Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
Volume 13, Number 1, Spring 2024, pp. 25-38
Received: 10/28/2023, Revised: 07/11/2024, Accepted: 10/03/2024

AoI-based Performance Optimization in RIS-Assisted mmWave Communications

Marzieh Sheikhi[1], Somayeh Asadi[2], and Vesal Hakami[3*]
1- School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran.
2- School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran.
3*- Center of Excellence in Future Networks, School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran.
[1] marzieh_sheikhi@comp.iust.ac.ir, [2] somaye_asadi@comp.iust.ac.ir, and [3*] vhakami@iust.ac.ir

Corresponding author's address: Vesal Hakami, Center of Excellence in Future Networks, School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran.

Abstract- With the advent of the next generation of wireless communication networks, the requirements for high-rate and timely information delivery, especially in real-time applications, have increased dramatically. In particular, Millimeter wave (mmWave) communications can provide a high data rate, and reconfigurable intelligent surfaces (RISs) can reduce the blockage sensitivity of mmWave links. Also, to quantify the data freshness, the age of information (AoI) metric is introduced in the recent literature, defined as the time elapsed since the last successfully received status update was generated. This paper aims to optimize the average AoI in a multi-user RIS-assisted millimeter wave network. For this purpose, the proposed problem was formulated as a Markov decision process (MDP) that aims to obtain the optimal control policy that minimizes the sum of the average AoI of all users. Under the unknown system statistics assumption, we propose a model-free deep reinforcement learning algorithm to schedule the users' transmission, adjust the transmit power, and configure the RIS to reflect the signals. The performance of the proposed method is evaluated in terms of convergence characteristics and the impact of changes in network parameters.

Keywords- Age of Information, Reconfigurable Intelligent Surface, Millimeter Wave, Model-free Deep Reinforcement Learning.

مجله علمی رایانش نرم و فناوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
جلد ۱۳، شماره ۱، بهار ۱۴۰۳ ، صفحه ۲۵-۳۸
دریافت: ۱۴۰۲/۰۸/۰۶، بازنگری: ۱۴۰۳/۰۴/۲۱، پذیرش: ۱۴۰۳/۰۷/۱۲

بهینه‌سازی کارایی مبتنی بر دیرینگی اطلاعات در ارتباطات امواج میلی‌متری و مجهز به صفحات قابل بازپیکربندی هوشمند

مرضیه شیخی[۱]، سمیه اسدی[۲] و وصال حکمی*[۳]
۱- دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران.
۲- دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران.
*۳- دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران.
[1]marzieh_sheikhi@comp.iust.ac.ir, [2]somaye_asadi@comp.iust.ac.ir, [3*]vhakami@iust.ac.ir

* نشانی نویسنده مسئول: وصال حکمی، تهران، میدان رسالت، خیابان هنگام، دانشگاه علم و صنعت ایران، دانشکده مهندسی کامپیوتر.

چکیده – با ظهور نسل جدیدی از شبکه‌های ارتباطی بی‌سیم، نیازمندی به نرخ داده بالا و دریافت به‌موقع اطلاعات به‌ویژه در برنامه‌های کاربردی بلادرنگ به‌طور چشمگیری افزایش‌یافته است. در تحقیقات اخیر، برای سنجش تازگی اطلاعات، از معیار دیرینگی اطلاعات (AoI) استفاده می‌شود که به‌عنوان زمان گذشته از تولید آخرین بسته موفق دریافت شده در مقصد تعریف می‌شود. به علت ماهیت تصادفی این معیار، در این مقاله، مسئله بهینه‌سازی متوسط تخفیف یافته هزینه مبتنی بر دیرینگی اطلاعات در شبکه‌های بی‌سیم چندکاربره مبتنی بر امواج میلی‌متری و مجهز به سطوح بازتابی هوشمند (RIS) موردبررسی قرار می‌گیرد. برای این منظور، با استفاده از چارچوب ریاضی فرآیند تصمیم مارکوف تخفیفی، مسئله مدل می‌شود که هدف آن بهینه‌سازی متوسط تخفیف یافته مبتنی بر AoI همه کاربران است. با توجه به ماهیت تصادفی مسئله و عدم در دسترس بودن اطلاعات کامل یا آماری از سیستم، از یک الگوریتم فارغ از مدل مبتنی بر یادگیری تقویتی عمیق بنام گرادیان سیاست قطعی عمیق (DDPG) برای حل مسئله استفاده می‌شود که بتوان زمان‌بندی کاربران، توان ارسالی آن‌ها، انتقال فاز در RIS و شکل‌دهی پرتوی دریافتی در BS را محاسبه کرد. عملکرد روش پیشنهادی ازنظر ویژگی همگرایی و تغییرات پارامترهای مختلف در شبکه موردبررسی قرار می‌گیرد و نتایج با روش‌های دیگر و شبکه‌های بدون RIS موردبررسی قرار می‌گیرد. نتایج به‌دست‌آمده نشان می‌دهد که عملکرد روش پیشنهادی نسبت به رویکردهای دیگر سبب کاهش متوسط تخفیف یافته AoI کاربران خواهد شد.

واژه‌های کلیدی: دیرینگی اطلاعات، امواج میلی‌متری، صفحات قابل بازپیکربندی هوشمند، یادگیری تقویتی عمیق فارغ از مدل.

۱- مقدمه

شبکه‌های تلفن همراه نسل جدید با استفاده از فناوری‌های پیشرفته، امکان ارائه برنامه‌های کاربردی حساس به زمان را فراهم می‌کنند. این برنامه‌ها نیازمند تازگی[۱] اطلاعات هستند، زیرا تصمیم‌گیری بر اساس داده‌های منسوخ می‌تواند عواقب جبران‌ناپذیری داشته باشد [۱]. برای سنجش تازگی داده، معیار دیرینگی اطلاعات[۲] به کار گرفته می‌شود که زمان گذشته از تولید آخرین بسته موفق دریافت شده در مقصد تعریف می‌گردد [۲]. این معیار با معیارهای سنتی مبتنی بر زمان مانند تأخیر[۳] و گذردهی[۴] تفاوت دارد. به عبارتی، به‌روزرسانی به‌موقع اطلاعات مبدأ در مقصد لزوماً به معنای بیشینه‌سازی بهره‌وری[۵] شبکه یا حتی تضمین دریافت پیام‌ها با حداقل تأخیر نیست [۳].

