تخصیص منابع مبتنی بر یادگیری تقویتی برای بهبود گذردهی در ارتباطات D2D سلولی

وصال حکمی[۱]، استادیار؛ سید اکبر مصطفوی[۲]، استادیار؛ زیبا عارفی نژاد[۳]، کارشناسی ارشد

۱- دانشکده مهندسی کامپیوتر - دانشگاه علم و صنعت ایران - تهران - ایران – vhakami@iust.ac.ir
۲- گروه مهندسی کامپیوتر - دانشگاه یزد - یزد - ایران – a.mostafavi@yazd.ac.ir
۳- دانشکده مهندسی کامپیوتر – دانشگاه علم و صنعت ایران - تهران - ایران – z_arefinezhad@comp.iust.ac.ir

چکیده: با توجه به تقا ضای روزافزون برای پهنای باند شبکه‌های سلولی، همزیستی ارتباطات د ستگاه به د ستگاه (D2D[۱]) با م شترکان مجوزدار شبکه سلولی می‌تواند به بهره‌وری کارآمد از طیف مغناطیسی منجر شده و گذردهی شبکه را افزایش دهد. در این شیوه، منابع به نحوی میان مشترکان مجوزدار شبکه سلولی و زوج دستگاه‌ها با ارتباط مستقیم به اشتراک گذاشته می‌شود که ضمن افزایش بهره‌وری طیف فرکانسی، خللی در کیفیت سرویس کاربران مجوزدار ایجاد نشود. اغلب روش‌های تخصیص منابع کنونی متکی به اطلاعات وضعیت کانال و بازخورد نرخ ارسال مشترکان شبکه سلولی هستند که این اطلاعات به شکل دقیق در دسترس نیست. در این مقاله، یک روش نوآورانه مبتنی بر یادگیری تقویتی برای تنظیم حالت کاری کاربران D2D و اختصاص طیف به گره‌ها پیشنهاد می شود که بدون نیاز به اطلاعات و ضعیت کانال، منابع به شکل بهینه بین کاربران ارتباطات D2D و مشترکان شبکه سلولی تقسیم شده و گذردهی شبکه را بیشینه می‌سازد. ارزیابی‌های انجام شده نشان می‌دهد که روش پیشنهادی علیرغم عدم دسترسی به اطلاعات وضعیت کانال و بازخورد نرخ ارسال مشترکان، به گذردهی نزدیک به حالت بهینه دست می‌یابد و نرخ قطعی آن بسیار نزدیک به حالت ایده آل است.

واژه‌های کلیدی: ارتباطات D2D، اطلاعات وضعیت کانال، تخصیص طیف فرکانسی، یادگیری تقویتی.

RL-based Resource Allocation for Improving Throughput in Cellular D2D Communications

V. Hakami[1], Assistant Professor; S. A. Mostafavi[2], Assistant Professor; Z. Arefinezhad[3], MSc.

1- School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran, Email: vhakami@iust.ac.ir
2- Department of Computer Engineering, Yazd University, Yazd, Iran, Email: a.mostafavi@yazd.ac.ir
3- School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran, Email: z_arefinezhad@comp.iust.ac.ir

Abstract: With increasing demand of bandwidth-intensive application in cellular networks, coexistence of Device-to-Device (D2D) communications with cellular subscribers is a promising solution for high spectrum efficiency and network throughput. In cellular D2D communications, intelligent resource sharing among the network subscribers and paired devices is of significant importance. The most state-of-the-art works are relied on the exact values of Channel State Information (CSI) and subscribers' transmission rate feedback which are not available in the real cases. In this paper, we propose a novel reinforcement-learning-based approach for mode selection and spectrum allocation called RL-D2D which shares efficiently resources amongst the D2D users and cellular subscribers with the need for CSI, achieving high network throughput. The results of evaluations show that RL-D2D achieves near-optimal performance and low outage rate in despite of lack of CSI and users' transmission rate feedback.

Keywords: Device-to-Device communications, Channel-state information, Spectrum Allocation, Reinforcement Learning.

تاریخ ارسال مقاله: ۱۳۹۷/۰۷/۱۸
تاریخ اصلاح مقاله: ۱۳۹۷/۱۰/۳۰
تاریخ پذیرش مقاله: ۱۳۹۸/۰۲/۱۳
نام نویسنده مسئول: وصال حکمی
نشانی نویسنده مسئول: ایران - تهران – رسالت – خیابان دانشگاه - دانشگاه علم و صنعت ایران - دانشکده مهندسی کامپیوتر.

۱- مقدمه

در سال‌های اخیر حجم ترافیک در شبکه‌های سلولی رشد بسیاری داشته است و هر روزه نیاز به برقراری ارتباط با نرخ‌های بالا جهت انتقال حجم ترافیک بالا بیشتر می شود، تاجایی که پیش‌بینی می شود الگوهای فعلی این شبکه‌ها (نسل چهارم و بالاتر) پاسخگوی نیاز کاربران نخواهند بود [۳-۱]. ازاین‌رو اپراتورهای شبکه‌های سلولی باید راهکاری برای افزایش امکان ارسال و دریافت ترافیک انجام دهند. اولین راه‌حل، افزایش پهنای باند مشترکین است، اما به دلیل محدودیت و ارزش طیف فرکانسی نمی‌توان با اختصاص پهنای باند بیشتر به شبکه‌های سلولی مشکل ترافیک زیاد را حل کرد. رویکردهای اخیر برای فائق آمدن بر این مشکل، بالا بردن بهره‌وری طیف فرکانسی در شبکه‌های سیار است [۴].

یک راه‌حل مؤثر در این راستا، کوچک‌سازی سلول‌های اپراتورهای سیار است، طوری که پهنای باند باریک‌تر و کمتری مصرف شود. همچنین، نزدیک‌تر شدن مشترکین به ایستگاه اصلی سلول، باعث بالا رفتن نرخ ارسال داده، کمتر شدن تأخیر، نویز و مصرف انرژی می شود. از راهکارهای اساسی برای افزایش گذردهی در شبکه‌های سیار، راه‌اندازی سلول‌های کوچک به نام فمتو سل [۵] است. در فمتو سل‌ها یک ایستگاه اصلی در یک محدوده کوچک نصب می‌شود و شعاع آنتن‌دهی آن داخل همان محدوده است. با این روش کاربران محدوده‌های مجاور می‌توانند از طیف فرکانسی یکسانی استفاده کنند که این به معنای افزایش کارایی شبکه در واحد مساحت است. همچنین در این تکنیک چون فرستنده و گیرنده نسبت به شبکه‌های سلولی بزرگ به یکدیگر نزدیک‌تر هستند، توان مصرفی کمتری برای ارسال نیاز دارند. اما راهکارهایی همچون کوچک کردن سلول، استفاده از چند ایستگاه اصلی در یک سلول و یا استفاده از فمتوسل به نوبه خود باعث افزایش هزینه خرید و نگهداری تجهیزات، پیچیدگی مدیریت تداخل و نهایتاً هدررفت انرژی و سرمایه می‌شود [۶،۵].

ایده همزیستی ارتباطات D2D در کنار ارتباط شبکه سلولی، در زمره جدیدترین رویکردهای افزایش بهره‌وری طیف فرکانسی است. نزدیک بودن کاربرها به یکدیگر، امکان برقراری ارتباطات D2D را فراهم نموده است که در آن ارتباط بین کاربرها از هسته شبکه سلولی و ایستگاه پایه به‌عنوان گره میانی عبور نمی‌کند. این ارتباطات باید به‌گونه‌ای برقرار شود که مشترکین اصلی شبکه سلولی با افت کیفیت سرویس مواجه نشوند و کوچک‌ترین خللی به ارتباطات مشترکین سلولی وارد نشود [۹-۷].

در این مقاله روشی برای اشتراک منابع میان مشترکین سلولی و زوج D2D با پیکربندی مناسب ارائه می شود تا اهداف سیستمی مانند گذردهی در کل شبکه بیشینه شود. نوآوری انجام شده در این تحقیق، عدم اتکا به پارامتر اطلاعات وضعیت کانال (CSI) و استفاده از بازخورد نرخ ارسال مشترکین شبکه است.

در مسئله پیش رو، متغیرهای تصمیم‌گیری در شبکه شامل تنظیم مُد کاری کاربران D2D و نیز اختصاص منابع شامل کانال و توان به گره‌ها است. با تنظیم بهینه این متغیرها می‌توان به حالتی دست یافت که هر دو گروه کاربری (ارتباطات سلولی و ارتباطات D2D) از مشارکت به وجود آمده بیشترین بهره را ببرند. روش‌های موجود برای محاسبه پیکربندی بهینه منابع و نیز تنظیم مُد کاری ارتباطات D2D متکی به جمع‌آوری اطلاعات کامل از کیفیت لینک‌های ارتباطی در سرتاسر شبکه هستند. اطلاعات و ضعیت کانال (CSI[۲]) در شبکه‌های بی سیم ماهیتی تصادفی و متغیر با زمان دارد و فرض دسترسی به مقدار دقیق این اطلاعات یک فرض محدودکننده و غیرواقع بینانه است [۱۲-۱۰]. اگر مقدار دقیق پارامتر CSI مشترکین در اختیار ایستگاه اصلی قرار داشت، اشتراک منابع سلولی و انتخاب مُدکاری مشترکین D2D قبل از شروع کار و به‌صورت آفلاین انجام می‌گرفت، اما در غیاب اطلاعات CSI، تخصیص منابع باید به‌صورت آنلاین انجام شود.

هدف ما در این مقاله، ارائه راهکاری است که علاوه بر پرداختن توأمان به مسئله مُد کاری و تخصیص منابع بتواند در غیاب CSI، پیکربندی بهینه شبکه را تعیین نماید. روش پیشنهادی ما (RL-D2D[۳]) مبتنی بر استفاده از بازخورد اطلاعات وضعیت گذردهی شبکه است. در روش پیشنهادی RL-D2D، میزان دستیابی به نرخ ارسال داده مشترکان و کاربران D2D محاسبه شده و در اختیار ایستگاه مرکزی قرار داده می‌شود تا بر اساس آن در مورد تخصیص منابع برای ارتباطات سلولی و D2D تصمیم‌گیری کند. اغلب روش‌های ارائه شده برای انتخاب پارامترهای سه‌گانه انتخاب مد کاری، تخصیص کانال و تعیین سطح توان سیگنال راهکار یکپارچه ارائه نمی‌دهند که این امر می‌تواند باعث عدم انعطاف‌پذیری شبکه در شرایط متفاوت شود. بررسی‌های انجام شده در این کار نشان می‌دهد که هر مُدکاری ممکن است در یک لحظه برای شبکه مناسب باشد. در نتیجه، ترکیب و انتخاب لحظه‌ای هرکدام از آن‌ها می‌تواند تأثیر بیشتری برای بالا بردن گذردهی شبکه داشته باشد.

