس، سرویس‌دهی کند؛ اما در صورتی که تعداد این درخواست‌ها از سمت کاربران خیلی زیاد باشد، سرور لبه به‌ناچار، میزانی از این درخواست‌ها را به سرور ابری، برای پردازش ارسال می‌کند. بر همین اساس فرض می‌شود که سرور لبه موجود، متشکل از c سرور همگن باشد. نرخ سرویس‌دهی در هر یک از سرورهای لبه با u^F نمایش داده می‌شود. همچنین حداکثر حجم بار قابل تحمل برای هر سرور را نیز با λ_{max}^F تعریف می‌کنیم. هدف از تعریف پارامتر λ_{max}^F این است که از تأخیر ناشی از اضافه‌بار در سرور لبه که منجر به افت کیفیت سرویس می‌شود، خودداری کنیم. هنگامی که کاربران شبکه درخواست خود را به سرور ابری ارسال می‌کنند، کل این درخواست‌ها باهم جمع می‌شود، لذا نرخ درخواست کل در سرور لبه با توجه به ویژگی توزیع پواسن از رابطه زیر قابل محاسبه است:

$$\lambda_{total}^M = \sum_{i=1}^N p_i^C \lambda_i \quad (12)$$

به این ترتیب، درصدی از پردازش کل که توسط سرور لبه، قابل پاسخ‌گویی است با ψ^F نمایش داده می‌شود. بر این اساس، نرخ واقعی پردازش در سمت سرور لبه به صورت زیر قابل تعریف است:

$$\lambda_p^F = \psi^F \lambda_{total}^M \quad (13)$$

بنابراین، میزان تأخیر زمانی ایجادشده در سرور لبه که شامل زمان موردنیاز برای پردازش سرویس در سرور لبه و زمان انتظار در صف بافر است، با توجه به مدل صف تعریف شده در سرور لبه و همچنین با توجه به رابطه ارلانگ، از رابطه زیر قابل بیان است [۳۴]:

$$T_{wait}^F(\lambda_p^F) = \frac{C(c, \rho^F)}{cu^F - \lambda_p^F} + \frac{1}{u^F} \quad (14)$$

که در آن

$$\rho^F = \frac{\lambda_p^F}{cu^F} \quad (15)$$

هدف استفاده شده است. حال برای این منظور مجموعه وزن توابع هدف به صورت $\alpha = \{\alpha_1, \alpha_2, \alpha_3\}$ معرفی می‌شده است به طوری که $\alpha_1 + \alpha_2 + \alpha_3 = 1$. به عنوان مثال اگر در شبکه، مینیمم کردن انرژی از اهمیت بیشتری برخوردار باشد، وزنی که به آن اختصاص پیدا می‌کند باید بیشتر باشد. با این روش مسئله بهینه‌سازی چندهدفه فوق می‌تواند به صورت یک مسئله تک هدفه واحد، به صورت زیر تبدیل شود.

$$P2: \min_{\{p^C, P, \psi^F\}} C(p^C, P, \psi^F) \\ \text{subject to: } C.1 \sim C.6$$

که در رابطه فوق، تابع $C(p^C, P, \psi^F)$ به صورت ترکیبی وزن دار از توابع هدف مورد نظر طبق رابطه (۲۵) قابل بیان است. قابل ذکر است که توابع هدف تعریف شده در مسئله P1، دارای معادله ابعادی^{۲۵} یکسان نیستند. در نتیجه با تعریف ضرایب ثابت \bar{E} ، \bar{T} و \bar{M} ، تابع هدف مورد نظر در مسئله P2 نرمالیزه و بدون دیمانسیون خواهد شد. با توجه به مسئله P2، متغیرهای بهینه‌سازی تعریف شده در مسئله عبارت‌اند از P_i ، p_i^C و ψ^F که این پارامترها با هم متقابلاً در ارتباط هستند، به طوری که با افزایش p_i^C ، میزان P_i که انرژی مورد نیاز برای انتقال داده به ایستگاه پایه است، افزایش می‌یابد، همچنین با افزایش P_i ، میزان زمان مورد نیاز برای داده به ایستگاه پایه کاهش می‌یابد و برعکس. به همین دلیل این پارامترها باید به گونه‌ای انتخاب شوند که تابع هدف تعریف شده در مسئله P2، به حداقل مقدار ممکن برسد.

۵- حل مسئله به روش SCA

مسئله P3 و P4 به دلیل نامحدب بودن تابع هدفشان، غیرمحدب هستند؛ لذا برای حل این مسئله بهینه‌سازی از روشی موسوم به SCA که در [۳۸] مطرح شده استفاده می‌شود. در این روش برای به دست آوردن جوابی از مسئله نامحدب از یک الگوریتم تکراری^{۲۶} استفاده می‌شود که در هر تکرار از الگوریتم برای مسئله نامحدب، تقریب محدب را به دست می‌آورد که این تقریب باید به گونه‌ای باشد که شرایط ذکر شده در [۳۸] را ارضا کند.

در مسئله P2 همان‌طور که دیده می‌شود، قیدها کاملاً محدب هستند و لذا در هر تکرار از الگوریتم تنها نیاز است که تقریب محدب مناسبی برای تابع هدف به دست آورده شود. در قسمت بعدی به محاسبه تقریب محدب پرداخته می‌شود.

۵-۱- تقریب محدب تابع هدف

برای محاسبه تقریب محدب تابع هدف، ابتدا به منظور سهولت در بیان ریاضی مسئله، کل متغیرهای بهینه‌سازی، با مجموعه Z نمایش داده شده است. به عبارت دیگر $Z = (p^C, P, \psi^F)$ که در این تعریف مسئله P3، حول نقطه $Z(v)$ با $\bar{C}(Z, Z(v))$ نمایش دهیم، این تقریب باید در شروط زیر صدق کند:

به صورت زیر می‌توان نوشت:

$$T_i(p_i^C, P_i) = (1 - p_i^C) T_i^M(p_i^C, P_i) + p_i^C T_i^F(p_i^C, P_i) + p_i^C \psi^F (T_{wait}^F + T_b^F) + p_i^C (1 - \psi^F) (T_{wait}^{CC} + T_b^{CC}) \quad (21)$$

بر اساس روابط فوق می‌توان گفت که متوسط میزان انرژی و تأخیر ایجاد شده در شبکه طبق روابط (۲۳) و (۲۴) قابل بیان است.