با پیشرفت فناوری، شبکه‌های تلفن همراه نسل جدید از فرکانس‌های بالاتر استفاده می‌کنند. این فرکانس‌ها مانند موج میلی‌متری[۶] در شبکه‌های 5G و امواج ترا هرتز[۷] در شبکه‌های 6G، سبب کاهش تأخیر و بهبود گذردهی می‌شوند. بااین‌حال ارتباطات مبتنی بر این امواج به دلیل پراکندگی[۸] و تضعیف بالا، در محیط‌های انتشار بی‌سیم با چالش‌های زیادی روبرو هستند [۴]. به‌عنوان‌مثال، وجود موانع و اثر انسداد[۹] باعث کاهش شدید کیفیت لینک ارتباطی می‌شود.

برای حل این چالش‌ها، استفاده از صفحات قابل بازپیکربندی هوشمند[۱۰] به‌عنوان یک راه‌حل جدید پیشنهاد شده است. RIS از عناصر منفعلی تشکیل‌شده است که قادر به تغییر فاز[۱۱] سیگنال‌های دریافتی هستند. با تغییر فاز مناسب عناصر RIS، سیگنال‌های منعکس‌شده را می‌توان به نقطه موردنظر هدایت کرد و کیفیت لینک را بهبود داد. درنتیجه، با استفاده از RIS، نرخ ارسال اطلاعات را می‌توان افزایش داد و فرآیند به‌روزرسانی به‌موقع اطلاعات را باکیفیت مناسب پشتیبانی کرد [۵-۷].

دیرینگی اطلاعات یک معیار کلیدی برای ارزیابی تازگی اطلاعات در شبکه‌های بی‌سیم است. در این زمینه، تاکنون پژوهش‌های متعددی برای بهینه‌سازی متوسط دیرینگی اطلاعات انجام‌شده است. برخی از این پژوهش‌ها مسئله را بدون بهره‌گیری از فناوری‌های مورداستفاده در نسل جدید و تنها با اتکا به استفاده از روش‌های مختلف سیاست زمان‌بندی حل می‌کنند. این کاربردها به دلیل ماهیت پویایی محیط انتشار بی‌سیم با چالش‌هایی تضعیف سیگنال و گذردهی پایین روبرو هستند. بر این اساس بهره‌گیری از فناوری‌های جدید، می‌توانند به بهبود کارایی شبکه‌های بی‌سیم در شرایط نامساعد محیط انتشار کمک کنند و باعث بهبود نرخ ارسال اطلاعات شود که درنتیجه آن فرآیند به‌روزرسانی به‌موقع اطلاعات به‌منظور ارائه کاربردهای بلادرنگ می‌تواند باکیفیت مناسبی صورت پذیرد.

برای بهینه‌سازی متوسط دیرینگی اطلاعات شبکه در بلندمدت نیازمند محاسبه سیاست کنترلی تطبیق‌پذیر با وضعیت جاری سیستم خواهیم بود. برای محاسبه چنین سیاستی، روش‌های مطرح‌شده در پژوهش‌های گذشته را می‌توان برحسب فرض اطلاعاتی به کار گرفته‌شده در آن‌ها به سه دسته تقسیم کرد که در زیر بخش بعدی به‌مرور آن‌ها می‌پردازیم.

۱-۱- مرور کارهای پیشین

در این بخش، بر اساس نوع فرض اطلاعاتی، کارهای پیشین را در زمینه بهینه‌سازی متوسط دیرینگی اطلاعات دسته‌بندی و مرور می‌نماییم:

فرض اطلاعاتی کامل (غیر سببی): در این دسته از پژوهش‌ها، بهینه‌سازی با این فرض انجام می‌گیرد که اطلاعات دقیق و کامل مربوط به پویایی سیستم در اختیار تصمیم‌گیرنده باشد.

برای مثال، در [۸]، نویسندگان مسئله بهینه‌سازی AoI را با در نظر گرفتن اطلاعات وضعیت کانال[۱۲] بررسی می‌کنند. هدف اصلی آن‌ها طراحی یک سیاست زمان‌بندی است که به‌طور رقابتی متوسط دیرینگی اطلاعات را برای همه کاربران بهینه می‌کند. به این منظور آن‌ها یک سیاست زمان‌بندی حریصانه[۱۳] به نام MA-CSIT را ارائه می‌دهند که بر اساس CSI، کاربران را با توجه به دیرینگی و کیفیت کانال اولویت‌بندی می‌کند. آن‌ها نشان می‌دهند که این سیاست در حداقل کردن متوسط AoI و حداکثر کردن سودمندی سیستم بهینه است و عملکرد بهتری نسبت به سایر سیاست‌ها دارد. مدل سیستم آن‌ها شامل چندین کاربر است که در هر شکاف زمانی، یک بسته به‌روزرسانی جدید از منبع به ایستگاه مرکزی ارسال می‌شود و سپس فرستنده بر اساس MA-CSIT، بسته جدید را به یک کاربر منتقل می‌کند.