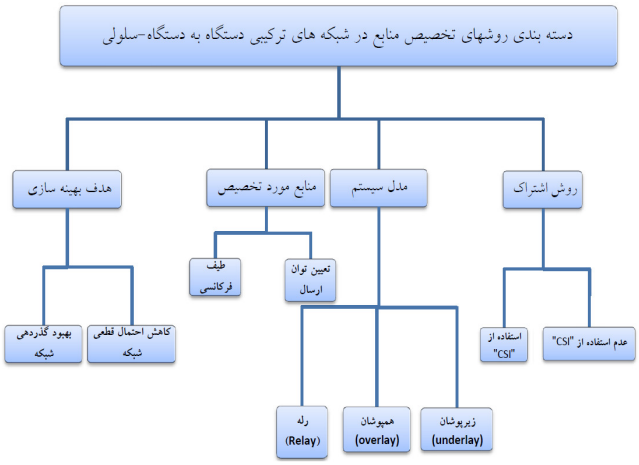
در این مقاله مسئله تخصیص منابع بر اساس پارامترهای مطرح شده به‌صورت یک مسئله راهزن چند دست ترکیباتی متمرکز (CMAB[۴]) فرمول‌بندی شده است و از روش یادگیری با پاداش خطی (LLR[۵]) برای حل آن استفاده شده است. پیاده سازی و شبیه سازی روش پیشنهادی با روش‌های مبتنی بر CSI نشان‌دهنده عملکرد مناسب الگوریتم پیشنهادی است.

در ادامه ساختار مقاله به‌صورت زیر است: در بخش ۲، کارهای مرتبط در زمینه تخصیص منابع برای ارتباطات D2D سلولی مطرح می شود. در بخش ۳، مسئله پیش رو فرمول‌بندی شده و راه‌حل پیشنهادی به‌تفصیل مطرح می‌شود. روش پیشنهادی در بخش ۴ مورد ارزیابی قرار می‌گیرد و نهایتاً نتایج حاصل از مقاله در بخش ۵ مطرح می‌شود.

۲- کارهای مرتبط

تاکنون روش‌ها و مدل‌های متفاوتی برای پیاده سازی ارتباطات D2D در شبکه‌های سلولی ارائه شده است. در هر یک از آن‌ها هدف خاصی دنبال شده و با توجه به اهداف ذکر شده، مدل سیستمی و روش تخصیص منابع در شبکه سلولی متفاوت بوده است. در برخی موارد برای ساده‌تر شدن پیکربندی، تنها از یک مُدکاری استفاده شده و برای اشتراک میان مشترکین سلولی و زوج ارتباط D2D هیچ سیاستی مدنظر قرار نگرفته است. در این دست از پیکربندی‌ها معیار فقط سادگی کار و کاهش سربار کنترلی شبکه بوده و گذردهی شبکه افزایش کمتری نسبت به راهکارهای دیگر داشته است.

در شکل ۱ دسته‌بندی انواع پیاده‌سازی‌های شبکه‌های ترکیبی D2D-سلولی به تصویر کشیده شده است. در بیشتر راهکارهای ارائه شده در شبکه‌های ترکیبی، یک یا چندین شاخص که در شکل نمایش داده شده است مدنظر قرار گرفته است. تعدد شاخص‌های در نظر گرفته شده باعث افزایش بهره‌وری و در عین حال پیچیدگی پیاده‌سازی شبکه ترکیبی می‌شود. به‌کارگیری توأمان تمامی این شاخص‌ها می تواند باعث بهبود حداکثری در گذردهی و کاهش احتمال قطعی در شبکه سلولی شود.



شکل ۱: دسته‌بندی روش‌های تخصیص منابع در شبکه‌های ترکیبی

در جدول ۱ به دسته‌بندی رویکردهای تخصیص منابع در ارتباطات D2D پرداخته شده و تأثیرات هر مدکاری و عوامل لازم جهت انتخاب آن بررسی شده است. در مد همپوشان و رله تقسیم زمانی میان مشترک سلولی و زوج مشترک D2D انجام می‌شود، و دیگر تداخل درون سلولی رخ نمی‌دهد. بهبود آنتن دهی در سلول و کاهش احتمال قطعی شبکه فقط در مدکاری رله اتفاق می‌افتد. امکان استفاده از الگوریتم مجارستانی و اشتراک تصادفی منابع فقط در مدکاری زیرپو شان که تق سیم زمانی صورت نمی‌گیرد امکان‌پذیر ا ست و باعث سادگی در پیاده‌سازی شبکه می‌شود.

کارهای انجام شده در این حوزه را می‌توان به دو دسته تقسیم کرد. در دسته اول راهکارها، مد کاری لحاظ نشده است که این امر باعث سادگی پیاده‌سازی شبکه ترکیبی می‌شود، ولی این شبکه‌ها نسبت به تغییر شرایط شبکه انعطاف‌پذیری کمتری دارند. در دسته دوم، علاوه بر تخصیص منابع، انتخاب مد کاری نیز در نظر گرفته شده است، بدین ترتیب ضمن پیچیدگی پیکربندی این شبکه‌ها، انعطاف‌پذیری بیشتری در آن‌ها حاصل خواهد شد.

جدول ۱: نحوه تخصیص منابع در شبکه‌های ترکیبی

مد کاری	رله[۶]	همپوشان[۷]	زیرپوشان[۸]
تقسیم زمانی منابع سلولی	✓	✓	×
بروز تداخل درون سلولی	×	×	✓
نیاز به تعیین قدرت سیگنال ارسالی	×	×	✓
بهبود گذردهی شبکه	✓	✓	✓
کاهش احتمال قطعی شبکه سلولی	✓	×	×
افزایش محدوده آنتن دهی سلول	✓	×	×
امکان اشتراک تصادفی منابع، میان دو گروه مشترکین	×	×	✓
امکان استفاده از الگوریتم مجارستانی، برای اشتراک منابع	✓	✓	✓

۲-۱- راهکارهای مبتنی بر مد کاری ثابت

هن و همکاران [۱۳] تخصیص بهینه منابع سلولی برای مشترکین D2D را بررسی کرده‌اند. طرح آن‌ها فقط مدکاری زیرپوشان را پوشش می‌دهد و مشترکان D2D در این مُد کاری از منابع سلولی استفاده می کنند. آن‌ها نرخ ارسال مشترکین سلولی و D2D را در هر دو سمت ارسال و دریافت مورد بررسی قرار داده‌اند. داپلر و همکاران [۷]، ارتباطات D2D در مُد زیرپوشان را بررسی کرده‌اند. این روش با در نظر گرفتن مکان فیزیکی دو مشترک متقاضی ارتباط D2D، فاصله دو مشترک را محاسبه می‌کند. در صورت نزدیک بودن آن‌ها به یکدیگر، برای آن‌ها ارتباط D2D انتخاب شده و منابع سلولی به‌صورت اشتراکی در اختیار آن‌ها قرار داده می‌شود، در غیر این‌صورت ارتباط این دو مشترک از طریق شبکه سلولی برقرار شده و اشتراک طیف فرکانسی انجام نمی‌شود. سی. یو و همکاران [۱۴]، تحقیقاتی در زمینه بهبود بهره‌وری در ارتباطات D2D در مُد زیرپوشان و همپوشان انجام داده‌اند. در این مقاله، تلاش محققان بر این بوده است تا با کنترل توان بتوانند از بروز تداخل در مُد زیرپوشان جلوگیری کنند. در مرجع [۱۴]، عدم دسترسی به CSI توسط ایستگاه اصلی مد نظر قرار نگرفته است و بر آن تأکید شده است که هر مشترک سلولی بتواند در هر زمان ممکن، ارتباط D2D برقرار سازد.

مین و همکاران [۱۵] یک راهکار کنترل محدوده تداخل برای بالا بردن تعداد ارتباطات D2D در مُد کاری زیرپوشان مطرح کرده‌اند. در این راهکار، ایستگاه اصلی، فضای سلول تحت پوشش خود را به چندین محدوده فرضی تقسیم می‌کند. سپس، طیف‌های فرکانسی را به مشترکین اصلی هر محدوده اختصاص می‌دهد و مشابه آن فرکانس‌ها را در محدوده‌های مقابل و با فاصله زیاد، به کاربران ارتباطات D2D

واگذار می‌کند. با این وجود، اساس این روش بر در اختیار داشتن CSI مشترکین توسط ایستگاه اصلی است.

وای. یی و همکاران [۱۶]، یک راهکار بر اساس مرتب کردن زیرساخت شبکه مطرح کرده‌اند که در مُد رله کار می‌کند. در این طرح، مشترکین اصلی توسط ایستگاه اصلی اولیه مدیریت می‌شوند و مشترکین ثانویه توسط ایستگاه اصلی ثانویه (ایستگاه کمکی) اداره می شوند. زیانگ لی یو و همکاران [۱۷]، مدلی شبیه‌سازی کرده‌اند که بر پایه IMT-Advanced بنا نهاده شده است و شامل شبکه‌های سلولی نسل ۴ و ۴،۵ می‌باشد. در این طرح دو مُد کاری زیرپوشان و همپوشان تحت یک سناریو بررسی شده است. برای تمامی عملیات در نظر گرفته شده در این مدل نیاز به داشتن پارامتر CSI است.

۲-۲- راهکارهای مبتنی بر انتخاب مد کاری

وای لیانگ و همکاران [۱۸] ارتباطات را به دو بخش ارتباطات اصلی و ارتباطات D2D تقسیم می‌کند. ابتدا برای هر ارتباط سلولی بین کاربر و ایستگاه اصلی یک زوجِ D2D انتخاب می شود که این انتخاب به‌وسیله پارامتر CSI و دانستن مکان جغرافیایی زوج دستگاه صورت می‌گیرد. ایستگاه اصلی می‌داند اطلاعات سلولی را می‌تواند با قدرت کمتری برای کاربران D2D ارسال کند و آن‌ها نیز با باز پخش کردن اطلاعات با قدرت پایین برای مشترک سلولی، اطلاعات را به آن مشترک برسانند.