علاوه بر این، کاربران در قبال استفاده از منابع موجود در سرورها ملزم به پرداخت هزینه می‌باشند که میزان واحد این هزینه برای سرور لبه با r^F و برای سرور ابری با r^{CC} نمایش داده می‌شود و به طور کلی $r^{CC} > r^F$. از طرفی میزان استفاده از منابع موجود در سرورها، تعیین کننده هزینه کل است که برای این منظور نرخ پردازش، به عنوان معیاری برای میزان استفاده از منابع ابری در نظر گرفته شده است و بر این اساس می‌توان میزان متوسط هزینه استفاده از خدمات سرورها را به صورت رابطه زیر نوشت:

$$M(p_i^C) = \frac{1}{N} \{ r^F \lambda_p^F(p_i^C) + r^{CC} [\lambda_{Total}^M(p_i^C) - \lambda_p^F(p_i^C)] \} \quad (22)$$

۴- بیان مسئله به فرم مسئله بهینه‌سازی

با توجه به روابط به دست آمده برای انرژی، تأخیر و هزینه مورد انتظار برای پردازش سرویس هر کاربر، می‌توان مسئله تخلیه بهینه را به صورت مینیمم‌سازی توأم مجموع انرژی، تأخیر و هزینه مصرفی توسط کل کاربران شبکه به بیان ریاضی به فرم مسئله بهینه‌سازی زیر بیان کرد.

$$P1: \min_{\{p^C, P, \psi^F\}} \{E(p_i^C, P_i), T(p_i^C, P_i), M(p_i^C)\} \\ \text{subject to}$$

$$C.1) \quad (1 - p_i^C) \lambda_i < u_i^M (1 - l_i^M)$$

$$C.2) \quad \lambda_p^F < cu^F$$

$$C.3) \quad \lambda_p^F < u_b^F$$

$$C.4) \quad \lambda_{Total}^M - \lambda_p^F < u_b^{CC}$$

$$C.5) \quad 0 \leq P_i \leq P_i^{\max} \quad \forall i \in \mathcal{N},$$

$$C.6) \quad 0 \leq p_i^C \leq 1 \quad \forall i \in \mathcal{N},$$

قیود C.1، C.2، C.3 و C.4 به ترتیب از روابط (۱)، (۱۵)، (۱۸) و (۲۰) استخراج شده‌اند. قید C.1 بیانگر این است که نرخ درخواست ورودی برای پردازش محلی نباید از نرخ پردازش در دستگاه کاربر مورد نظر بیشتر باشد. قید C.2 نشان می‌دهد که نرخ درخواست در سمت سرور لبه نباید از نرخ سرویس دهی در این سرور بیشتر باشد. قید C.3 و C.4 هم به ترتیب بیان کننده این است که نرخ پردازش از نرخ انتقال در سرور لبه و سرور ابری نباید تجاوز کند. مسئله فوق، یک مسئله بهینه‌سازی غیرخطی نامحدب با چندین تابع هدف است. برای حل مسائل بهینه‌سازی چندهدفه، دودسته روش وجود دارد که روش اول همان روش‌های متداول بوده که شامل روش‌های برنامه‌ریزی خطی، روش‌های وزن دار کردن توابع هدف و غیره بوده و روش‌های دسته دوم مرتبط با روش‌های ابتکاری مانند تئوری بازی است. در این مقاله برای حل این مسئله از روش‌های متداول، به طور خاص وزن دار کردن توابع

$$E(p_i^c, P_i) = \frac{1}{N} \sum_{i=1}^N E_i(p_i^c, P_i) = \frac{1}{N} \left\{ \sum_{i=1}^N [(1-p_i^c) E_i^M(p_i^c, P_i) + p_i^c E_i^S(p_i^c, P_i)] \right\} = \frac{1}{N} \left\{ \sum_{i=1}^N \left[\frac{\kappa_i (1-p_i^c)}{u_i^M (1-l_i^M) - (1-p_i^c) \lambda_i} + p_i^c \frac{P_i p_i^c \lambda_i \theta_i}{R_i} \right] \right\} \quad (23)$$

$$T(p_i^c, P_i) = \frac{1}{N} \sum_{i=1}^N T_i(p_i^c, P_i) = \frac{1}{N} \sum_{i=1}^N \left\{ \begin{aligned} & (1-p_i^c) T_i^M(p_i^c, P_i) + p_i^c T_i^I(p_i^c, P_i) \\ & + p_i^c \psi^F (T_{wat}^F + T_b^F) + p_i^c (1-\psi^F) (T_{wat}^{CC} + T_b^{CC}) \end{aligned} \right\} = \frac{1}{N} \sum_{i=1}^N \left\{ \begin{aligned} & \frac{(1-p_i^c)}{u_i^M (1-l_i^M) - (1-p_i^c) \lambda_i} + p_i^c \frac{P_i p_i^c \lambda_i \theta_i}{R_i} \\ & + p_i^c \psi^F \left(\frac{C(c, \rho^F)}{cu^F - \lambda_p^F} + \frac{1}{u^F} + \frac{1}{u_b^F - \lambda_p^F} \right) \\ & + p_i^c (1-\psi^F) \left(T^O + \frac{1}{u^{CC}} + \frac{1}{u_b^{CC} - (\lambda_{total}^M - \lambda_p^F)} \right) \end{aligned} \right\} \quad (24)$$

$$C(p^c, P, \psi^F) \triangleq \alpha_1 \frac{E(p^c, P)}{\bar{E}} + \alpha_2 \frac{T(p^c, P, \psi^F)}{\bar{T}} + \alpha_3 \frac{M(p^c, \psi^F)}{\bar{M}} = \alpha_1 \frac{1}{N} \frac{1}{\bar{E}} \left\{ \sum_{i=1}^N \left[(1-p_i^c) \frac{\kappa_i}{u_i^M (1-l_i^M) - (1-p_i^c) \lambda_i} + p_i^c \frac{P_i p_i^c \lambda_i \theta_i}{R_i} \right] \right\} \\ + \alpha_2 \frac{1}{N} \frac{1}{\bar{T}} \sum_{i=1}^N \left\{ \begin{aligned} & (1-p_i^c) \frac{1}{u_i^M (1-l_i^M) - (1-p_i^c) \lambda_i} + p_i^c \frac{P_i p_i^c \lambda_i \theta_i}{R_i} \\ & + p_i^c \left[\frac{\lambda_{max}^F}{\sum_{i=1}^N p_i^c \lambda_i} \left(\frac{C(c, \rho^F)}{cu^F - \lambda_{max}^F} + \frac{1}{u^F} + \frac{1}{u_b^F - \lambda_{max}^F} \right) \right] \\ & + p_i^c \left(1 - \frac{\lambda_{max}^F}{\sum_{i=1}^N p_i^c \lambda_i} \right) \left(T^O + \frac{1}{u^{CC}} + \frac{1}{u_b^{CC} - (\lambda_{total}^M - \lambda_{max}^F)} \right) \end{aligned} \right\} + \alpha_3 \frac{1}{N} \frac{1}{\bar{M}} \left[r^F \lambda_{max}^F + r^{CC} \left[\sum_{i=1}^N p_i^c \lambda_i - \lambda_{max}^F \right] \right] \quad (25)$$

$$\bar{C}(Z, Z(v)) \triangleq \frac{\gamma_{pc}}{2} \|p^c - p^c(v)\|^2 + \frac{\gamma_p}{2} \|P - P(v)\|^2 \quad (27)$$

$$lin_tem(P) = \sum_{j=1, j \neq i}^N \left\{ \begin{aligned} & (p_j^c)^2 \lambda_j \theta_j (1+P_j) \times \\ & \frac{d}{dP_i} \left(\frac{1}{R_j} \right) (P_i - P_i(v)) \end{aligned} \right\} \quad (28)$$

در رابطه بالا سه جمله اول سمت راست تساوی به منظور محذب سازی تابع هدف است؛ جمله $\bar{C}(Z, Z(v))$ به منظور قویاً محذب ساختن تقریب تابع هدف و جمله $lin_tem(P)$ نیز برای خطی سازی به عبارت اضافه شده است.