در کار [۹]، نویسندگان مسئله زمان‌بندی بهینه را در یک شبکه بی‌سیم با استفاده از یک RIS برای کاهش آسیب‌های ناشی از انتشار و افزایش کیفیت سیگنال بررسی می‌کنند. آن‌ها یک مدل شبکه با یک ایستگاه پایه و چندین جریان ترافیک را در نظر می‌گیرند که از یک RIS برای تغییر فاز سیگنال‌ها و هدایت آن‌ها به مقصد استفاده می‌کنند. هدف آن‌ها یافتن زمان‌بندی بهینه جریان‌های ترافیک و ماتریس تغییر فاز RIS است که مجموع AoI را به حداقل برساند. آن‌ها یک الگوریتم کارآمد مبتنی بر تغییر متغیرها با آزادسازی نیمه معین (SDR) را برای حل مسئله پیشنهاد می‌دهند که با استفاده از متغیرهای کمکی و SDR، مسئله را به یک مسئله بهینه‌سازی محدب تبدیل و با یک روش تکراری حل می‌کند. آن‌ها نشان می‌دهند که الگوریتم آن‌ها عملکرد بهتری نسبت به الگوریتم‌های ساده‌تر دارد و RIS می‌تواند عملکرد شبکه بی‌سیم را بهبود دهد. بااین‌حال، در این کار پیچیدگی و ابعاد فضای بهینه‌سازی بالاست و عدم قطعیت‌ها و محدودیت‌های شبکه در نظر گرفته نشده است.

اطلاعات مبتنی بر مدل (آماری): درصورتی‌که اطلاعات اولیه آماری (توزیع‌های احتمال) از پویایی سیستم مانند تغییرات کانال بی‌سیم و وقوع رویدادهای محیطی از پیش در اختیار تصمیم‌گیرنده باشد و

اجرای الگوریتم تنها نیازمند وقوع رویدادهای تصادفی (به‌عنوان ورودی سیاست از پیش محاسبه‌شده) باشد، از آن به‌عنوان اطلاعات مبتنی بر مدل یاد می‌شود.

مثلاً، در [۱۰]، نویسندگان مسئله زمان‌بندی بهینه را در یک شبکه حسگر بی‌سیم تغذیه‌شده[۱۴] بررسی می‌کنند. آن‌ها یک سیاست زمان‌بندی جدید به نام اولین تأخیر با آگاهی از مهلت[۱۵] را ارائه می‌دهند که بر اساس تأخیر و ضرب‌الاجل به‌روزرسانی‌ها را اولویت‌بندی می‌کند. آن‌ها نشان می‌دهند که این سیاست در حداقل کردن متوسط AoI و حداکثر کردن سودمندی سیستم بهینه است و عملکرد بهتری نسبت به سایر سیاست‌های زمان‌بندی دارد. مدل سیستم آن‌ها شامل چندین حسگر و محرک است که با یک ایستگاه مرکزی از طریق یک کانال بی‌سیم که در معرض تداخل و محوشدگی است ارتباط برقرار می‌کنند. همچنین از اندازه بافر محدود در ایستگاه مرکزی و ضرب‌الاجل سخت برای به‌روزرسانی و فرمان‌ها فرض می‌شود.

در مقاله [۱۱]، نویسندگان مسئله زمان‌بندی در یک شبکه IoT را به‌صورت انحصاری با ورودی‌های اولویت‌بندی شده و مهلت زمانی بررسی می‌کنند. آن‌ها یک مدل شبکه با دو کلاس از حسگرهای IoT با اولویت‌های بالا و پایین را در نظر می‌گیرند که داده‌های خود را به یک مرکز نظارت با یک سرور و یک بافر ارسال می‌کنند. آن‌ها فرض می‌کنند که بسته‌ها دارای ضرب‌الاجل‌های سخت هستند و بافر می‌تواند بسته‌های اولویت پایین را با بسته‌های اولویت بالا جایگزین کند. آن‌ها از AoI و نرخ اطلاعات[۱۶] به‌عنوان معیارهای تازگی و توان عملیاتی سیستم استفاده می‌کنند و یک معیار جدید به نام نسبت دیرینگی به نرخ اطلاعات[۱۷] را پیشنهاد می‌دهند که نسبت AoI به RoI را نشان می‌دهد. آن‌ها از یک مدل مارکوف برای استخراج عبارات فرم بسته با در نظر گرفتن مهلت‌های قطعی و تأثیر بسته‌های اولویت بالا و پایین استفاده می‌کنند. نتایج شبیه‌سازی نشان می‌دهد که ARRoI معیار مناسبی برای یافتن نرخ‌های انتقال مناسب است.

اطلاعات فارغ از مدل[۱۸]: در شرایطی که دانش آماری از شرایط تصادفی محیط عملیاتی در زمان طراحی در اختیار تصمیم‌گیرنده نباشد، ناگزیر از ارائه راهکارهای مبتنی بر یادگیری ماشین برای محاسبه سیاست بهینه هستیم. استفاده از تکنیک‌های یادگیری از اهمیت ویژه‌ای برخوردار است چراکه در بسیاری از سناریوهای کاربردی، امکان مدل‌سازی دقیق ساختار احتمالاتی سیستم میسر نیست. از این گذشته، جواب نظیر یک مدل خاص با تغییر شرایط، اعتبار خود را از دست می‌دهد. این فرض اطلاعاتی واقع‌بینانه است و به توزیع آماری شرایط مسئله وابسته نیست.

در مقاله [۱۲]، نویسندگان مسئله ایمنی وسایل نقلیه خودکار متصل[۱۹] را با در نظر گرفتن اشتراک‌گذاری داده‌های خام قابل‌اعتماد و کارآمد از چندین نوع حسگر بررسی می‌کنند. آن‌ها یک مدل شبکه با یک ایستگاه پایه[۲۰] و چندین CAV را در نظر می‌گیرند که از کانال mmWave یا کانال زیر ۶ گیگاهرتز برای انتقال داده‌های سنجش خام استفاده می‌کنند. آن‌ها فرض می‌کنند که بسته‌ها دارای ضرب‌الاجل‌های سخت هستند و ایستگاه پایه می‌تواند ساختار فریم را به‌صورت پویا بر اساس پروتکل رادیوی جدید 5G پیکربندی کند. هدف آن‌ها یافتن بهترین طرح تخصیص منابع برای حداقل کردن مجموع وزنی AoI و مصرف انرژی CAVها است. آن‌ها از یک بازی غیر همکارانه با استفاده از اطلاعات متقابل رادار برای مدل‌سازی مسئله استفاده می‌کنند و الگوریتم‌های یادگیری تقویتی تکرار ارزش و یادگیری Q را برای حل آن توسعه می‌دهند. آن‌ها نشان می‌دهند که الگوریتم‌های آن‌ها عملکرد بهتری نسبت به الگوریتم‌های ساده‌تر دارند.