چیترا و همکاران [۱۹] در مقاله خود ارتباط چندگانه D2D را بررسی کرده‌اند. در این طرح اگر دو مشترک در فاصله کمی از یکدیگر واقع شده باشند و امکان ارتباط D2D برایشان مقدور باشد، ایستگاه اصلی پیوند ارتباطی این دو را بر اساس ارتباط مستقیم D2D و در مد کاری زیرپوشان با استفاده از منابع سلولی برقرار می‌کند. در صورتی‌که ارتباط مستقیم D2D مقدور نباشد، ارتباط دو مشترک از طریق ایستگاه اصلی و مانند مشترکین عادی سلولی برقرار می شود. یانگ کااو و همکاران [۱۳] در تحقیقاتشان برای بهبود عملکرد شبکه‌های سلولی نسل چهارم به بالا از ارتباط مستقیم D2D استفاده کرده‌اند. در این راهکار، هر سه مُد کاری زیرپوشان، همپوشان و بازپخشی (رله) برای ارتباطات D2D انتخاب می‌شود. چنانچه منابع فرکانسی تخصیص یافته بیش از نیاز مشترک اصلی بود، ایستگاه مرکزی آن را در اختیار کاربر ارتباطی D2D قرار می‌دهد، که این ارتباط می‌تواند در مُد زیرپوشان و یا همپوشان باشد. این طرح کامل‌تر از دیگر مدل‌های ذکر شده است و با انتخاب مُدکاری می‌تواند راهکار مناسب‌تری را برای بالابردن توان عملیاتی شبکه بکار گیرد. ایراد عمده روش مرجع [۱۳]، اتکای آن بر دسترسی به CSI به‌صورت دقیق و کامل می‌باشد. در مرجع [۲۰] یک راهکار مبتنی بر الگوریتم ژنتیک برای تخصیص بلوک‌های منبع در شبکه‌های مبتنی بر OFDMA ارائه شده است. روش پیشنهادی این مقاله برای مسئله تخصیص هم‌زمان منابع در محیط ناشناخته به کاربران متعدد قابل به‌کارگیری نیست. مرجع [۲۱] مسئله انتخاب مجدد سلول و تخصیص کاربران به هر سلول در شبکه‌های سلولی مبتنی بر 3GPP مورد برر سی قرار گرفته است.

در شرایط واقعی، ایستگاه اصلی برای جمع‌آوری این اطلاعات از کل شبکه سربار کنترلی زیادی را ایجاد می‌کند. همچنین، شرایط کانال تصادفی بوده و با زمان متغیر است. بنابراین، محاسبه پیکربندی بهینه بر اساس اطلاعات لحظه‌ای کفایت نمی‌کند و باید با تغییر شرایط، پیکربندی شبکه و نحوه تخصیص منابع نیز تطبیق داده شوند. جدول ۲ خلاصه‌ای از تحقیقات انجام شده در این زمینه را به همراه ویژگی‌های هر روش نشان می‌دهد.

با توجه به مطالب ذکر شده، هر یک از مدهای کاری دارای محاسن و معایبی هستند. پیاده سازی ایده همزیستی در هر یک از این سه مُد به‌طور ایستا منجر به پیکربندی با بالاترین کارایی نمی شود. بلکه، باید بسته به شرایط سلول‌ها، کانال‌های ارتباطی و تراکم کاربران فعال، مُد بهینه با هدف حداکثرسازی گذردهی شبکه تعیین شود.

جدول ۲: مقایسه ویژگی‌های کارهای مرتبط در زمینه شبکه‌های ترکیبی

مرجع	عدم نیاز به CSI	مُد کاری پشتیبانی شده	منبع مورد تخصیص	کاهش احتمال قطعی
[۱۶]	خیر	زیرپوشان	طیف فرکانسی	خیر
[۷]	خیر	زیرپوشان	طیف فرکانسی و توان	خیر
[۱۴]	خیر	زیرپوشان، همپوشان	طیف فرکانسی، توان و تایم اسلات زمانی	خیر
[۱۵]	خیر	زیرپوشان	طیف فرکانسی و توان	خیر
[۲۲]	خیر	زیرپوشان، همپوشان، رله	طیف فرکانسی و توان	خیر
[۲۴]	خیر	زیرپوشان	طیف فرکانسی و توان	خیر
[۲۵]	خیر	زیرپوشان	طیف فرکانسی، توان و تایم اسلات زمانی	خیر
[۲۶]	خیر	رله	طیف فرکانسی و تایم اسلات زمانی	خیر
[۱۷]	خیر	رله	طیف فرکانسی، توان و تایم اسلات زمانی	بله
[۱۸]	خیر	زیرپوشان، همپوشان، رله	طیف فرکانسی، توان و تایم اسلات زمانی	بله
[۱۹]	خیر	زیرپوشان و رله	طیف فرکانسی، توان و تایم اسلات زمانی	بله
[۱۳]	خیر	زیرپوشان، همپوشان، رله	طیف فرکانسی، توان و تایم اسلات زمانی	بله
روش پیشنهادی	بله	زیرپوشان، همپوشان، رله	طیف فرکانسی، توان و تایم اسلات زمانی	بله

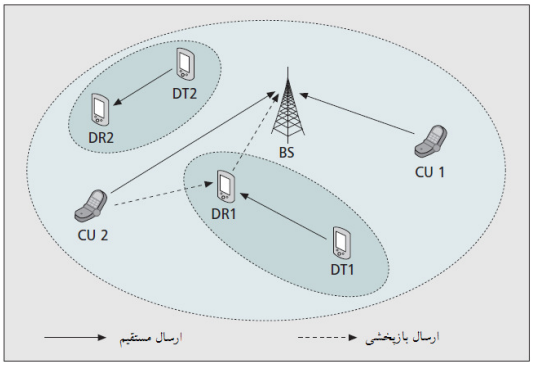
۳- روش پیشنهادی: RL-D2D

در این بخش، یک روش نوآورانه پیشنهادی مبتنی بر یادگیری تقویتی به نام RL-D2D برای افزایش گذردهی شبکه‌های ترکیبی سلولی شرح داده می شود. در ابتدا، مدل سیستم ارائه شده و نحوه اشتراک‌گذاری

منابع و انتخاب مد کاری تعریف می‌شود. سپس، مسئله انتخاب مد کاری به‌صورت یک مسئله یادگیری تقویتی فرمول‌بندی شده و الگوریتم کارآمدی برای حل آن ارائه می‌شود.

۳-۱- مدل سیستم و فرضیات

دریک شبکه تک سلولی، تعداد m مشترک سلولی CU[۹] و n زوج مشترک ارتباط D2D به صورت فرستنده-گیرنده (DT-DR[۱۰]) هستند. تعداد مشترکین سلولی بیشتر از زوج ارتباط D2D است، پس n زوج مشترک D2D نیاز به منابع برای ارتباط دارند و باید با n مشترک سلولی اشتراک منابع داشته باشند (شکل ۲). برای به اشتراک‌گذاری منابع، باید انتخاب مُدکاری و سطح قدرت سیگنال ارسالی نیز در نظر گرفته شود. مُدکاری مشترکین می‌تواند مُدزیرپوشان، مُدهمپوشان و یا مُدکاری رله باشد. وظیفه مدیریت منابع، انتخاب اشتراک منابع سلولی با ارتباط D2D و انتخاب مُدکاری بهینه بر عهده ایستگاه پایه BS می باشد.



شکل ۲: مدل شبکه فرض شده در روش پیشنهادی

۳-۱-۱- محاسبه نرخ ارسال مشترکین در مُدکاری زیرپوشان

در این مُدکاری چون مشترک سلولی و مشترک D2D، در یک طیف فرکانسی و هم‌زمان سیگنال خود را ارسال می‌کنند، تداخل درون سلولی به وجود می‌آید، لذا در گیرنده باید تداخل سیگنال این دو مشترک با هم مدنظر قرار داده شود. برای محاسبه نرخ ارسالِ مشترکین سلولی خواهیم داشت [۱۳]:

$$\gamma_{C_i} = \frac{P_{C_i} \times g_{C_i}}{P_{D_j} \times I_{D_j} + Noise},\ R_{UC_i} = B \times log_2(1 + \gamma_{C_i}) \qquad (۱)$$

در رابطه (۱)، R_{UC_i} نرخ ارسال مشترک سلولی i برای ایستگاه اصلی در مُدکاری زیرپوشان است، که از رابطه شانون به دست می‌آید، B پهنای باند اختصاص داده شده به مشترک برحسب هرتز، γ_{C_i} نسبت سیگنال به نویز در بسته ارسالی از مشترک سلولی i برای ایستگاه اصلی، P_{C_i} توان سیگنال ارسالی مشترک سلولی i به ایستگاه اصلی، g_{C_i} میزان بهره کانال مشترک سلولی i، P_{D_j} توان سیگنال ارسالی مشترک j $D2D$ که از همان فرکانس مشترک سلولی استفاده می‌کند و بر روی سیگنال مشترک j $D2D$ تداخل ایجاد می‌کند، I_{D_j} میزان تداخل مشترک j $D2D$ بر روی سیگنال مشترک سلولی در گیرنده، $Noise$ چگالی طیفی توان نویز گوسی سفید است.

همچنین برای مشترک $D2D$، نرخ ارسال به‌صورت زیر محاسبه می شود:

$$\gamma_{D_j} = \frac{P_{D_j} \times g_{D_j}}{P_{C_i} \times I_{C_i} + Noise},\ R_{UD_j} = B \times log_2(1 + \gamma_{D_j}) \qquad (۲)$$

در رابطه (۲)، R_{UD_j} نرخ ارسال مشترک D2D j فرستنده برای مشترک گیرنده در مُدکاری زیرپوشان، γ_{D_j} نسبت سیگنال به نویز در بسته ارسالی از مشترک D2D فرستندهِ j برای گیرنده، P_{D_j} توان سیگنال ارسالی مشترک D2D j، g_{D_j} میزان بهره کانال مشترک ارسال کننده D2D j، P_{C_i} توان سیگنال ارسالی مشترک سلولی i که از همان فرکانس مشترک D2D استفاده می‌کند و بر روی سیگنال مشترک D2D تداخل ایجاد می‌کند، I_{C_i} میزان تداخل مشترک سلولی i بر روی سیگنال مشترک D2D در گیرنده می‌باشد.

۳-۱-۲- محاسبه نرخ ارسال مشترکین در مُدکاری همپوشان

در این مُد کاری، طیف فرکانسی میان مشترکین سلولی و مشترکین D2D تقسیم زمانی می‌شود، بدین ترتیب دیگر تداخل درون سلولی وجود نخواهد داشت و میزان تداخل خارج سلولی را صفر درنظر می گیریم، پس محاسبه میزان گذردهی برای دو مشترک سلولی و D2D به شکل زیر خواهد بود [۱۳] :

$$\gamma_{C_i} = \frac{P_{C_i} \times g_{C_i}}{Noise},\ R_{OC_i} = \theta \times B \times \log_2(1 + \gamma_{C_i}) \qquad (۳)$$

$$\gamma_{D_j} = \frac{P_{D_j} \times g_{D_j}}{Noise},\ R_{OD_j} = (1 - \theta) \times B \times \log_2(1 + \gamma_{D_j}) \qquad (۴)$$

$$\theta = \frac{R_{\text{Threshold}}}{R_{Maximum}} \qquad (۵)$$

در رابطه (۳) R_{OC_i} نرخ ارسال مشترکین سلولی i در مُدکاری همپوشان و در رابطه (۴) R_{OD_j} نرخ ارسال مشترکین D2D j در مُدکاری همپوشان می‌باشند. در رابطه (۵)، $R_{Maximum}$ حداکثر میزان نرخ ارسال قابل‌دستیابی برای مشترک سلولی و $R_{Threshold}$ حداقل نرخ ارسال مورد نیاز مشترک سلولی است، چنانچه $R_{Max} < R_{Thr}$ باشد، امکان اشتراک منابع سلولی با ارتباط D2D، در این مُدکاری نخواهد بود [۱۳].