۲-۵- تقریب محذب مسئله

با محاسبه تقریب‌های محذب تابع هدف حول نقطه شدنی $Z(v)$ می‌توان به جای حل مسئله نامحذب $P2$ ، با الگوریتم تکراری SCA به حل تقریب محذب آن پرداخت.

$$P3: \min_{(p^c, P, \psi^F)} \bar{C}(p^c, P, \psi^F) \\ \text{subject to : } C.1 \sim C.6$$

با توجه به فرم مسئله بهینه‌سازی $P3$ ، دیده می‌شود که این مسائل دارای فرم پیوسته و محذب است. با بکارگیری روش SCA و استفاده

(۱) تقریب $\bar{C}(\bullet, Z(v))$ روی مجموعه \mathcal{K} باید پیوسته و قویاً محذب باشد.

$$\nabla_{p^c} \bar{C}(Z(v), Z(v)) = \nabla_{p^c} C(Z(v)) \quad (2)$$

$$\nabla_P \bar{C}(Z(v), Z(v)) = \nabla_P C(Z(v))$$

$$\nabla_{\psi^F} \bar{C}(Z(v), Z(v)) = \nabla_{\psi^F} C(Z(v))$$

(۳) تابع $\nabla_Z \bar{C}(\bullet, \bullet)$ باید دارای پیوستگی لیپ‌شیتز باشد

با توجه به شرایط ذکر شده می‌توان این تقریب را از رابطه زیر

محاسبه کرد طوری که شرایط فوق را ارضا کند.

(۲۶)

$$\bar{C}(Z, Z(v)) = \alpha_1 \frac{1}{N} \frac{1}{\bar{E}} \left\{ \begin{aligned} & \sum_{i=1}^N E_i(p_i^c, P_i(v)) \\ & + \sum_{i=1}^N E_i(p_i^c(v), P_i) \end{aligned} \right\} \\ + \alpha_2 \frac{1}{N} \frac{1}{\bar{T}} \sum_{i=1}^N \left\{ \begin{aligned} & T_i(p_i^c, P_i(v), \psi^f) + T_i(p_i^c(v), P_i, \psi^f) \\ & + T_i(p_i^c, P_i, \psi^f(v)) \end{aligned} \right\} \\ + \alpha_3 \frac{1}{M} \sum_{i=1}^N M_i(p_i^c, \psi^f) + \bar{C}(Z, Z(v)) \\ + lin_tem(P),$$

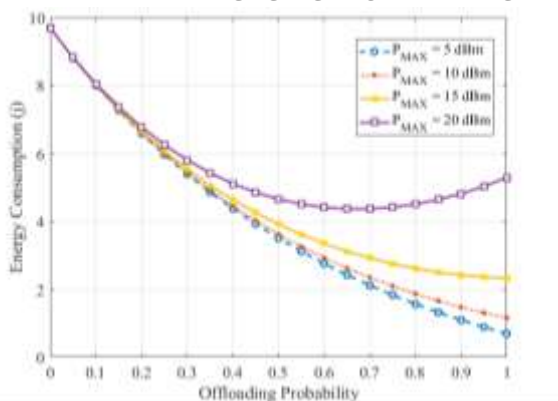
که در آن

جدول شماره ۲: پارامترهای مورداستفاده در شبیه‌سازی

N	r^F	r^{CC}	κ_i	P_i^{\max} (dBm)	c	T^O	l_i^M
۴	۰/۰۰۱	۰/۰۱	۱۶	۲۳	۴	۲	۰/۳
u^F (MIPS)		u_b^F (MIPS)		u^{CC} (MIPS)		u_b^{CC} (MIPS)	
۱۰		۱۰		۳۰۰		۳۰۰	
λ_i (MIPS)		u_i^M (MIPS)		θ_i (Mbits)			
~uniform (۱،۲)		~uniform (۴/۴، ۴/۸)		~uniform (۲،۵)			

شکل ۳، تأثیر احتمال تخلیه بر میزان تأخیر ایجاد شده در شبکه برای مقادیر مختلف توان ارسالی را نشان می‌دهد. همان‌طور که در این شکل دیده می‌شود، به ازای مقدار مشخص از توان ارسالی، با افزایش احتمال تخلیه کاربر، میزان تأخیر دریافت سرویس از سرور لبه افزایش می‌یابد، به این دلیل که با افزایش احتمال تخلیه، طول بافر موجود در سرور لبه افزایش پیدا کرده و این امر منجر به این می‌شود که کاربران شبکه، سرویس خود را با تأخیر بیشتری دریافت کنند. در شکل ۴، به تأثیر حداکثر توان ارسالی بر میزان تأخیر ایجاد شده به ازای مقادیر مختلف احتمال تخلیه پرداخته شده است. با توجه به شکل دیده می‌شود که با افزایش میزان توان ارسالی، میزان تأخیر در دریافت سرویس کاهش می‌یابد، زیرا با افزایش میزان توان ارسالی توسط هر کاربر میزان نرخ ارسال اطلاعات افزایش پیدا کرده که با توجه به رابطه (۱۰) این امر موجب کاهش تأخیر در شبکه می‌شود. از طرفی دیده می‌شود که به ازای مقدار ثابت توان ارسالی، با افزایش احتمال تخلیه میزان تأخیر در شبکه افزایش پیدا می‌کند که این امر نیز بدین خاطر است که با افزایش احتمال تخلیه، میزان درخواست‌های بیشتری به سمت سرور لبه ارسال می‌شود که این امر خود منجر به افزایش طول صف در سمت سرور شده و موجب افزایش تأخیر در شبکه می‌شود.

تأثیر پارامتر حداکثر توان ارسالی بر انرژی مصرف شده در شبکه در شکل ۵ بررسی شده است. همان‌طور که در این شکل قابل مشاهده است در ابتدا با افزایش میزان حداکثر توان ارسالی کاربر، مصرفی در شبکه کاهش می‌یابد اما از یک مقدار مشخص به بعد میزان انرژی مصرفی شبکه در سطح ثابتی باقی می‌ماند.