در کار [۱۳]، نویسندگان مسئله زمان‌بندی بهینه را در یک WPSN با استفاده از یک RIS برای کاهش آسیب‌های ناشی از انتشار و افزایش کیفیت سیگنال بررسی می‌کنند. آن‌ها یک مدل شبکه با یک ایستگاه پایه چند آنتنی و چندین جریان ترافیک را در نظر می‌گیرند که از یک RIS برای تغییر فاز سیگنال‌ها و هدایت آن‌ها به مقصد استفاده می‌کنند. هدف آن‌ها حداقل کردن مجموع AoI است که نشان‌دهنده تازگی اطلاعات در مورد وضعیت شبکه است. بهینه‌سازی شامل برنامه‌ریزی انتقال گره‌ها و استراتژی‌های کنترل انتقال مانند بردار شکل‌دهی پرتو[۲۱] AP و ماتریس تغییر فاز RIS است. آن‌ها یک چارچوب یادگیری تقویتی عمیق سلسله مراتبی[۲۲] را برای حل مسئله پیشنهاد می‌دهند که شامل دو گام است. گام اول استفاده از یک عامل DRL حلقه بیرونی است که با در نظر گرفتن بازخورد مجموع AoI، تصمیم می‌گیرد کدام گره‌ها باید اطلاعات خود را در هر شکاف زمانی ارسال کنند. گام دوم استفاده از روش بهینه‌سازی حلقه داخلی است که با توجه به نوع جریان ترافیک، سطح توان و فاز را برای افزایش نرخ اطلاعات فراسو[۲۳] یا انرژی برداشت‌شده[۲۴] تطبیق می‌دهد. آن‌ها همچنین روش تقریبی ساده و کارآمدی را برای کاهش پیچیدگی محاسباتی بهینه‌سازی حلقه داخلی پیشنهاد می‌کنند. آن‌ها عملکرد چارچوب خود را با شبیه‌سازی‌های گسترده‌ای ازنظر مجموع AoI و عدالت نسبی در بین گره‌های مختلف، با رویکردهای رایج مقایسه می‌کنند و نشان می‌دهند که RIS می‌تواند عملکرد شبکه بی‌سیم را بهبود دهد.

در مقاله [۱۴]، نویسندگان مسئله زمان‌بندی بهینه را در یک WPSN با استفاده از یک RIS نصب‌شده بر روی پهپاد برای کاهش تأخیر انتقال و افزایش تازگی اطلاعات بررسی می‌کنند. آن‌ها یک مدل شبکه با یک ایستگاه پایه و چندین دستگاه اینترنت اشیا را در نظر می‌گیرند که از پهپاد به‌عنوان یک گره رله و RIS برای تغییر فاز سیگنال‌ها و هدایت آن‌ها به مقصد استفاده می‌کنند. هدف آن‌ها حداقل کردن مجموع AoI است که نشان‌دهنده دیرینگی اطلاعات در مورد وضعیت شبکه است. بهینه‌سازی شامل پیکربندی RIS، قرار دادن RIS بر پهپاد و زمان‌بندی انتقال حسگر است که با استفاده از روش‌های یادگیری تقویتی عمیق به دست می‌آید. آن‌ها دو روش یادگیری عمیق Q[۲۵] خارج از سیاست و بهینه‌سازی سیاست تقریبی[۲۶] را برای حل مسئله پیشنهاد می‌دهند که روش دوم در افزایش به‌موقع بودن اطلاعات ارسالی بسیار مؤثر است. آن‌ها عملکرد روش‌های خود را با شبیه‌سازی‌های گسترده‌ای ازنظر مجموع AoI و انصاف متناسب در بین دستگاه‌های مختلف، با رویکردهای استاندارد مقایسه می‌کنند. بااین‌حال، در این کار فضای اعمال به‌صورت گسسته در نظر گرفته‌شده است که جواب بهینه دقت کمتری دارد و تقریبی است. جواب دقیق‌تر و مناسب‌تری در صورتی قابل‌دستیابی است که فضای اعمال پیوسته در نظر گرفته شود.

۱-۲- انگیزه ارائه راهکار جدید و نوآوری‌ها

در پژوهش‌های پیشین از میان سه رویکرد شرح داده‌شده در مورد فرض اطلاعاتی، رویکرد فارغ از مدل کاربردی‌تر و واقع‌بینانه‌تر است. ما نیز در این مقاله، رویکردی فارغ از مدل ارائه خواهیم داد. به بیان مشخص‌تر، ابتدا یک فرمول‌بندی رسمی از مسئله حداقل‌سازی مجموع تخفیفی AoI همه کاربران ارائه می‌دهیم. فرمول‌بندی ارائه‌شده و راهکار پیشنهادی برای حل مسئله نسبت به کارهای پیشین از چند جهت دارای نوآوری است:

- اول اینکه، ما برای کاهش متوسط تخفیف یافته مبتنی بر AoI در یک شبکه بی‌سیم از فناوری RIS با پیوند mmWave و ارتباطات چند نقطه‌به‌نقطه استفاده کردیم که با استفاده از یادگیری ماشین می‌توان یک سیستم هوشمند و تطبیقی طراحی کرد که می‌تواند به‌صورت پویا پارامترهای انتقال و استراتژی‌های چندگانه را برای به حداقل رساندن تخفیفی AoI تحت شرایط مختلف شبکه انتخاب کند. به‌عنوان‌مثال، سیستم می‌تواند از RIS برای ایجاد مسیرهای انتشار اضافی و غلبه بر انسداد یا محو شدن سیگنال در ارتباطات mmWave که مستعد افت مسیر بالا و همچنین حساس به موانع هستند، استفاده کند. همچنین، به‌کارگیری ارتباطات چند نقطه‌به‌نقطه این قابلیت را فراهم می‌آورد که منابع متعدد به‌طور هم‌زمان، به‌روزرسانی‌های وضعیت خود را به یک مقصد واحد ارسال کنند که این امر موجب بهبود کارایی استفاده از طیف فرکانسی و اتصالات شبکه‌ای می‌گردد.
- دوم اینکه ما یک الگوریتم یادگیری تقویتی عمیق به نام DDPG برای حل مسئله بهینه‌سازی AoI طراحی کردیم. DDPG می‌تواند سیاست‌های بهینه برای تنظیم زمان‌بندی کاربران، توان ارسالی، فاز RIS و شکل پرتو در BS را بدون دانستن مدل سیستم یا اطلاعات وضعیت کانال یاد بگیرد. DDPG همچنین دارای ویژگی همگرایی سریع و راه‌حل بهینه است.
- سومین نوآوری، ارزیابی دقیق عملکرد روش پیشنهادی در سناریوهای مختلفی مانند تعداد کاربران، اندازه بسته‌ها، تعداد عناصر RIS و غیره می‌باشد که این ارزیابی با الگوریتم‌های معتبر رایج مقایسه شده است.

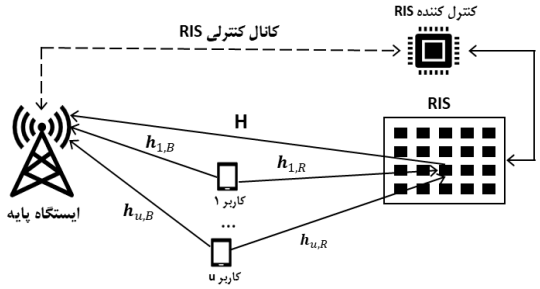
۱-۳- ساختار مقاله

ساختار ادامه مقاله به این شرح است: در بخش ۲، به معرفی مدل سیستم پرداخته می‌شود. در بخش ۳ فرمول‌بندی مسئله مبتنی بر فرآیند مارکوف را شرح می‌دهیم و همچنین در بخش ۴ به معرفی روش‌های پیشنهادی می‌پردازیم سپس به شبیه‌سازی و نتایج حاصل از آن در بخش ۵ اشاره می‌شود و در آخر به نتیجه‌گیری از پژوهش انجام‌شده خواهیم پرداخت.

۲- مدل سیستم و فرضیات

مطابق شکل ۱، سناریوی پیشنهادی در یک شبکه بی‌سیم مبتنی بر امواج میلی‌متری فراسو به کمک RIS موردبررسی قرار می‌گیرد که در آن U کاربر با تک آنتن وظیفه دارند که پویایی فیزیکی محیط خود را اندازه‌گیری و با استفاده از مکانیزم دسترسی چندگانه تقسیم زمان[۲۷] آن‌ها را در قالب یک بسته به‌روزرسانی با کمک یک RIS مجهز به M عنصر کم‌هزینه بازتاب‌دهنده و یک کنترل‌کننده به یک ایستگاه پایه با N آنتن ارسال کنند. فرض می‌شود که BS بر اساس معماری شکل‌دهی پرتوی آنالوگ[۲۸] طراحی‌شده است و نیازمند یک زنجیره فرکانس رادیویی است. معماری ABF تکنیک پردازش سیگنال است که برای انتقال یا دریافت سیگنال جهت‌دار استفاده می‌شود. این فناوری سبب دریافت سیگنال متمرکزتر در سمت گیرنده خواهد شد. هر آنتن با استفاده از یک انتقال‌دهنده فاز و یک تقویت‌کننده کم نویز[۲۹] به RFC متصل می‌شود؛ بنابراین BS مجهز به N انتقال‌دهنده فاز و N تقویت‌کننده کم نویز است. به‌منظور

کاهش پیچیدگی پیاده‌سازی سخت‌افزاری همه LNA ها یک ضریب تقویت‌کننده خواهند داشت، به‌طوری‌که برای بردار $\boldsymbol{w} \in \mathbb{C}^{N\times 1}$ شرط مدول ثابت[30] برقرار باشد. به عبارتی، $|[\boldsymbol{w}]_n| = \frac{1}{\sqrt{N}}$ به ازای هر $n \in N$. ضرایب بازتاب‌دهنده RIS توسط BS محاسبه‌شده و از طریق یک کانال اختصاصی به کنترل‌کننده RIS اطلاع داده می‌شود. تغییر فاز RIS در زمان واقعی توسط کنترل‌کننده پیکربندی می‌شود. ما فرض می‌کنیم که در هر شکاف زمانی اطلاعات وضعیت کانال به‌صورت کامل در دسترس BS است که فرض معمولی در مقالات نیز است [۱۸-۱۵] و ممکن است به‌صورت عملی با استفاده از روش‌های تخمین کانال قابل دستیابی باشد.



شکل ۱: مدل سیستم برای دیرینگی اطلاعات.