همان‌طورکه از روابط (۳) و (۴) مشهود است، بدون تداخل درون سلولی هر چه قدرت سیگنال بالاتر باشد، میزان گذردهی بیشتر خواهد بود، لذا در این مُدکاری نیازی به تعیین قدرت سیگنال نیست و میزان آن حداکثر در نظر گرفته می‌شود.

۳-۱-۳ محاسبه نرخ ارسال مشترکین شبکه در مُدکاری رله

در این مُدکاری، طیف فرکانسی به سه زمان تقسیم می‌شود: ۱- ارسال اطلاعات مشترک سلولی برای ایستگاه اصلی و مشترک D2D جهت رله، ۲- بازپخش اطلاعات مشترک سلولی توسط مشترک D2D رله برای ایستگاه اصلی، ۳- ارسال اطلاعات مشترکین D2D برای یکدیگر. به‌اندازه α از زمان طیف برای زمان اول، همان مقدار برای زمان دوم و زمان $(1\text{-}2\alpha)$ باقیمانده برای زمان سوم در نظر گرفته می‌شود. بدیهی است مقدار α نمی‌تواند بیشتر از ۰/۵ باشد، حتی درصورتی‌که این مقدار ۰/۵ باشد، زمانی برای ارسال مشترکین $D2D$ باقی نمی‌ماند. معمولاً مقدار این پارامتر ۰/۴ در نظر گرفته می‌شود[۱۳].

بدین ترتیب برای محاسبه نرخ ارسال خواهیم داشت:

$$\gamma_{C_{c2B}} = \frac{P_{C_i} \times g_{C_i}}{Noise},\ R_1 = B \times \log_2(1 + \gamma_{C_{c2B}}) \quad (۶)$$

$$\gamma_{C_{c2d}} = \frac{P_{C_i} \times g_{C_{i2}}}{Noise},\ R_2 = B \times log_2(1 + \gamma_{C_{c2d}}) \quad (۷)$$

$$\gamma_{D_{d2B}} = \frac{P_{d_i} \times g_{d_i}}{Noise},\ R_{d2B} = B \times log_2(1 + \gamma_{D_{d2B}}) \quad (۸)$$

$$\gamma_{D_j} = \frac{P_{D_j} \times g_{D_j}}{Noise},\ R_{RD_j} = (1 - 2\alpha) \times B \times log_2\left(1 + \gamma_{D_j}\right) \quad (۹)$$

$$R_{RC_i} = \alpha \times max\{R_1, min(R_2, R_{d2B})\} \quad (۱۰)$$

در رابطه (۶)، $\gamma_{C_{c2B}}$ نسبت سیگنال به نویز اطلاعات ارسالی از مشترک سلولی برای ایستگاه اصلی و g_{C_i} بهره کانال مشترک سلولی i در ارسال به ایستگاه اصلی است. در رابطه (۷)، $\gamma_{C_{c2d}}$ نسبت سیگنال به نویز اطلاعات ارسالی از مشترک سلولی برای مشترک رله کننده و $g_{C_{i2}}$ بهره کانال مشترک سلولی i در ارسال به مشترک D2Dj جهت رله می‌باشد. در رابطه (۸)، $\gamma_{D_{d2B}}$ نسبت سیگنال به نویز اطلاعات رله شده از مشترک D2D برای ایستگاه اصلی و g_{d_i} بهره کانال ارسال مشترک D2Dj برای ایستگاه اصلی در هنگام رله اطلاعات مشترک سلولی i می‌باشد. در رابطه (۹)، γ_{D_j} نسبت سیگنال به نویز اطلاعات ارسالی دو مشترک D2D و R_{RD_j} نرخ ارسال مشترک D2D j در مُدکاری رله می‌باشند. در رابطه (۱۰)، R_{RC_i} نرخ ارسال مشترک سلولی i در مُدکاری رله است. همان‌طورکه در مورد مُدکاری همپوشان توضیح داده شد، در این مُدکاری نیز به دلیل تقسیم فرکانسی، تداخل درون سلولی نداشته و مقدار قدرت سیگنال باید حداکثر باشد.

۳-۱-۴- محاسبه احتمال قطعی کاربران سلولی در شبکه

یکی از پارامترهای مهم در تعیین کیفیت سرویس در شبکه‌های سلولی، احتمال قطعی در این شبکه است. هر چه این شاخص پایین‌تر باشد، کیفیت سرویس شبکه بالاتر بوده و مشترکین سلولی رضایتمندی بالاتری خواهند داشت. یکی از مزایای استفاده از ارتباطات D2D این است که با رله اطلاعات مشترکین سلولی احتمال قطعی کاهش یافته و در نتیجه، کیفیت سرویس افزایش می‌یابد [۲۳]. برای محاسبه این پارامتر، یک نرخ ارسال آستانه برای مشترکین سلولی در نظر گرفته می شود. چنانچه نرخ ارسال اطلاعاتِ مشترک سلولی، برای ایستگاه اصلی پایین‌تر از آن باشد، قطعی اتفاق می‌افتد.

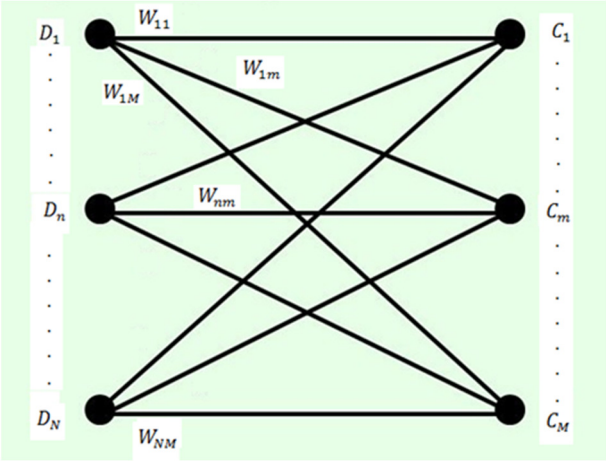
در ارتباط مشارکتی سلولی با D2D، چنانچه از مُدکاری رله استفاده شود، مشترک ارتباط D2D، داده ارسالی مشترک سلولی برای ایستگاه اصلی را باز پخش می‌کند. چون اطلاعات مشترک سلولی از دو مسیر جدا به ایستگاه اصلی می‌رسد، لذا احتمال اینکه نرخ ارسال بیشتر از $R_{Threshold}$ بشود، بالاتر رفته و به‌تبع آن، احتمال قطعی کاهش می‌یابد. برای محاسبه احتمال قطعی از رابطه زیر استفاده می‌شود [۲۳]:

$$outage - ratio = \frac{\left|\{i: R_{Cu_i} < R_{Threshold}\}\right|}{M}, i \in \{1, .., m\} \quad (۱۱)$$

در رابطه (۱۱)، صورت کسر بیانگر تعداد مشترکین سلولی است که نرخ ارسال آن‌ها کمتر از نرخ ارسال آستانه است و مخرج کسر M نشان‌دهنده تعداد کل مشترکین سلولی می‌باشد.

۳-۲- تعریف مسئله تخصیص منابع و انتخاب مد کاری

تخصیص منابع سلولی به ارتباط D2D می‌تواند با مسئله تطابق دوبخشی وزن‌دار حل شود. در یک سوی مشترکین شبکه سلولی و در سوی دیگر اعضای ارتباط D2D قرار دارند، ارتباط دو طرف با یک یال نشان داده می‌شود و نرخ ارسال به دست آمده از اشتراک منابع هر مشترک سلولی با هر زوج ارتباط D2D، وزن آن یال است (شکل ۳). وزن یال‌های گراف دوبخشی، نرخ شانون در پیکربندی‌های مختلف (زیرپوشان، همپوشان و رله) است.



شکل۳: مدل‌سازی مسئله تخصیص منابع و انتخاب مُد به‌صورت یک مسئله تطابق دوبخشی وزن‌دار

محاسبه مجموع نرخ ارسال برای هر اشتراک منبع بین مشترکین سلولی و D2D از روابط زیر محاسبه می‌شود [۱۳]:

$$W_{ijOver} = R_{OC_i} + R_{OD_j} \quad (۱۲)$$

$$W_{ijUnder} = R_{UC_i} + R_{UD_j} \quad (۱۳)$$

$$W_{ijRelay} = R_{RC_i} + R_{RD_j} \quad (۱۴)$$

در رابطه (۱۲)، W_{ijOver} مجموع نرخ ارسال مشترک سلولی i با زوج مشترک j در مُدکاری همپوشان، در رابطه (۱۳) $W_{ijUnder}$ مجموع نرخ ارسال مشترک سلولی i با زوج مشترک j در مُدکاری زیرپوشان و در رابطه (۱۴)، $W_{ijRelay}$ مجموع نرخ ارسال مشترک سلولی i با زوج مشترک j در مُدکاری رله می‌باشند.

$$W_{ij} = max(W_{ijOver}, W_{ijUnder}, W_{ijRelay}), i, j \in \{1,2,..,n\} \quad (۱۵)$$

در رابطه (۱۵)، برای هر اشتراک منابع میان مشترک سلولی i و زوج مشترک D2D j (وزن هر یال) هرکدام از سه مُدکاری که نرخ گذردهی بالاتری داشته باشد به‌عنوان شاخص برای ارتباط i و j در نظر گرفته می‌شود و تعیین کننده مقدار W_{ij} خواهد بود.

اگر در زمان محاسبه پارامتر CSI موجود باشد و شرایط شبکه تغییر نکند، با استفاده از روابط ذکر شده برای هر مُدکاری و هر یک از توان‌های ارسال سیگنالِ مشترکین، میزان نرخ ارسال حداکثری برای

اشتراک دو نوع مشترک محاسبه می‌شود. با استفاده از الگوریتم مجارستانی، بیشترین وزن‌های یال برای اشتراک منابع سلولی با مشترکین D2D انتخاب شده و شبکه به حداکثر میزان گذردهی دست خواهد یافت. به‌عبارت‌دیگر، مسئله (۱۶) باید برای تعیین انتساب بهینه زوج مشترک *D2D* به کانال‌های سلولی حل شود.

$$\underset{j\in\{1,2,..,n\}}{arg\,max}\left(\sum_{i=1}^{n} W_{ij}\right) \qquad (۱۶)$$

با حل رابطه (۱۶) تطابق بهینه برای بیشینه کردن تابع $\left(\sum_{i=1}^{n} W_{ij}\right)$ به دست می‌آید، که باعث رسیدن به حداکثر گذردهی شبکه می شود. این مسئله در مقاله [۱۳] مطرح شده و با استفاده از الگوریتم مجارستانی و به‌صورت آفلاین محاسبه و حل شده است. حال در شرایط نبود CSI در بخش بعدی فرم آنلاین (برخط) مسئله بیان می‌شود.