شکل ۲: تأثیر احتمال تخلیه بر میزان انرژی مصرف شده در شبکه

از الگوریتم نقاط داخلی در هر تکرار از الگوریتم SCA، اثبات می‌شود که دو مسئله مطرح شده، به نقاط بهینه سراسری همگرا می‌شود [۳۸]. الگوریتم SCA به صورت خلاصه در زیر بیان شده است.

Algorithm 1: SCA Solution for P5 and P6**Initialization:** $\mathbf{Z}(0) \in \mathcal{Z}; \{\gamma(v)\}_v \in (0, 1]; v = 0;$ **1:** If $\mathbf{Z}(v)$ satisfies the termination criterion, stop.**2:** Compute $\tilde{\mathbf{Z}}(\mathbf{Z}(v))$ from P5 and P6.**3:** Set $\mathbf{Z}(v+1) = \mathbf{Z}(v) + \gamma(v) (\tilde{\mathbf{Z}}(\mathbf{Z}(v)) - \mathbf{Z}(v))$.**4:** $v \leftarrow v + 1$, and return to step 1.**Output:** $\mathbf{Z} = (p^c, P, \psi^f)$.

در این الگوریتم $\mathbf{Z}(0)$ نقطه شروع برای الگوریتم است که از ناحیه شدنی مسئله یعنی \mathcal{Z} انتخاب می‌شود. همچنین پارامتر γ تعیین کننده میزان گام الگوریتم است به طوری که:

$$\gamma(v) = \gamma(v-1)(1 - \alpha\gamma(v-1)) \quad (29)$$

$$\gamma(0) \in (0, 1] \quad (30)$$

از طرفی $\alpha \in (0, 1/\gamma(0))$. شرط توقف الگوریتم نیز به صورت $|\tilde{\mathbf{C}}(\mathbf{Z}(v+1)) - \tilde{\mathbf{C}}(\mathbf{Z}(v))| \leq \delta$ است که در این رابطه δ تعیین کننده میزان دقت الگوریتم است.

۶ - نتایج شبیه‌سازی

در این قسمت، به ارائه نتایج حاصل از شبیه‌سازی پرداخته خواهد شد. پارامترهای مورداستفاده در شبیه‌سازی در جدول شماره ۲ قابل مشاهده است. لازم به ذکر است که برای شبیه‌سازی از نرم‌افزار متلب ۲۰۱۴ نسخه a در این مقاله استفاده شده است.

در ابتدا، در شکل‌های ۲ تا ۵ به بررسی تأثیر پارامترهای احتمال تخلیه (p_i^c) و توان ارسالی (P_i) بر انرژی مصرف شده در شبکه و تأخیر دریافت سرویس از سرور پرداخته شده است. در این قسمت برای سهولت، فرض می‌شود سیستم تک کاربره است، چون هدف اصلی در این قسمت بررسی تأثیر پارامترهای مذکور در شبکه است. در شکل ۲، به بررسی تأثیر احتمال تخلیه بر میزان مصرف شده در شبکه به ازای مقادیر مختلف توان ارسالی پرداخته شد. همان‌طور که در شکل دیده می‌شود، به ازای مقادیر ثابت توان ارسالی، با افزایش احتمال تخلیه، میزان انرژی مصرف شده در شبکه کاهش پیدا می‌کند و این امر به این دلیل است که با افزایش احتمال تخلیه به ازای مقدار ثابت توان ارسالی، میزان درخواست بیشتری به سرور لبه ارسال می‌شود و منجر به این می‌شود که انرژی کمتری نسبت به حالت پردازش محلی صرف شود. همان‌طور که در این شکل پیداست، میزان انرژی مصرفی در حالت پردازش در سرور به مراتب از حالت پردازش محلی کمتر است. به عنوان مثال، برای احتمال تخلیه ۰.۸، و توان ارسالی ۵ dBm میزان مصرف انرژی نسبت به حالت پردازش محلی به میزان ۸۰ درصد کاهش یافته است که این نشان‌دهنده مزیت استفاده از پردازش ابری نسبت به پردازش محلی است.

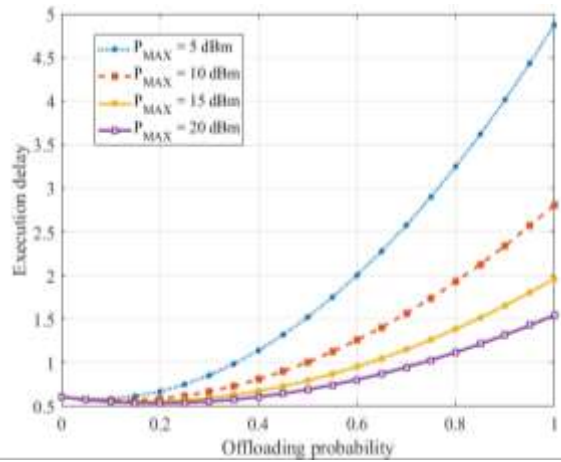
می‌شود که همان‌طور که قبلاً در شکل شماره ۳ بیان شد پردازش در سمت سرور لبه نسبت به پردازش محلی انرژی کمتری مصرف می‌کند.

با مقایسه نتایج حاصل از منحنی‌ها لزوم ایجاد یک توازن بین انرژی مصرف‌شده در شبکه، تأخیر در دریافت سرویس از سرور و میزان هزینه پرداختی در قبال این سرویس‌ها به‌خوبی مشهود است. برای قسمت بعدی شبیه‌سازی، فرض می‌شود تعداد کاربران شبکه برابر با ۴ باشد و با این فرض به حل مسئله بهینه‌سازی چندهدفه پرداخته شده است. در جدول شماره ۳ مقادیر بهینه توان ارسالی، احتمال تخلیه و درصد حجم پردازش در سرور لبه برای سه وزن مختلف قابل‌مشاهده است که در حالت اول فرض شد میزان انرژی مصرفی کل شبکه دارای اهمیت بیشتری نسبت به تأخیر و هزینه مصرفی باشد. با این وزن دهی به توابع هدف، همان‌طور که دیده می‌شود برای کاربر شماره ۱، ۸۰ درصد پردازش به سرور ارسال می‌شود که این امر موجب می‌شود که حجم کمی از پردازش به‌صورت محلی اجرا شود و انرژی کمتری مصرف شود. از طرفی برای حالتی که بیشترین وزن به تأخیر داده‌شده است برای کاربر شماره ۱ دیده می‌شود درصد کمتری از پردازش نسبت به حالت قبلی به سرور ارسال می‌شود. از طرف دیگر چون نحوه دسترسی به طیف رادیویی به‌صورت نامتعاد در نظر گرفته‌شده است و کاربران بر اساس شرایط کانال شماره بندی شدند (کاربر شماره ۱ دارای قوی‌ترین کانال و کاربر شماره ۴ دارای ضعیف‌ترین کانال در شبکه است) دیده می‌شود کاربر شماره ۱ که دارای قوی‌ترین کانال است نسبت به سایر کاربران دارای توان ارسالی کمتری است و ضعیف‌ترین کاربر شبکه (کاربر شماره ۴) دارای بیشترین توان برای ارسال است.