در سمت فرستنده، سیگنال ارسال‌شده توسط کاربر u-ام، $u \in \mathcal{U}$، از رابطه (۱) محاسبه می‌شود:

$$s_u = x_u\sqrt{p_u}, \forall u \in \mathcal{U} \qquad (۱)$$

به‌طوری‌که، p_u و x_u به ترتیب توان و سیگنال ارسالی کاربر u-ام خواهد بود و $\mathcal{U} \triangleq \{1, \dots, u\}$. با توجه به مکانیزم دسترسی به رسانه TDMA، زمان به شکاف‌های مشخصی به طول τ تقسیم می‌شود، که در هر شکاف زمانی، فقط یکی کاربر اجازه ارسال داده‌ها را خواهد داشت. بردار $Z = [z_u]_{u\in\mathcal{U}} \in \{0,1\}^u$ زمان‌بندی ارسال کاربران نشان خواهد داد به‌طوری‌که $\sum_{u=1}^{u} z_u = 1$ خواهد بود.

بنابراین با توجه به رابطه ۱، سیگنال دریافتی در BS از کاربر $u \in \mathcal{U}$ بعد از پردازش ABF از رابطه (۲) محاسبه می‌شود:

$$y = \boldsymbol{w}^H\left(\boldsymbol{h}_{u,B} + \widetilde{\boldsymbol{H}}_{\boldsymbol{u}}\boldsymbol{\theta}\right)s_u + \boldsymbol{n}, \forall \in \mathcal{U} \qquad (۲)$$

به‌طوری‌که، $\boldsymbol{n} \in \mathbb{C}^{N\times 1} \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_N)$ بردار نویز، $\boldsymbol{\Theta} \in \mathbb{C}^{M\times M} = diag(\boldsymbol{\theta})$ ماتریس انتقال فاز RIS است، که $\boldsymbol{\theta} = \left[e^{j\theta_1}, \dots, e^{j\theta_M}\right]^T$ و θ_m برابر m-امین انتقال فاز ضریب بازتاب $\forall m \in \mathcal{M} \triangleq \{1, \dots, M\}$ در نظر گرفته‌شده است و دامنه ضرایب بازتاب به مقدار یک لحاظ می‌گردد تا بازتاب سیگنال حداکثر شود. $\widetilde{\boldsymbol{H}}_u = \boldsymbol{H}diag(h_{u,R})$ بهره کانال زنجیره‌ای[31] است. بردارهای بهره کانال mmWave بین کاربر u-ام و BS، $\boldsymbol{h}_{u,B}$، بین کاربر u-ام و RIS، $\boldsymbol{h}_{u,R}$، و بین RIS و BS، $\boldsymbol{H}$ همگی مبتنی بر مدل Saleh-Valenzuela (SV) فرموله شده است [۲۰,۱۹].

۱-۱- مدل کانال بی‌سیم

به‌طور مشخص، لینک مستقیم $\boldsymbol{h}_{u,B} \in \mathbb{C}^{N\times 1}$ که مبتنی بر مدل کانال سه بعدی SV می باشد از رابطه (۳) محاسبه می‌گردد:

$$\boldsymbol{h}_{u,B} = \varrho_U\varrho_B\sqrt{\frac{N}{L_{u,B}}}\sum_{l=1}^{L_{u,B}}\alpha_{u,l}\mathrm{a}_B\left(N, \phi_{u,l}^{(AoA)}\right) \qquad (۳)$$

به‌طوری‌که، ϱ_U و ϱ_B به ترتیب، بهره المان‌های ارسال و دریافت هستند. $L_{u,B}$ تعداد مسیرهای کانال mmWave بین کاربر u-ام و BS است. $\alpha_{u,l}$ و $\phi_{u,l}^{(AoA)}$ به ترتیب، ضریب بهره مختلط و زاویه رسیدن[32] برای مسیر l-ام کاربر u-ام است. $\mathrm{a}_B\left(N, \phi_{u,l}^{(AoA)}\right)$، بردار آرایه هدایت کردن بین کاربر u-ام و BS است، به‌طوری‌که، $\mathrm{a}(N,\Delta) = \frac{1}{\sqrt{N}}\left[1, e^{j\pi\sin\Delta}, \dots, e^{j\pi(N-1)\sin\Delta}\right]^T$، که $\Delta \sim U[0,2\pi]$. کانال mmWave بین کاربر u-ام و RIS، $\boldsymbol{h}_{u,R} \in \mathbb{C}^{M\times 1}$ از رابطه (۴) محاسبه می‌شود:

$$\boldsymbol{h}_{u,R} = \varrho_U\varrho_R\sqrt{\frac{M}{L_{u,R}}}\sum_{l=0}^{L_{u,R}-1}\beta_{u,l}\mathrm{a}_R\left(M, \psi_{u,l}^{(AoA)}\right) \qquad (۴)$$

به‌طوری‌که، ϱ_R بهره دریافت در RIS و $L_{u,R}$ تعداد مسیرهای mmWave بین کاربر u-ام و RIS است. هم‌چنین، $\beta_{u,l}$ و $\psi_{u,l}^{(AoA)}$ به ترتیب، ضریب بهره مختلط و زاویه رسیدن برای مسیر l-ام کاربر u-ام است. $\mathrm{a}_R\left(M, \psi_{u,l}^{(AoA)}\right)$ آرایه هدایت کردن بین کاربر u-ام و RIS است.

کانال mmWave بین RIS و BS، $\boldsymbol{H} \in \mathbb{C}^{N\times M}$ از رابطه (۵) به دست می‌آید:

$$\boldsymbol{H} = \varrho_R\varrho_B\sqrt{\frac{MN}{L_{R,B}}}\times \sum_{l=0}^{L_{R,B}-1}\beta_l\mathrm{a}_B\left(N, \psi_l^{(AoA)}\right)\mathrm{a}_R^H\left(M, \psi_l^{(AoD)}\right) \qquad (۵)$$

به‌طوری‌که، $L_{R,B}$ تعداد مسیرهای mmWave بین RIS و BS است. هم‌چنین، β_l و $\psi_l^{(AoA)}$ و $\phi_l^{(AoD)}$ به ترتیب، ضریب بهره مختلط و زاویه رسیدن و زاویه خروج برای مسیر l-ام کاربر u-ام است. $\mathrm{a}_B\left(N, \psi_l^{(AoA)}\right)$ و $\mathrm{a}_R\left(M, \psi_{k,l}^{(AoD)}\right)$ آرایه هدایت کردن بین RIS و BS است. در روابط (۳)-(۵)، $l = 0$ نشان‌دهنده مسیر LoS و $l > 0$ مسیرهای NLoS را نشان می‌دهد.