۳-۳- تخصیص منابع بدون CSI : فرمول‌بندی مبتنی بر CMAB

با توجه به تصادفی و ناشناخته بودن پارامتر اطلاعات وضعیت کانال (CSI) و عدم اطلاع از توزیع احتمال آن، مقدار SNR و در نتیجه مقدار نرخ ارسال به‌صورت تصادفی خواهد بود. هدف ما، یافتن یک تطابق بهینه وزن‌دار در گراف دوبخشی با وزن‌های ناشناخته است. روش پیشنهادی برای مدل‌سازی مسئله تطابق بهینه برخط، استفاده از چارچوب فرمول‌بندی Combinatorial multi-armed bandit [۲۷-۲۹] است. این فرمول‌بندی برخلاف فرمالیسم یادگیری تقویتی استاندارد (MDP) بدون حالت است و برای مدل‌سازی مسائلی مانند تطابق گراف بکار می‌رود که در آن‌ها انتخاب‌ها به‌صورت برداری و دارای ساختار ترکیباتی هستند. این چارچوب به‌طورکلی می‌تواند برای حل گستره وسیعی از مسائل بهینه سازی ترکیباتی در محیط‌های ناشناخته مورد استفاده قرار گیرد.

در مسئله CMAB، تصمیم گیرنده می‌بایست هر بار از میان مجموعه انتخاب (عمل) های موجود، بر اساس سیاست تعیین شده ترکیبی از انتخاب‌ها را برگزیند به‌گونه‌ای که پاداش آن بیشینه شده و یا به عبارتی پشیمانی آن کمینه شود. مجموعه انتخاب‌ها (اعمال) بازی به‌صورت مجموعه $K = \{1, \dots, k\}, |K| = Z$ نمایش داده می شود. $a_{k,t}$ مجموعه اعمال انتخاب شده در لحظه t می‌باشد. هنگامی‌که در زمان t عمل $a_{k,t}$ توسط تصمیم گیرنده انتخاب می شود، پاداش $X_{k,t}$ به ازای هر عمل انتخاب شده حاصل می‌شود که این پاداش تصادفی و مستقل از زمان است. مجموعه اعمال انتخاب شده در بازی بر اساس سیاست $\pi(t)$ تعیین می شود. در مسئله CMAB سیاستی مناسب است که تعادل سرمایه‌گذاری میان دو وضعیت بهره‌برداری و اکتشاف را لحاظ کرده باشد.

مجموع کلیه پاداش‌های در زمان t برای تمامی انتخاب‌ها از رابطه (۱۷) به دست می‌آید:

$$Reward_{a(t)}(t) = \sum_{k=1}^{z} a_k(t).X_k(t) \qquad (۱۷)$$

حداکثر پاداش به دست آمده برای تمامی اعمال از رابطه (۱۸) به دست می‌آید:

$$\mu^* = \max_{a\in f} \sum_{k=1}^{z} a_k \times \mu_k \qquad (۱۸)$$

در مسئله CMAB معیار بهینه بودن انتخاب‌ها، میزان پشیمانی نام دارد که اختلاف میان پاداش به دست آمده با پاداش مورد انتظار است، و هر چه کمتر باشد و به سمت صفر میل کند، بدان معناست که انتخاب‌ها بهتر بوده و پاداش بیشتری به دست آمده است.

$$Regret_t^{\pi} = T \times \mu^* - \mathbb{E}^{\pi}[\sum_{t=1}^{T} Reward_{\pi(t)}(t)] \qquad (۱۹)$$

اگر اعمال از طریق یک روش درست و سیاست مناسب برای بازی انتخاب شوند، مقدار پشیمانی با گذشت زمان سیر نزولی داشته و به صفر نزدیک می‌شود.

برای اتصال دو به دوی n تا از زوج مشترکین (تطابق دوبخشی مشترکین سلولی با زوج مشترک D2D)، n^2 اشتراک وجود دارد. با در نظر گرفتن مدکاری و سطوح قدرت سیگنال، کل انتخاب‌ها از رابطه (۲۰) حاصل می‌شود:

$$K = O(n!) \times Modes \times Power\ levels \qquad (۲۰)$$

بدین ترتیب تعداد اعمال در مسئله CMAB همان تعداد کل انتخاب‌ها در نظر گرفته می‌شود. هرکدام از این انتخاب‌ها نشان‌دهنده اشتراک منابع یک مشترک سلولی با زوج مشترک D2D با تعیین یک مُدکاری خاص و تعیین سطح قدرت ارسالی هر دو گروه مشترک توسط ایستگاه اصلی می‌باشد. منابع مشترکین سلولی به‌گونه‌ای با مشترکین D2D به اشتراک گذاشته می شود که هر مشترک سلولی تنها با یک جفت مشترک D2D اشتراک طیف فرکانسی داشته باشد. پاسخ به دست آمده از رابطه (۲۰) یک عدد بسیار بزرگ است که امکان آزمایش و بررسی تمامی این ترکیبات برای انتخاب بهترین ترکیب عملاً امکان‌پذیر نیست. در بخش بعدی نشان داده می‌شود که می‌توان با محدودسازی انتخاب‌ها و با حل برخط مسئله CMAB، بالاترین مجموع گذردهی درشبکه را به دست آورد.

میزان نرخ ارسالی از این اشتراک منبع به هر ارتباط توأمان سلولی و D2D در یک مدکاری خاص، پاداش به دست آمده از انتخاب هر اهرم در نظر گرفته می‌شود که در مسئله CMAB همان $X_{k,t}$ است. $\mu_{k,t}$ میانگین نرخ‌های ارسال اطلاعاتِ انتخاب یک منبع طیف، با یک سطح توان انتخابی و در یکی از سه مُدکاری (اهرم k)، تا زمان t می‌باشد. با توجه به اینکه تعداد اشتراک منابع برای n زوج ارتباط D2D لازم است، پس در هر زمان t باید n اهرم انتخاب شود که انتخاب این یال‌ها از $a_{k,t}$ به دست می‌آید. همچنین پاداش حاصل از انتخاب یال‌ها $Reward_{a(t)}(t)$ برابر میزان مجموع گذردهی شبکه در زمان t است.

برای به دست آوردن میزان پشیمانی در بازه زمانی t نیز خواهیم داشت

$$Regret(t) = Max\_Throughput - Avg\_Throughput_t \qquad (۲۱)$$

در رابطه (۲۱) نرخ ارسالی به دست آمده کلیه مشترکین فعال در هر لحظه به‌وسیله سیگنال‌های کنترلی به اطلاع ایستگاه اصلی می‌رسد. برای محاسبه حداکثر نرخ ارسالی قابل‌دستیابی نیز بهترین گذردهی ممکن در نظر گرفته می‌شود.

حال با توجه به فرمول‌بندی مسئله با CMAB، باید تعادل میان دو وضعیت بهره‌برداری و اکتشاف در سیاست‌گذاری مد نظر قرار گیرد، بدین صورت که ۱) با انتخاب بهترین مدکاری و اشتراک منابع سلولی (طیف وتوان) بین مشترکین سلولی و زوج D2D، به حداکثر میزان گذردهی در شبکه دست یافت و ۲) با توجه به تصادفی بودن نرخ ارسال به دست آمده از انتخاب‌ها، برای حداکثر ماندن گذردهی، اشتراک‌ها و انتخاب‌های جدید در شبکه سلولی کشف و در صورت بهینه بودن اجرا شوند.

در مسئله CMAB به شکل کلاسیک هر انتخاب به شکل یک موجودیت مستقل در نظر گرفته می‌شود. در این شرایط، مجموعه اعمال بسیار بزرگ خواهد بود و در کاربردهای واقعی با تعداد زیاد (نمایی) انتخاب نمی‌توان ازاین‌روش استفاده کرد. درصورتی‌که در ساختار مسئله از وابستگی اعمال استفاده شود، فضای عمل به شکل قابل‌توجهی کاهش یافته و مسئله در زمان قابل‌قبول و با کارایی مناسب قابل‌حل است. یکی از روش‌های بهینه‌سازی ترکیباتی با *CMAB* با کارایی مناسب، روش یادگیری با پاداش‌های خطی (LLR) است که در آن در هر زمان، یک ترکیب وزن‌دار خطی از انتخاب‌ها منجر به پاداش می‌شود [۳۱].

۳-۴- حل مسئله تخصیص منابع با الگوریتم LLR

الگوریتم LLR در مرجع [۳۱] برای همگرایی برخط و model-free به ساختارهای ترکیباتی بهینه‌ای ارائه شده است که برای آن‌ها الگوریتم‌های دقیق یا تقریبی با زمان چندجمله‌ای وجود داشته باشد. مسئله تطابق بیشینه گراف کامل وزن‌دار دوبخشی در نظریه گراف، یک مسئله ترکیباتی با کلاس پیچیدگی P است که برای آن الگوریتم زمان چندجمله‌ای مجار ستانی یا Kuhn-Mankres ارائه شده است. ازاین‌رو، می‌توان مسئله فوق را به شکل برخط با LLR حل کرد.

شبه کد الگوریتم پیشنهادی مبتنی بر LLR برای اختصاص منابع شامل توان و طیف برای مشترکین D2D در شکل ۴ ارائه شده است. این الگوریتم بر اساس فرمول‌بندی CMAB می‌باشد. در ادامه به توضیح روش LLR بر طبق الگوریتم ارائه شده پرداخته می‌شود [۳۱].

نمادهای مورد استفاده در الگوریتم تخصیص منابع LLR در جدول ۳ نشان داده شده است.

نحوه عملکرد الگوریتم به این صورت است که در هر بار اجرای حلقه اول (خطوط ۳ تا ۶)، یک انتخاب برای هر یک از مشترکان انجام شده و بقیه انتخاب‌ها به‌صورت تصادفی خواهد بود. با پایان یافتن حلقه، تمامی انتخاب‌ها حداقل یک‌بار بررسی شده‌اند و LLR اطلاعات کلیه انتخاب‌ها را جمع‌آوری کرده است. در حلقه اصلی برنامه (خطوط ۱۰ تا ۱۴)، در هر بار از بین تخصیص‌های ممکن، حالتی انتخاب می شود که رابطه (۱۹) را کمینه کرده و بیشترین گذردهی را برای تخصیص انجام شده به دست آورد.

```
1:  // Initialization
2:  N=max_a |A_a| ;
3:  for k=1 to Z
4:     select matching "a" such that k ∈ A_a ;
5:     update (X̄_k , T_k) that was selected ;
6:  end for
7:  t=Z ;
8:
9:  //Main Loop
10: while 1 do
11:    t=t+1 ;
12:    select the matching a, which solves the maximization
       problem
       a = arg max_{a∈f} Σ_{k∈A_a} a_k × ( X̄_k + sqrt( (N+1) ln t / T_k ) );   (۲۲)
13:    update (X̄_k, T_k) that was selected in a;
14: end while
```