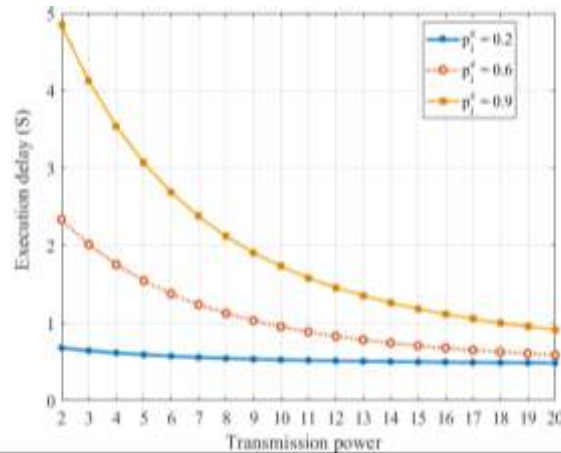
در شکل‌های ۶، ۷ و ۸ اثر افزایش تعداد کاربران شبکه بر انرژی مصرف‌شده در شبکه، تأخیر دریافت سرویس و تابع هزینه کلی ایجادشده در دو حالت دسترسی به منبع طیفی را به ترتیب نشان می‌دهد. همان‌طور که از این شکل‌ها نیز پیداست، استفاده از دسترسی نامتعاد، نسبت به حالت تعاد دارای عملکرد بهتر بوده و میزان هزینه کل شبکه را به مقدار قابل‌توجهی کاهش می‌دهد. به‌عنوان مثال، در شکل ۶، ۷ و ۸ در صورتی که تعداد کاربران شبکه برابر با ۱۵ باشد، استفاده از دسترسی نامتعاد نسبت به تعاد، میزان ۵۴ درصد انرژی مصرفی شبکه، ۶۶ درصد تأخیر پردازش و حدود ۵۷ درصد هزینه کل شبکه را به ترتیب کاهش می‌دهد که نشان‌دهنده بهبود عملکرد سیستم در صورت استفاده از دسترسی نامتعاد را نشان می‌دهد. و دارای عملکرد بهتری است.

جدول شماره ۳: مقادیر بهینه توان ارسالی، احتمال تخلیه و درصد حجم پردازش در سرور لبه برای سه وزن مختلف

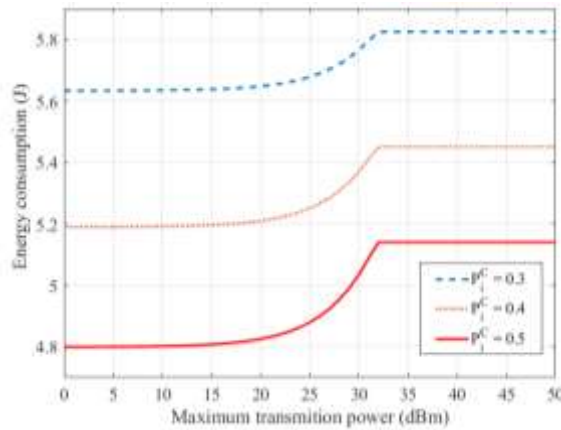
ψ^f	(P_4^*, p_4^{c*})	(P_3^*, p_3^{c*})	(P_2^*, p_2^{c*})	(P_1^*, p_1^{c*})	مجموعه وزن توابع هدف
۰/۲۹۷۲	(۰/۳۱۶۲، ۰/۶۹۱۶)	(۰/۱۵۷۹، ۰/۶۸۷۷)	(۰/۱۰۱۸، ۰/۸۲۳۶)	(۰/۰۷۰۵، ۰/۸۴۴۳)	(۰/۷/۰/۲۰/۱۱)
۰/۲۶۶۲	(۰/۳۱۶۲، ۰/۲۵۶۵)	(۰/۳۱۶۲، ۰/۳۰۵۹)	(۰/۳۱۶۲، ۰/۴۷۷۳)	(۰/۱۸۸۰، ۰/۴۹۲۷)	(۰/۲۰/۱۶۰/۱۲)
۰/۵۲۴۳	(۰/۳۱۶۲، ۰/۴۴۱۳)	(۰/۳۱۰۰، ۰/۵۰۷۰)	(۰/۲۰۳۷، ۰/۶۵۱۰)	(۰/۲۱۴۱، ۰/۶۶۵۸)	(۰/۲۰/۲۰/۱۶)



شکل ۳: تأثیر احتمال تخلیه بر میزان تأخیر پردازش برنامه

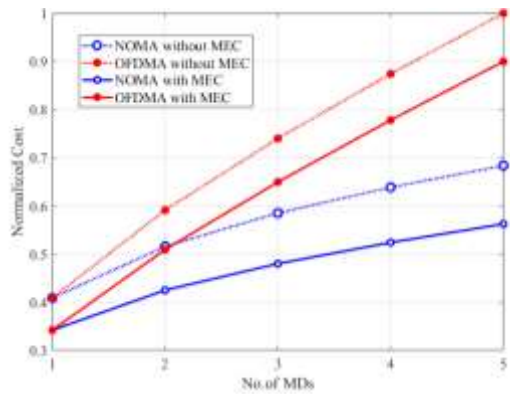


شکل ۴: تأثیر حداکثر توان ارسالی بر میزان انرژی مصرفی شبکه



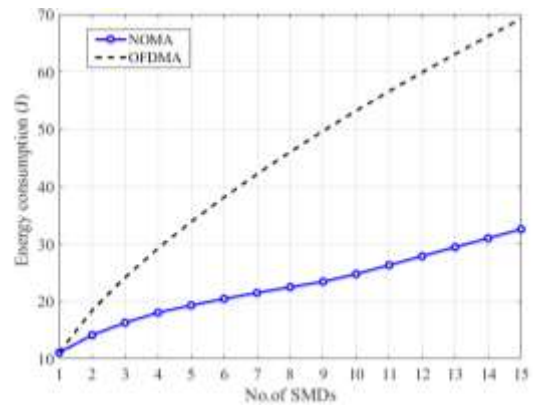
شکل ۵: تأثیر حداکثر توان ارسالی بر تأخیر پردازش برنامه