با توجه به رابطه (۲) نسبت سیگنال به‌علاوه نویز[۳۳] کاربر u-ام از رابطه (۶) محاسبه می‌شود:

$$SNR_u = \frac{p_u \left|\boldsymbol{w}^H\left(\boldsymbol{h}_{u,B} + \widetilde{\boldsymbol{H}}_{\boldsymbol{u}}\boldsymbol{\theta}\right)\right|^2}{\sigma^2} \qquad (۶)$$

در این رابطه هریک از کاربران با توان p_u بسته خود را ارسال خواهند کرد. توان ارسال هریک از کاربرها باید کمتر از حد مجاز P_{MAX} انتخاب گردد. بنابراین نرخ قابل‌دستیابی برای ارسال داده اندازه‌گیری شده کاربر u-ام از رابطه (۷) به دست می‌آید.

$$\begin{aligned} R_u &= W\log_2(1+SNR_u) \\ &= W\log_2\left(1+\frac{p_u\left|\boldsymbol{w}^H\left(\boldsymbol{h}_{u,B}+\widetilde{\boldsymbol{H}}_{\boldsymbol{u}}\boldsymbol{\theta}\right)\right|^2}{\sigma^2}\right) \end{aligned} \qquad (۷)$$

در این رابطه W پهنای باند کانال را مشخص می‌کند.

۲-۲- مدل دیرینگی اطلاعات

با در نظر گرفتن تعریف رسمی ریاضی دیرینگی اطلاعات $\Delta_u(t) = t - u_u(t)$ که زمان گذشته از تولید آخرین بسته موفق دریافت شده در مقصد از هریک از کاربرها تعریف می‌شود. در این مسئله فرض شده است که در هر شکاف زمانی، کاربران پویایی فیزیکی محیط اطراف خود را اندازه گیری کرده و قصد دارند اندازه‌گیری خود را در قالب بسته‌ای به ایستگاه پایه ارسال کنند. درصورتی‌که بسته کاربر u-ام به‌اندازه f_u در طول یک شکاف زمانی به BS نرسد دیرینگی اطلاعات یک واحد افزایش می‌یابد و بسته دور ریخته می‌شود، در غیر این صورت به ۱ واحد زمانی کاهش می‌یابد. به عبارتی دیرینگی اطلاعات برای کاربر u بر اساس رابطه (۸) به‌روزرسانی می‌شود.

$$\Delta_u(t+1) = \begin{cases} \Delta_u(t)+1, & R_u.\tau < f_u \\ 1, & R_u.\tau \geq f_u \end{cases} \qquad (۸)$$

۲-۳- فرمول‌بندی مسئله

در این مقاله هدف مسئله ما حداقل‌سازی متوسط هزینه تجمعی تخفیفی مجموع دیرینگی اطلاعات همه کاربران با ضریب تخفیف $\gamma \in [0,1]$ است. به‌منظور بهینه‌سازی هدف ما زمان‌بندی ارسال کاربرها، توان ارسالی آن‌ها، انتقال فاز در RIS و شکل‌دهی پرتوی دریافتی در BS را با شرایط عنوان‌شده در رابطه (۹) فرموله می‌کنیم.

$$\begin{aligned} &\max_{\boldsymbol{Z},\boldsymbol{p},\boldsymbol{w},\boldsymbol{\Theta}} \bar{\Delta} = \lim_{T\to\infty}\left[\frac{1}{T}\sum_{t=1}^{T}\gamma^{t-1}\left(-1\times\sum_{u=1}^{U}\Delta_u\right)\right] \\ &s.t. \quad |[\boldsymbol{w}]_n| = \frac{1}{\sqrt{N}} \\ &R_u.\tau \geq f_u, \forall u \in \mathcal{U} \\ &\boldsymbol{\theta} = \left[e^{j\theta_m}\right]_{m\in\{1,\dots,M\}}, \theta_m \in [0,2\pi) \\ &\sum_{u=1}^{U} z_u = 1 \\ &0 \leq p_u \leq P_{MAX} \end{aligned} \qquad (۹)$$

بر اساس شرایط قید شده در هر شکاف زمانی فقط یک کاربر اجازه ارسال داده روی کانال دارد. اندازه بسته ارسالی کاربر u-ام در شکاف زمانی t نباید بیشتر از نرخ قابل دستیابی کاربر باشد، به این معنا که بسته باید در یک شکاف زمانی به مقصد برسد. همچنین توان ارسالی هر کاربر نباید از حد مجاز P_{MAX} عبور کند. بردارهای $\boldsymbol{w}$ و $\boldsymbol{\theta}$ نیز که به ترتیب نشان‌دهنده بردار شکل‌دهی پرتوی دریافتی در BS و ضرایب بازتاب‌دهنده RIS می‌باشند، باید دارای المان‌هایی به‌صورت اعداد مختلط در بازه $[0,2\pi]$ تعریف شوند.

۳- فرمول‌بندی مسئله با استفاده از فرآیند مارکوف تخفیفی

در مسئله موردبررسی، با توجه به تصادفی بودن ماهیت محیط می‌خواهیم ضمن تأمین قیود تعریف‌شده در رابطه (۹)، مجموع تخفیفی دیرینگی اطلاعات همه کاربران را حداقل کنیم. ازاین‌رو مسئله در قالب MDP به‌صورت چندتایی $(S, A, \mathbb{P}, R)$ مدل می‌شود که به ترتیب از چپ به راست شامل: وضعیت، اقدامات امکان‌پذیر، تابع احتمال گذار و پاداش است.