شکل ۴: شبه کد الگوریتم تخصیص منابع با روش LLR

جدول ۳: نمادهای مورد استفاده در الگوریتم LLR

نماد	تعریف
Z	تعداد تمامی حالات ممکن انتخاب طیف، توان و مُدکاری (تمامی اهرم‌های بازی).
A_a	مجموعه انتخاب‌هایی که در لحظه باید انجام شود، که با تعداد زوج ارتباط *D2D* برابر است
k	هریک از انتخاب‌ها، که شامل طیف، توان و مُد کاری است (هر یک از اهرم‌های بازی)
t	تعداد دفعات انجام انتخاب تا آخرین لحظه (تعداد دفعاتی که بازی شده است)
$\overline{X_k}$	میانگین نرخ‌های ارسال به دست آمده از انتخاب k ام تا آخرین لحظه (میانگین پاداش به دست آمده از اهرم k)
T_k	تعداد دفعاتی که اهرم k تا آخرین لحظه انتخاب شده است
f	مجموعه سیاست انتخاب اهرم‌ها در هر لحظه
a	انتخاب مجموعه اهرم‌ها در هر لحظه(تعیین اشتراک زوج مشترک *D2D* با مشترک سلولی و تعیین مدکاری)

در رابطه (۲۲) برای تعیین سیاست انتخاب اهرم‌های بازی از روش UCB کمک گرفته شده است [۳۲]. در این روش برای هر اهرم یک شاخص تعیین می‌شود، در هر بار بازی اهرم‌هایی که شاخص آن‌ها بیشترین بوده انتخاب شده و بر روی آن‌ها سرمایه‌گذاری می شود. این شاخص باید بگونه ای باشد که تعادل میان سرمایه‌گذاری و اکتشاف در CMAB را رعایت کند. شاخص هر انتخاب در UCB از رابطه زیر محاسبه می‌شود:

$$Indicator_{k,t} = \bar{X}_{k,t} + \sqrt{\frac{(N+1)\ln t}{T_{k,t}}} \qquad (۲۳)$$

در رابطه (۲۳)، t تعداد دفعات اجرای انتخاب از ابتدا تاکنون، N تعداد اهرم‌هایی که باید در هر بار انتخاب باید بازی شوند، $\bar{X}_{k,t}$ ،

میانگین پاداش به دست آمده از اهرم k تا زمان t و $T_{k,t}$، تعداد دفعاتی است که اهرم k از ابتدا تا زمان t انتخاب شده است.

عبارت ماکزیمم آرگومان در رابطه (۲۲) بدان معناست که اهرم‌ها (اشتراک منابع سلولی برای زوج مشترک D2D و تعیین مُدکاری در شبکه) در لحظه t بگونه ای انتخاب شوند که بتوان به حداکثر مقدار پاداش (گذردهی در شبکه) دست یافت. حل مسئله ماکزیمم آرگومان به‌وسیله الگوریتم مجارستانی انجام می‌شود.

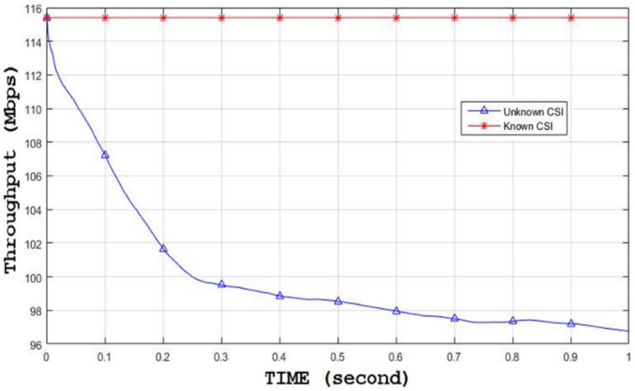
۴- ارزیابی روش پیشنهادی و نتایج

در این بخش مدل ریاضی و الگوریتم محاسبه گذردهی و نسبت قطعی شبکه در هنگام اشتراک منابع سلولی با مشترکین پیوند D2D شبیه‌سازی شده و نتایج به دست آمده ارائه می‌شود. میزان گذردهی و نسبت قطعی شبکه در حالت واقعی و بدون داشتن پارامتر CSI بررسی می‌شود و درنهایت شبکه بدون داشتن پارامتر CSI شبیه‌سازی می‌شود و با استفاده از روش LLR عملیات انتخاب مُدکاری و اختصاص منابع (شامل طیف و توان) صورت می‌گیرد.

در این بخش، ابتدا نمودارهای شبیه سازی همکاری D2D با شبکه سلولی با فرض داشتن پارامتر CSI [۱۳] ترسیم می‌شود. سپس شبیه‌سازی بر اساس روش پیشنهادی انجام شده و این دو با هم مقایسه می‌شوند.

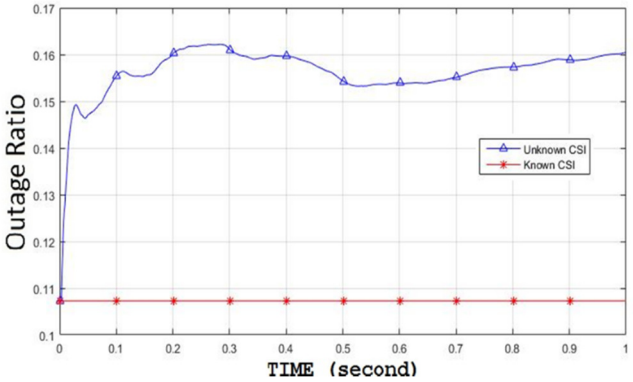
۴-۱- ارزیابی روش پیشنهادی مرجع [۱۳]

اگر ایستگاه اصلی پارامتر CSI را در اختیار داشته باشد، ظرفیت هر کانال ارتباطی در سه مُد کاری و بر اساس سه سطح مختلف توان محاسبه شده و وزن همه حالت‌های اشتراک منابع سلولی با ارتباط D2D به دست آورده می‌شود. سپس با استفاده از الگوریتم مجارستانی بهترین تطابق دوبخشی صورت گرفته و منابع سلولی به اشتراک گذاشته می‌شود [۱۳]. اما همان‌گونه که در بخش ۲ اشاره شد، پارامتر CSI هر لحظه در حال تغییر است و تخمین آن نیز دقت بسیار پایینی دارد[۱۱،۱۰]. در نتیجه حتی اگر در ابتدا پارامتر CSI درست تخمین زده شده باشد، بعد از گذشت زمان بسیار اندک با تغییر پارامترهای شبکه، میزان گذردهی کاهش چشم‌گیری خواهد داشت (شکل ۵).



شکل ۵: مقایسه میزان گذردهی در دو حالت دانستن و ندانستن CSI

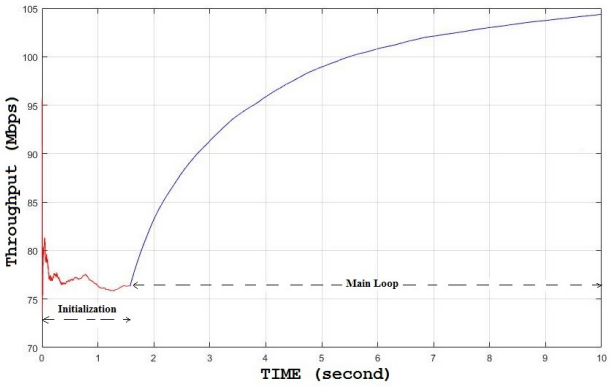
در حالت همکاری مشترکین سلولی و مشترکین D2D، امکان رله کردن بسته ارسالی مشترکین سلولی به ایستگاه اصلی فراهم می‌شود، بدین ترتیب دو مسیر برای ارسال اطلاعات مشترک سلولی فراهم است که این امر باعث کاهش احتمال قطعی در شبکه سلولی و بالا رفتن نرخ ارسال مشترکین آن خواهد شد، اما به دلیل تخمین نادرست CSI، احتمال قطعی شبکه نیز بعد از چند میلی‌ثانیه بالارفته و کیفیت سرویس در شبکه به حالت عادی باز می‌گردد (شکل ۶).



شکل ۶: نمودار مقایسه نسبت قطعی شبکه در دو حالت دانستن و ندانستن CSI

۴-۲- عدم نیاز به CSI در الگوریتم LLR

الگوریتم مطرح شده در شکل ۷، یک راهکار بهینه برای اشتراک منابع (شامل طیف فرکانسی و توان) بین ارتباطات سلولی و ارتباطات *D2D* است. در شکل ۷، این الگوریتم برای اشتراک منابع مورد استفاده قرار گرفته است و نمودار گذردهی شبکه ترسیم شده است.



شکل ۷: میزان گذردهی شبکه با الگوریتم LLR

همان‌گونه که در شکل ۷ مشهود است، در ۱۵۷۵ میلی‌ثانیه اول هیچ بهبودی در میزان گذردهی صورت نمی‌گیرد، می‌توان گفت الگوریتم مشغول جمع‌آوری اطلاعات از شبکه و به دست آوردن میزان گذردهی در تمامی حالت‌های ممکنِ اشتراک‌های منابع سلولی با ارتباط *D2D* است. سپس با به دست آوردن اطلاعات لازم روند اشتراک منابع با روش بهینه‌سازی صورت می‌گیرد و بهبود چشمگیری در میزان گذردهی شبکه رخ می‌دهد. همان‌طورکه در بخش قبل ذکر شد، این

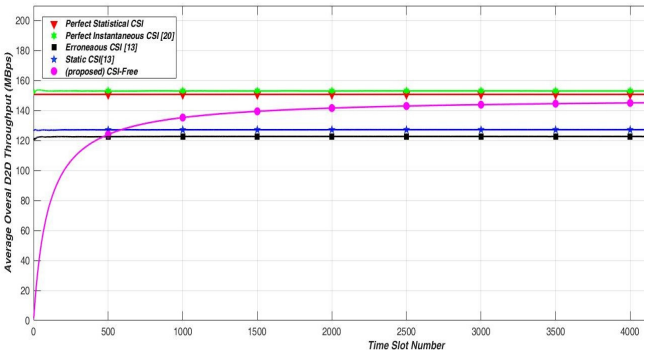
الگوریتم با گذشت زمان عملکرد بهتری پیدا می‌کند و می‌تواند به بالاترین حد خود برسد.

۴-۳- مقایسه روش پیشنهادی با سایر روش‌ها

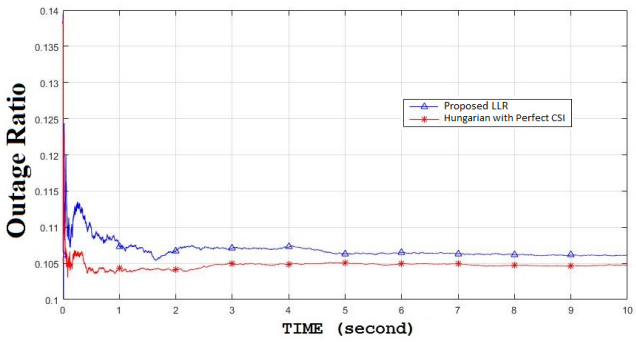
روش پیشنهادی ما نیازی به داشتن پارامتر "اطلاعات وضعیت کانال" ندارد، و میزان گذردهی از طریق روش LLR و بازخورد نرخ ارسال هر انتخاب بهینه می‌شود. شکل ۸ میزان گذردهی روش پیشنهادی را با الگوریتم مطرح در مرجع [۱۳] و همچنین بهینه‌سازی گذردهی در شرایط در اختیار داشتن مقادیر دقیق لحظه‌ای CSI [۲۲] مقایسه کرده است. در مقاله [۱۳] فرض بر داشتنِ پارامتر "اطلاعات وضعیت کانال" و عدم تغییر آن است. همان‌طورکه در شکل ۸ مشهود است، گذردهی کل شبکه در وضعیت داشتن پارامتر CSI حدود ۴/۵ درصد بالاتر از روش پیشنهادی ما است، این در حالی است که دستیابی به این پارامتر کاملاً فرضی است و جنبه اجرایی ندارد. همچنین، الگوریتم پیشنهادی ما عملکرد بهتری نسبت به روش تخمین CSI در مرجع [۱۳] از خود نشان می‌دهد.