همچنین با ازای مقدار ثابتی از حداکثر توان ارسالی، با افزایش احتمال تخلیه میزان مصرف انرژی در شبکه کاهش پیدا می‌کند زیرا با افزایش احتمال تخلیه، میزان بیشتری از پردازش به سرور لبه ارسال

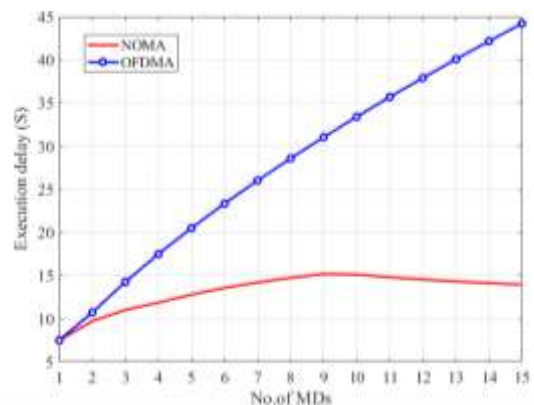


شکل ۱۰: مقایسه تابع هزینه نرمالیزه شده در حالت استفاده توأم از سرور ابری و لبه باحالت سرور ابری تنها

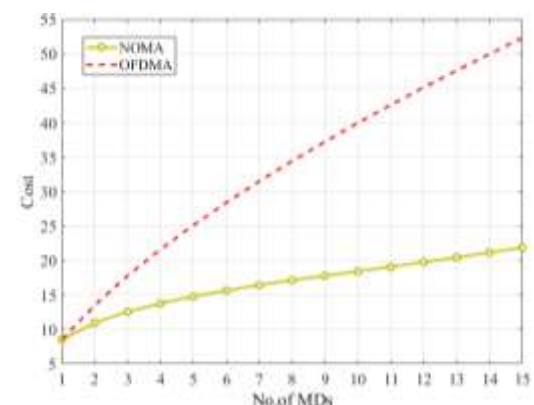
در شکل ۹ سعی شده کار پیشنهادی با دو حالت مختلف بررسی شود و تأثیر تخصیص بهینه توأم پارامترهای استفاده‌شده در مدل پیشنهادی بر هزینه کل شبکه برای چند سناریوی مختلف مورد بررسی قرار گیرد. در حالت اول فرض بر این است که میزان احتمال تخلیه داده برای تمامی کاربران ثابت بوده و با این فرض مقادیر بهینه حداکثر توان ارسالی و درصد حجم پردازش بین سرورها به گونه‌ای به دست آورده شده که میزان تابع هزینه کل حداقل شود، در حالت دیگر فرض بر این است که حداکثر توان ارسالی در سمت کاربران ثابت باشد و بقیه پارامترها به صورت بهینه به دست بیایند و در انتها تمام پارامترها، باهم به صورت بهینه به کاربران تخصیص داده می‌شود. طبق شکل ۹، دیده می‌شود تخصیص توأم این پارامترها نسبت به دو حالت دیگر، به مقدار قابل توجهی تابع هزینه تعریف‌شده در شبکه را کاهش می‌دهد. به عنوان مثال در حالتی که تعداد کاربران برابر با ۵ است؛ دیده می‌شود تخصیص بهینه توأم پارامترهای استفاده‌شده با دسترسی نامتعاد نسبت به حالتی که تنها احتمال تخلیه به صورت بهینه به کاربران شبکه اختصاص پیدا می‌کند؛ با فرض دسترسی متعادل، میزان هزینه کل شبکه در حدود ۴۷ درصد کاهش پیدا می‌کند. در شکل ۱۰ به بررسی مزیت مشارکت سرور ابری و لبه در شبکه پرداخته شد. در یک حالت فرض شد شبکه فقط شامل سرور ابری باشد و در حالت دوم فرض شد علاوه بر سرور ابری، سرور لبه نیز در شبکه وجود داشته باشند که این سرورها در ایستگاه‌های پایه تعبیه شدند. با این فرض در دو حالت دسترسی به منابع طیفی میزان هزینه نرمالیزه شده کل، برحسب تعداد کاربران شبکه در این شکل رسم شد. استفاده توأم سرور ابری و لبه و روش پیشنهادی میزان تابع هزینه شبکه به مقدار قابل توجهی کاهش پیدا می‌کند. مثلاً در این شکل، اگر تعداد کاربران شبکه برابر با ۵ باشد دیده می‌شود با فرض استفاده از دسترسی نامتعادل برای کاربران شبکه، زمانی که سرور ابری و لبه باهم مشارکت داشته باشند نسبت به حالتی که تنها سرور ابری در شبکه وجود داشته باشد هزینه کل شبکه به اندازه ۲۱ درصد کاهش پیدا می‌کند. همچنین دیده می‌شود با فرض در نظر گرفتن دسترسی متعادل برای کاربران و عدم حضور سرور لبه هزینه شبکه به میزان ۴۲ درصد نسبت به حالت اول افزایش پیدا می‌کند.



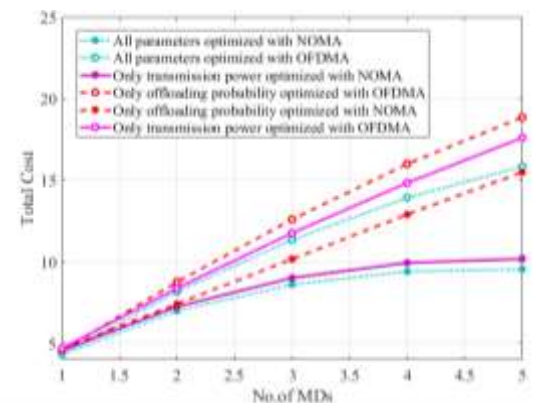
شکل ۶: تأثیر تعداد کاربران شبکه بر میزان مصرف انرژی در شبکه



شکل ۷: تأثیر تعداد کاربران شبکه بر مجموع تأخیر ایجاد شده در شبکه



شکل ۸: تأثیر تعداد کاربران شبکه بر میزان تابع هزینه کلی



شکل ۹: مقایسه تابع هزینه روش پیشنهادی با سایر روش‌های قبلی

۵ - نتیجه‌گیری و کارهای آینده

در این مقاله، با تعریف تابع هزینه مطلوب شامل ترکیب وزن‌دار از انرژی مصرف‌شده، تأخیر ایجادشده در شبکه برای دریافت سرویس و هزینه استفاده از خدمات سرورهای ابری و لبه، سعی بر مینیمم کردن آن تحت قیود مناسب، از جمله قید انرژی و تأخیر کردیم. فرض شد که علاوه بر سرور ابری، سرور لبه نیز در شبکه موجود باشد؛ به طوری که این دو سرور با مشارکت هم به درخواست کاربران پاسخ دهند. توان ارسالی، احتمال تخلیه و محل انجام پردازش به عنوان پارامترهای بهینه‌سازی تعریف شدند. در این مدل فرض شد که تخلیه پردازش کاربران به صورت جزئی باشد؛ یعنی قسمتی از پردازش به صورت محلی در دستگاه کاربر و قسمتی دیگر از آن در سرور انجام شود. همچنین فرض شد نحوه دسترسی کاربران به منابع طیفی شبکه به صورت نامتعاد باشد. برای مدل‌سازی مسئله با در نظر گرفتن بافر در سمت کاربران شبکه و سرورهای لبه و ابری، از روابط حاکم بر تئوری صف استفاده شده است، با مدل‌سازی مسئله به فرم یک مسئله بهینه‌سازی مناسب، به دلیل نامحدوب بودن مسئله مطرح شده، از روش SCA برای حل استفاده شد. این الگوریتم به دلیل محاسبه تقریب محدب مناسب برای مسئله در هر تکرار از الگوریتم دارای سرعت همگرایی بالایی بوده و به جواب بهینه مسئله نامحدب همگرا می‌شود. نتایج حاصل از شبیه‌سازی نشان داده است که با در نظر گرفتن مشارکت بین سرور ابری، سرور لبه، پردازش محلی، تخصیص بهینه پارامترهای تعریف شده و با به کارگیری دسترسی نامتعاد برای کاربران شبکه به طیف رادیویی نسبت به حالتی که مشارکت بین سرورها وجود ندارد و نحوه دسترسی کاربران به طیف رادیویی به صورت متعادل است، میزان انرژی مصرفی کل شبکه، تأخیر در دریافت سرویس و هزینه کلی شبکه به مقدار قابل توجهی کاهش پیدا کرده است. همچنین تأثیر تخصیص توأم پارامترهای تعریف شده در مسئله به صورت بهینه برای کاربران نسبت به حالات دیگر در شبیه‌سازی کاملاً مشخص شده است و دیده شد که با فرضیات وارد شده در مسئله، میزان انرژی مصرفی شبکه، تأخیر و هزینه کل شبکه به مقدار قابل توجهی کاهش پیدا می‌کند.

بررسی سیستم فوق با توپولوژی پیچیده‌تر و حالت چند سلولی، استفاده از پردازش دستگاه به دستگاه و حتی پردازش کلاستری، تعیین الگوی مناسب برای مرتب کردن کاربران بر حسب قید انرژی و توان و تعریف یک الگوریتم برای سرویس‌دهی به کاربران با حفظ کیفیت سرویس‌دهی، در نظر گرفتن مدل دینامیک برای شبکه و دخیل کردن پارامتر زمان و مدل‌سازی آن به فرم مسئله بهینه‌سازی دینامیکی با فرم تصادفی و حل آن با روش‌های موجود مانند روش لیپانوف، وارد کردن فرضیاتی همچون فرض چندهسته‌ای بودن دستگاه کاربران و تعداد سرورهای بیشتر و طراحی الگوریتمی برای اتصال کاربر به بهترین سرور موجود و دریافت سرویس از آن، تعدادی از کارهای تحقیقاتی است که می‌تواند در ادامه پژوهش به عنوان کارهای آتی، پیشنهاد شود.

مراجع

- [1] J. Zhang, W. Xia, F. Yan, L. Shen, "Joint computation offloading and resource allocation optimization in heterogeneous networks with mobile edge computing," IEEE Access, vol. 6, pp. 19324-19337, 2018.
- [2] X. Lyu, H. Tian, W. Ni, Y. Zhang, P. Zhang, and R. P. Liu, "Energy-efficient admission of delay-sensitive tasks for mobile edge computing," IEEE Trans. on Communications, vol. 66, no. 6, pp. 2603-2616, 2018.
- [3] F. Wang, J. Xu, X. Wang, S. Cui, "Joint offloading and computing optimization in wireless powered mobile-edge computing systems," IEEE Trans. on Wireless Communications, vol. 17, no. 3, pp. 1784-1797, 2018.
- [4] N. Nouri, and A. Tadaion. "Energy optimal resource allocation for mobile edge computation offloading in presence of computing access point." In Communication and Information Theory (IWCIT), 2018 Iran Workshop on, pp. 1-6, 2018.
- [5] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswmi, "Internet of things (IoT): a vision, architectural elements, and future directions," ELSEVIER Future Gener. Comp. Syst., vol. 29, no. 7, pp. 1645-1660, 2013.
- [6] N. Nouri, P. Rafiee and A. Tadaion. "NOMA-based energy-delay trade-off for mobile edge computation offloading in 5G networks," International Symposium on Telecommunications (IST), pp. 522-527, 2018.
- [7] S. Abolfazli, Z. Sanaei, E. Ahmed, A. Gani, and R. Buyya, "Cloudbased augmentation for mobile devices: motivation, taxonomies, and open challenges," IEEE Communications Surveys & Tutorials, vol. 16, no. 1, pp. 337-368, 2014.
- [8] سیمین قاسمی فلاورجانی، محمدعلی نعمتبخش، بهروز شاهقلی قهفرخی، «تخصیص وظایف چندهدفه در واگذاری به ابر سیار» مجله مهندسی برق دانشگاه تبریز، جلد ۴۶، شماره ۴، صفحات ۲۱۷-۲۳۲، زمستان ۱۳۹۵.
- [9] وحید ستاری نائینی، یاسمین سالم، عصمت راشدی، «بهره‌گیری از الگوریتم پرش ترکیبی قورباغه جهت کاهش مصرف انرژی مراکز داده ابری از طریق بهینه‌سازی مدیریت زمانبندی کارها و ترکیب مؤثر ماشینهای مجازی» مجله مهندسی برق دانشگاه تبریز، جلد ۴۸، شماره ۲، صفحات ۶۸۷-۶۹۸، تابستان ۱۳۹۷.
- [10] N. Fernando, S. Loke, and W. Rahayu, "Mobile cloud computing: A survey," Future Generation Computer Systems, vol. 29, no. 1, pp. 84-106, 2013.
- [11] Y. Liu, M. J. Lee, and Y. Zheng. "Adaptive multi-resource allocation for cloudlet-based mobile cloud computing system," IEEE Trans. on Mobile Computing, vol. 15, no. 10, pp. 2398-2410, 2016.
- [12] W. Song, and X. Su, "Review of mobile cloud computing," Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on. IEEE, pp 1-4, 2011.
- [13] M. Satyanarayanan, V. Bahl, R. Caceres and N. Davies, "The case for vm-based cloudlets in mobile computing," IEEE Pervasive Computing, no. 4, pp. 14-23, 2009.
- [14] M. Patel, B. Naughton, C. Chan, N. Sprecher, S. Abeta, and A. Neal, "Mobile-edge computing," Mobile-Edge Comput-Introductory, Tech.Rep, pp. 5896-5907, 2014.
- [15] J. Liu, Y. Mao, J. Zhang, and K. B. Letaief, "Delay-optimal computation task scheduling for mobile-edge computing systems," IEEE International Symposium on Information Theory (ISIT), pp. 1451-1455, 2016.
- [16] Y. Mao, J. Zhang, K. B. Letaief, "Dynamic computation offloading for mobile-edge computing with energy harvesting devices," IEEE Journal on Selected Areas in Communications, vol. 34, no. 12, pp. 3590-3605, 2016.
- [17] M. Kamoun, W. Labidi, and M. Sarkiss, "Joint resource allocation and offloading strategies in cloud enabled cellular networks," IEEE International Conference on Communications (ICC), pp. 5529-5534, 2015.
- [18] W. Labidi, M. Sarkiss, and M. Kamoun, "Energy-optimal resource scheduling and computation offloading in small cell networks,"