نسبت قطعی در شبکه‌های سلولی، یک پارامتر تأثیرگذار بر روی کیفیت سرویس شبکه‌های سیار است. لذا این شاخص در شکل ۹ در دو حالت داشتن پارامتر CSI [۱۳] و بهینه‌سازی شبکه با روش پیشنهادی مقایسه شده است. کاهش نسبت قطعی در شبکه سلولی بر پایه داشتن CSI معتبر نیست.

در روش پیشنهادی، کاهش احتمال قطعی به‌وسیله تخصیص منابع با روش LLR انجام شده است. قطعی در شبکه سلولی زمانی رخ می دهد که نرخ ارسال برای مشترک سلولی کمتر از ۴ مگابیت بر ثانیه شود [۲۸].



شکل ۸: مقایسه گذردهی شبکه روش پیشنهادی با سایر روش‌ها

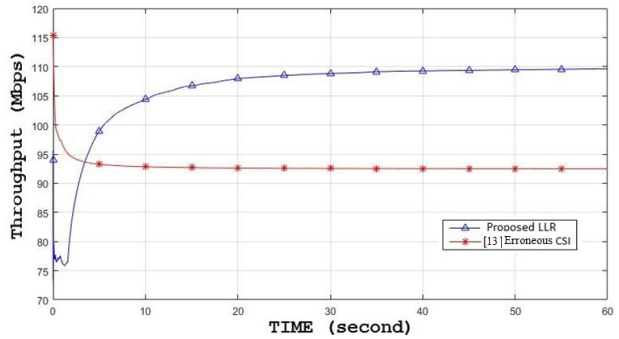


شکل ۹: مقایسه نسبت قطعی شبکه در مرجع [۲۰] با وضعیت بهینه‌سازی با روش پیشنهادی

نتیجه شبیه‌سازی نشان می‌دهد که روش "LLR" در مورد شاخص نسبت قطعی شبکه، به‌خوبی عمل کرده و روش پیشنهادی تفاوت چندانی با بهترین حالت که داشتن پارامتر CSI است، ندارد (شکل۹).

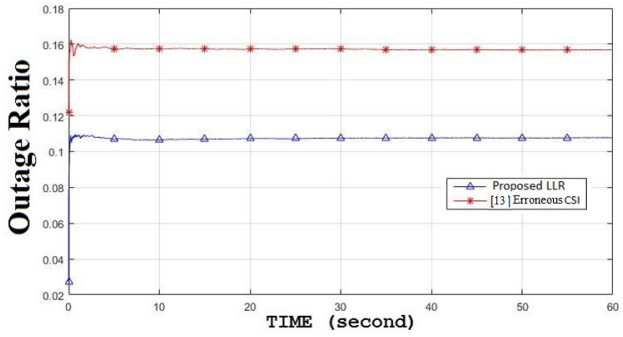
۴-۴- مقایسه روش پیشنهادی با روش تخمین CSI

در این بخش روش پیشنهادی پایان‌نامه با دو روش متکی بر CSI ایستا و CSI تخمینی مقاله [۱۳] مقایسه شده است. در مقاله [۱۳] تخمین پارامتر CSI درست نبوده و اتکا بر این پارامتر باعث بهبود حداکثری گذردهی شبکه نمی‌شود. در نمودار شکل ۱۰ میزان گذردهی در هر دو وضعیت ذکر شده، شبیه‌سازی و رسم شده است.



شکل ۱۰: مقایسه گذردهی در روش پیشنهادی با تخمین CSI [۱۳]

نتایج به دست آمده نشان می‌دهد که در عمل به دلیل عدم توانایی در تخمین درستِ پارامتر CSI، افت شدید میزان گذردهی شبکه روی می‌دهد و این پارامتر تا سطح ۹۳ مگابیت برثانیه کاهش می‌یابد. اما با استفاده از روش پیشنهادی و عدم نیاز به پارامتر CSI می‌توان گذردهی را تا ۱۱۰ مگابیت بر ثانیه افزایش داد. در شکل ۱۱ نسبت قطعی شبکه سلولی در دو روش فوق ترسیم شده است. بررسی بیشتر و مقایسه این دو حالت نشان می‌دهد که نسبت قطعی شبکه در روش پیشنهادی ۵ درصد نسبت به حالت نداشتن پارامتر CSI بهبود یافته است.

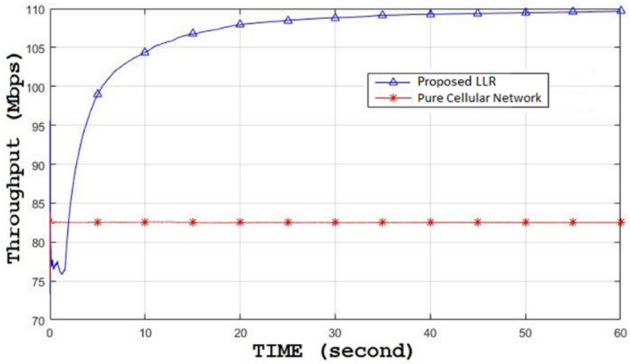


شکل ۱۱: مقایسه نسبت قطعی شبکه در روش پیشنهادی با روش تخمین CSI [۱۳]

۴-۵- مقایسه روش پیشنهادی با شبکه سلولی بدون مشارکت D2D

در این قسمت میزان گذردهی شبکه در دو وضعیت سلولی بدون مشارکت D2D و شبکه در حالت اشتراک منابع سلولی با زوج ارتباط D2D مقایسه می‌شود. اشتراک منابع به روش بهینه‌سازی با روش پیشنهادی انجام می‌شود (شکل ۱۲).

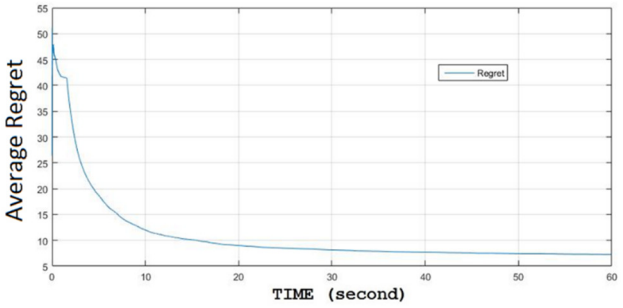
همان‌طورکه در شکل ۱۲ مشهود است، گذردهی بدون اشتراک منابع با پیوند D2D مقدار ثابتی داشته و از ۸۳ مگابیت برثانیه بالاتر نمی‌رود، ولی با اشتراک منابع با پیوند D2D این مقدار به ۱۱۰ مگابیت بر ثانیه افزایش داشته است که نشان‌دهنده افزایش بهره‌وری شبکه و کارایی بالای طیفی در روش پیشنهادی است.



شکل۱۲: مقایسه گذردهی در روش پیشنهادی با شبکه بدون اشتراک

۴-۶- بررسی بهینگی روش پیشنهادی

در الگوریتم‌های بهینه سازی شاخصی برای میزان بهینه بودن عملکرد الگوریتم در نظرگرفته می‌شود. همان‌طورکه در بخش ۳ توضیح داده شد، شاخص بهینه بودن الگوریتم برای راهکارهای مبتنی بر MAB محاسبه میزان پشیمانی است. هرچه مقدار این شاخص کمتر و به صفر نزدیک‌تر باشد، الگوریتم عملکرد بهتری دارد. بر این اساس در شکل ۱۳ نمودار شاخص پشیمانی برای میزان گذردهی شبکه رسم شده است.

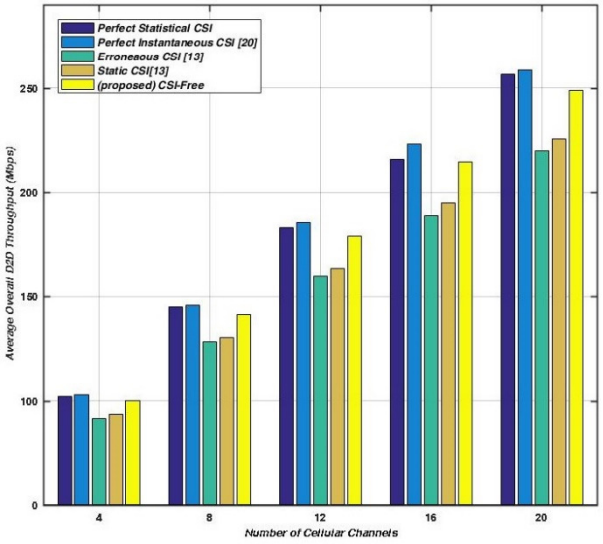


شکل ۱۳: تغییرات پشیمانی روش پیشنهادی برای گذردهی شبکه

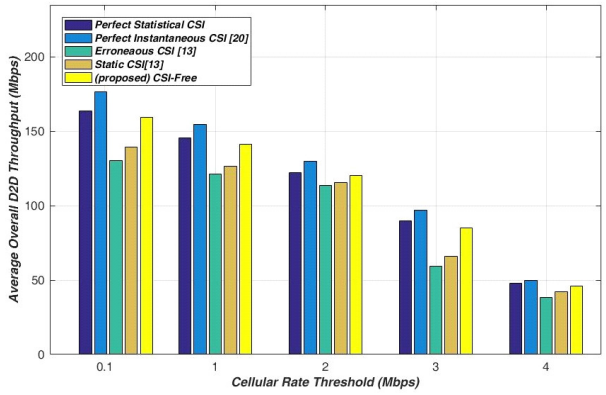
براساس نتایج شبیه‌سازی، میزان پشیمانی در بخش اولیه الگوریتم بالاست، ولی بعد از جمع‌آوری اطلاعات اولیه و شروع بهینه‌سازی، پشیمانی به‌شدت سیر نزولی دارد و به صفر میل می‌کند.

۴-۷- تحلیل حساسیت روش پیشنهادی

در این بخش، ما دو پارامتر تعداد کانال‌های سلولی و آستانه نرخ سلولی را تغییر داده و حساسیت روش پیشنهادی به این پارامترها را بررسی کرده‌ایم. همان‌گونه که در شکل ۱۴ نشان داده شده است، روش پیشنهادی در تعداد کانال‌های سلولی مختلف گذردهی بیشتری نسبت به روش‌های مبتنی بر CSI ثابت و یا خطادار [۱۳] از خود نشان می دهد. همچنین گذردهی الگوریتم پیشنهادی به الگوریتم مرجع [۲۲] که مبتنی بر فرض داشتن اطلاعات لحظه‌ای وضعیت کانال است، نزدیک است. با تغییر آستانه نرخ سلولی در شکل ۱۵ نیز مشاهده می شود که RL-D2D علیرغم در اختیار نداشتن اطلاعات وضعیت کانال، به نرخ گذردهی بالاتری نسبت به روش تخمین CSI [۱۳] دست یافته است و گذردهی آن به شرایط بهینه (در اختیار داشتن CSI لحظه‌ای مرجع [۲۲]) بسیار نزدیک است. این ارزیابی‌ها نشان می‌دهد که روش پیشنهادی می‌تواند علیرغم ناشناخته بودن CSI، ترکیب بهینه پارامترهای شبکه را انتخاب کرده و به نرخ گذردهی مناسب دست یابد.



شکل ۱۴: مقایسه گذردهی روش پیشنهادی با سایر روش‌ها با تعداد کانال‌های سلولی متغیر



شکل ۱۵: مقایسه گذردهی روش پیشنهادی با سایر روش‌ها با آستانه نرخ متغیر

۵- نتیجه‌گیری

در تحقیقات و مدل‌های موجود ارائه شـــده برای شـــبکه‌های ترکیبی D2D- سلولی، فرض می شود که پارامتر CSI موجود است. تخمین این پارامتر در شبکه‌های سلولی، کاری دشوار و نادقیق است. در بسیاری از مدل‌های ارائه شده پیشین، انتخاب مُدکاری مدنظر قرار نگرفته است و معمولاً از یک مد کاری ثابت در شبکه ترکیبی استفاده شده است. در این مقاله، یک شبکه سلولی با فرض نبود پارامتر CSI مدل سازی شد. برای این منظور، ابتدا منابع شبکه سلولی به اشتراک گذاشته شده و مدکاری انتخاب شـــد. با گذشـــت زمان و اســـتفاده از بازخورد میزان گذردهی تمامی مشترکین در شبکه، با استفاده از یک روش یادگیری تقویتی، اشـــتراک منابع میان مشـــترکین بهبود داده شـــد و گذردهی مجموع مشترکین به حداکثر مقدار ممکن نزدیک شد. برای انجام این کار، مسئله تخصیص منابع سلولی و اشتراک مشترکین عادی سلولی با زوج مشترک D2D در قالب CMAB فرمول‌بندی شد. سپس مسئله انتخاب بهینه برای رسـیدن به بیشـینه مجموع گذردهی مشـترکین با استفاده از روش LLR حل شد. نتایج شبیه سازی و نمودارهای ترسیم شـده از آن نشـان داد که گذردهی تا حد زیادی به مقدار بیشـینه در شبکه نزدیک شده است.

۶- مراجع

[1] M. Rebato,M. Mezzavilla,S. Rangan,M. Zorzi, "Resource sharing in 5G mmWave cellular networks," *IEEE Conference on Computer Communications Workshops,* 2016.

[2] M. Salahuddin, K. Alam, "Information and Communication Technology, electricity consumption and economic growth in OECD countries: A panel data analysis," *International Journal of Electrical Power & Energy Systems,* vol. 76, pp. 185-193, 2016.

[3] S. Zhang , Y. Hou , X. Xu , X. Tao, "Resource allocation in D2D-based V2V communication for maximizing the number of concurrent transmissions," *IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications,* pp. 1-6, 2016.

[4] J. Sachs, I. Maric , A. Goldsmith, "Cognitive Cellular Systems within the TV Spectrum," *IEEE Symposium on New Frontiers in Dynamic Spectrum (DySPAN),* pp. 1-12, 2010.

[5] V. Chandrasekhar,J.G.Andrew,Femtocell network: A Survey,IEEE comminucations Magazine, Vol 46, no.9.2008.

[6] A. Abdelhadi , T. C. Clancy, "Optimal context-aware resource allocation in cellular networks," *International Conference on Computing, Networking and Communications (ICNC),* pp. 1-5, 2016.

[7] K. Doppler , M. Rinne, C. Wijting,C. B. Ribeiro, "Device-to-device communication as an underlay to LTE-advanced networks," *IEEE Communications Magazine,* vol. 47, no. 12, pp. 42-49, 2009.

[8] R. Wang,D. Cheng,G. Zhang,Y. Lu,J. Yang,L. Zhao,K. Yang, "Joint relay selection and resource allocation in cooperative device-to-device communications," *AEU - International Journal of Electronics and Communications,* vol. 73, pp. 50-58, 2017.

[9] S. Mallick,R. A. Loodaricheh ,K. N. R. Surya Vara PrasadVijay Bhargava, "Resource Allocation for Cooperative D2D Communication Networks," *5G Mobile Communications* , pp. 531-570, 2016.

[10] B. S. Thian,A. Goldsmith, "Decoding for MIMO Systems with Imperfect Channel State Information," *IEEE Global Telecommunications Conference GLOBECOM,* 2010.

[11] K.Ghavame,M.Naraghi, , "MIMO Detection With Imperfect Channel State Information Using Expectation Propagation," *IEEE Transactions on Vehicular Technology,* vol. 66, no. 9, 2017.

[12] S. A. Ramprashad , G. Caire, "Cellular vs. Network MIMO: A comparison including the channel state information overhead," IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications, 2009.

[13] J.Han , Q. Cui ,C.Yang ,X.Tao , "Bipartite matching approach to optimal resource allocation in device to device underlaying cellular network, Electronics letters ," vol. 50, no. 3, pp. 212-214.,2014.

[14] C. H. Yu,K. Doppler , C. B. Ribeiro ,O. Tirkkonen , "Resource Sharing Optimization for Device-to-Device Communication Underlaying Cellular Networks," *IEEE Transactions on Wireless Communications,* vol. 10, no. 8, pp. 2752 - 2763, 2011.

[15] H. Min,J. Lee,S. Park,D. Hong, "Capacity Enhancement Using an Interference Limited Area for Device-to-Device Uplink Underlaying Cellular Networks," *IEEE Transactions on Wireless Communications,* vol. 10, no. 12, pp. 3995 - 4000, 2011.

[16] Y. Yi, J. Zhang , Q. Zhang , T. Jiang , J. Zhang, "Cooperative Communication-Aware Spectrum Leasing in Cognitive Radio Networks," *IEEE Symposium on New Frontiers in Dynamic Spectrum (DySPAN),* pp. 1-11, 2010.

[17] Z. Liu,T. Peng, S. Xiang, W. Wang, "Mode selection for Device-to-Device communication under LTE-Advanced networks," *IEEE International Conference on Communications,* pp. 5563-5567, 2012.

[18] Y. Pei , Y.C. Liang, "Resource Allocation for Device-to-Device Communications Overlaying Two-Way Cellular Networks," *IEEE Transactions on Wireless Communications,* vol. 12, no. 7, pp. 3611-3621, 2013.

[19] R. Chithra, R. Bestak,S. Patra, "Hungarian Method Based Joint Transmission Mode and Relay Selection in Device-to-Device Communication," *8th IFIP Wireless and Mobile Networking Conference (WMNC),* pp. 261-268, 2015.

[20] علیرضا عبدالله پوری، گلاله عزیزی ، «تخصیص منابع در شبکه‌های WiMax مبتنی بر OFDMA برای سیستم‌های IPTV با استفاده از الگوریتم ژنتیک»، مجله مهندسی برق دانشگاه تبریز، جلد ۴۶، شماره ۳، ص. ۲۶۷-۲۷۶، ۱۳۹۵.

[21] فرهاد دانائی یگانه، افشین ابراهیمی، «مدیریت انتخاب مجدد سلول در نسل‌های مختلف شبکه‌های سلولی مبتنی بر 3GPP و تحلیل دو چالش یک اپراتور داخلی»، مجله مهندسی برق دانشگاه تبریز، جلد ۴۶، شماره ۳، ص. ۱۶۱-۱۷۹، ۱۳۹۵.

[22] Y. Cao, T. Jiang and C. Wang, "Cooperative device-to-device communications in cellular networks," IEEE Wireless Communications, Vol. 22, No. 3, pp. 124-129, 2015.

[23] H. Shin, Y. Jin Sang, and J. G. Andrews., Outage probability for heterogeneous cellular networks with biased cell association, IEEE Global Telecommunications Conference GLOBECOM, pp. 1-5 , 2011.

[24] L. Lei and Z. Zhong.,, Operator Controlled Device-to-Device Communications in LTE-Advanced Networks, IEEE Wireless Communications, Vol. 19, No. 3, pp. 96–104,2012.

[25] N. Chen , H. Tian , Z. Wang, "Resource Allocation for Intra-Cluster D2D Communications Based on Kuhn-Munkres Algorithm," IEEE 80th Vehicular Technology Conference, pp. 1-5, 2014.
[26] B. Zhou , H. Hu, S. Q. Huang, H. H. Chen, "Intracluster Device-to-Device Relay Algorithm With Optimal Resource Utilization," *IEEE Transactions on Vehicular Technology,* vol. 62, no. 5, pp. 2315-2326, 2013.
[27] S. Chen, T. Lin, I. King, M. R. Lyu, W. Chen, "Combinatorial Pure Exploration of Multi-Armed Bandits," *Advances in Neural Information Processing Systems 27 (NIPS),* pp. 379-387, 2014.
[28] Y. Gai, B. Krishnamachari, R. Jain, "Learning Multiuser Channel Allocations in Cognitive Radio Networks: A Combinatorial Multi-Armed Bandit Formulation," *IEEE Symposium on New Frontiers in Dynamic Spectrum,* pp. 1-6, 2010.
[29] Y. Gai , B. Krishnamachari , M. Liu, "On the Combinatorial Multi-Armed Bandit Problem with Markovian Rewards," *IEEE Global Telecommunications Conference ,* pp. 1-6, 2011.
[30] Ontanón, S., The combinatorial multi-armed bandit problem and its application to real-time strategy games In Ninth Artificial Intelligence and Interactive Digital Entertainment Conference, 2013.
[31] Y.Gai, B. Krishnamachari, and R. Jain, Combinatorial network optimization with unknown variables: Multi-armed bandits with linear rewards and individual observations, IEEE/ACM Transactions on Networking, Vol. 20, No.5, pp. 1466-1478, 2012.
[32] Garivier, A., & Moulines, E., On upper-confidence bound policies for switching bandit problems, In International Conference on Algorithmic Learning Theory, pp. 174-188, 2011.

۱ Device to Device
۲ Channel State Information
۳ Reinforcement-Learning-based D2D communications
٤ Combinatorial Multi-Armed Bandit
٥ Learning with Linear Rewards
٦ Relay
۷ Overlay
۸ Underlay
۹ Cellular User
۱۰ Device Transmitter-Device Receiver