- [۲۹] نرگس پرهیزی، موسی مرزبند، سید مازیار میرحسینی مقدم، بهنام محمدی ایواتلو، فاطمه آذرینژادیان، «پایاده‌سازی عملی یک سیستم مدیریت انرژی برای یک ریزشبکه متصل به شبکه سراسری با استفاده از الگوریتم رقابت استعماری چندبعدی» مجله مهندسی برق دانشگاه تبریز، جلد ۴۶، شماره ۱، صفحات ۲۵-۴۰، بهار ۱۳۹۵.
- [30] H. Zhang, S. Huang, C. Jiang, K. Long, V. C. Leung and H. V. Poor, "Energy efficient user association and power allocation in millimeter-wave-based ultra dense networks with energy harvesting base stations," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1936-1947, 2017.
- [31] W. Hao and S. Yang, "Small cell cluster-based resource allocation for wireless backhaul in two-tier heterogeneous networks with massive MIMO," *IEEE Trans. on Vehicular Technology*, vol. 67, no. 1, pp. 509-523, 2018.
- [32] X. Chen, "Decentralized computation offloading game for mobile cloud computing," *IEEE Trans. on Parallel and Distributed Systems*, vol. 26, no. 4, pp. 974-983, 2015.
- [33] S. Sardellitti, G. Scutari, and S. Barbarossa. "Joint optimization of radio and computational resources for multicell mobile-edge computing," *IEEE Trans. on Signal and Information Processing over Networks*, vol. 1, no. 2, pp. 89-103, 2015.
- [34] M. Jia, J. Cao, W. Liang, "Optimal cloudlet placement and user to cloudlet allocation in wireless metropolitan area networks," *IEEE Trans. on Cloud Computing*, vol. 5, no. 4, pp. 725-737, 2017.
- [35] F. Fang, H. Zhang, J. Cheng, V. C. Leung, "Energy-efficient resource allocation for downlink non-orthogonal multiple access network," *IEEE Trans. on Communications*, vol. 64, no. 9, pp. 3722-3732, 2016.
- [36] Y. Wang, X. Lin, M. Pedram, "A nested two stage game-based optimization framework in mobile cloud computing system," *IEEE Seventh International Symposium on Service-Oriented System Engineering*, pp. 494-502, 2013.
- [37] J. Zhu, J. Wang, Y. Huang, S. He, X. You, L. Yang, "On optimal power allocation for downlink non-orthogonal multiple access systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 12, pp. 2744-2757, 2017.
- [38] G. Scutari, F. Facchinei, L. Lampariello, and P. Song, "Parallel and distributed methods for nonconvex optimization-Part I&II: Theory & Applications," *IEEE Trans. on Signal Processing*, vol. 65, no. 8, pp. 840-844, 2017.
- International Conference on Telecommunications (ICT), pp. 313-318, 2015.
- [19] K. Zhang, Y. Mao, S. Leng, Q. Zhao, L. Li, X. Peng, L. Pan, S. Maharjan, and Y. Zhang, "Energy-efficient offloading for mobile edge computing in 5G heterogeneous networks," *IEEE Access*, vol. 4, pp. 5896-5907, 2016.
- [20] X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient multi-user computation offloading for mobile-edge cloud computing," *IEEE/ACM Trans. on Networking*, vol. 24, no. 5, pp. 2795-2808, 2016.
- [21] M. H. Chen, B. Liang, and M. Dong, "A semi definite relaxation approach to mobile cloud offloading with computing access point," *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 186-190, 2015.
- [22] M. H. Chen, M. Dong, and B. Liang, "Joint offloading decision and resource allocation for mobile cloud with computing access point," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3516-3520, 2016.
- [23] S. Cao, X. Tao, Y. Hou, and Q. Cui, "An energy-optimal offloading algorithm for mobile computing based on hetnets," *International Conference on Connected Vehicles and Expo (ICCVE)*, 254-258, 2015.
- [24] Y. Zhao, S. Zhou, T. Zhao, and Z. Niu, "Energy-efficient task offloading for multiuser mobile cloud computing," *IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 1-5, 2015.
- [25] M. Deng, H. Tian, and B. Fan, "Fine-granularity based application offloading policy in small cell cloud-enhanced networks," *IEEE International Conference on Communications Workshops (ICC)*, pp. 638-643, 2016.
- [26] Y. Mao, J. Zhang, S.H. Song, and K. B. Letaief, "Power-delay tradeoff in multi-user mobile-edge computing systems," *IEEE Global Communications Conference (GLOBECOM)*, pp. 1-6, 2016.
- [27] C. C. Coskun and E. Ayanoglu, "Energy- and spectral-efficient resource allocation algorithm for heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 1, pp. 590-603, 2018.
- [۲۸] محمد لاری، «تخصیص منابع جهت کمینه‌سازی تأخیر ارسال در سامانه‌های مخابراتی تغذیه‌شونده به‌صورت بی‌سیم» مجله مهندسی برق دانشگاه تبریز، جلد ۴۷، شماره ۳، صفحات ۱۲۰۵-۱۲۱۲، پاییز ۱۳۹۶.

زیر نویس‌ها

¹⁴ Partial Offloading

¹⁵ Edge Server

¹⁶ Non-Orthogonal Multiple Access

¹⁷ Base Station

¹⁸ Local

¹⁹ Successive Convex Approximation

²⁰ Backhaul

²¹ Superposition Coding

²² Successive Interference Cancellation

²³ Offload Rate

²⁴ Central Processing Unit

²⁵ Dimension

²⁶ Iterative Algorithm

¹ Mobile Cloud Computing

² Cloud Computing

³ Cloud Server

⁴ Quality of Service

⁵ European Telecommunications Standards Institute

⁶ Full Offloading

⁷ Online

⁸ Offline

⁹ Deterministic

¹⁰ Stochastic

¹¹ Energy Efficient Computation Offloading

¹² Tradeoff

¹³ Heuristic