فضای وضعیت سیستم یک فضای پیوسته به‌صورت ضرب دکارتی AoI همه کاربران است که به‌صورت یک بردار چندتایی در نظر گرفته می‌شود. در هر شکاف زمانی، واحد تصمیم‌گیرنده بردار وضعیت را به‌عنوان ورودی دریافت و اقدامات امکان‌پذیر در وضعیت جاری را شناسایی می‌کند. فضای اقدامات امکان‌پذیر در هر شکاف زمانی، یک فضای پیوسته و متناهی به‌صورت $A = (Z, \boldsymbol{p}, \boldsymbol{w}, \boldsymbol{\theta})$ است که در آن بردار Z زمان‌بندی ارسال کاربران، $\boldsymbol{p}$ توان ارسالی آن‌ها، $\boldsymbol{w}$ به‌عنوان بردار شکل‌دهی پرتو دریافتی توسط BS و $\boldsymbol{\theta}$ ماتریس ضرایب بازتاب‌دهنده RIS است. هر یک از مؤلفه‌های اقدام که در شرط صدق کند مجموعه A را تشکیل می‌دهند.

در وضعیت s_t با انجام اقدام، وضعیت سیستم با احتمال $\mathbb{P}(s_{t+1}|s_t, a_t)$ به وضعیت جدید گذار می‌کند که در رابطه (۱۰) ماتریس گذار سیستم بیان شده است:

$$\mathbb{P}(s_{t+1}|s_t, a_t) = \mathbb{P}(\Delta_{t+1}|\Delta_t, a_t) \qquad (۱۰)$$

پاداش آنی ناشی از اجرای اقدام a_t در وضعیت s_t، را با نماد r_t به‌صورت رابطه (۱۱) نشان می‌دهیم.

$$r_t(s_t, a_t) = -1 \times \sum_{u=1}^{U} \Delta_u \qquad (۱۱)$$

همچنین به‌منظور بررسی هدف مسئله سیاست کنترلی $\pi: S \rightarrow \Delta(A)$ جهت نگاشت یک وضعیت مشخص $s_t \in S$ به توزیع احتمال بر روی مجموعه اقدامات کنترلی امکان‌پذیر، $\Delta(A)$ تعریف‌شده است.

بازده از یک وضعیت به‌عنوان مجموع پاداش آنی با تخفیف به‌صورت $R_t = \sum_{i=t}^{T} \gamma^{(i-t)} r_i(s_i, a_i)$ که ضرب هر پاداش در ضریب تخفیف $\gamma \in [0, 1]$ انجام می‌شود. به این معنا که اگر عامل از یک وضعیت خاص شروع کند و از یک سیاست خاص پیروی کند، چقدر می‌تواند در آینده پاداش بگیرد. بازده بستگی به اعمال انتخاب‌شده، سیاست π و احتمال تصادفی بودن دارد.

برای حل مسئله، در روش پیشنهادی، گره ایستگاه پایه به‌عنوان واحد تصمیم‌گیرنده باید سیاست بهینه (که با نماد μ^* نشان داده می‌شود) را برای دستیابی به جواب بهینه که هدف مسئله است، یاد بگیرد به‌طوری‌که برای همه $s \in S$ در رابطه (۱۲) برقرار باشد.

$$\mu^* = \underset{a \epsilon A(s)}{\operatorname{argmax}} Q^{\mu}(s, a) \qquad (۱۲)$$

۴- روش پیشنهادی مبتنی بر یادگیری تقویتی

محاسبه سیاست بهینه با روش‌های مبتنی بر مدل در شرایط تغییرپذیر و نامعلوم سخت و ناممکن است. همچنین، در بسیاری از سناریوهای کاربردی، دقت و کامل بودن اطلاعات آماری و احتمالاتی سیستم قابل‌دسترس نیست. با توجه به ماهیت تصادفی مسئله و عدم در دسترس بودن اطلاعات کامل یا آماری از سیستم ما از یک الگوریتم فارغ از مدل مبتنی بر یادگیری تقویتی عمیق بنام گرادیان سیاست قطعی عمیق[۳۴] برای یافتن سیاست بهینه برای فضاهای اقدام پیوسته استفاده می‌کنیم.

۴-۱- روش یادگیری DDPG

با توجه به فرمول‌بندی مبتنی بر MDP که در بخش ۳ بیان شد، ما می‌توانیم از تکنیک‌های یادگیری تقویتی برای یادگیری سیاست بهینه استفاده کنیم. یادگیری تقویتی عمیق یک روش قدرتمند برای حل مسائل کنترل پیوسته است که نیاز به یافتن سیاست‌های بهینه برای فضاهای اقدام بزرگ و پیچیده دارند. این روش می‌تواند با استفاده از تجربه و بازخورد، سیاست‌های هوشمند و تطبیقی را برای هر محیط یاد بگیرد. به‌طور خاص، ما از الگوریتم DDPG به‌عنوان مبنای یادگیری فارغ از مدل با فضای اقدامات پیوسته استفاده می‌کنیم.

DDPG بر ایده استفاده از دو شبکه عصبی، یکی به‌عنوان نقاد و دیگری به‌عنوان عملگر، برای تقریب تابع ارزش-اقدام بهینه و تقریب سیاست بهینه تکیه دارد. تابع ارزش-اقدام بهینه، بازده مورد انتظار را پس از اجرای اقدام a_t در وضعیت s_t و پیروی از سیاست μ محاسبه می‌کند و سیاست بهینه، بهترین اقدام را که بازده مورد انتظار را به حداکثر می‌رساند، به دست می‌آورد.

یادگیری تابع ارزش-اقدام بهینه در DDPG بر اساس معادله بلمن است که با استفاده از آن، می‌توانیم بهترین اقدام را در هر وضعیتی پیدا کنیم [۲۱].

بر اساس MDP فرموله شده، ما استفاده از رویکرد DDPG خود را برای حل مشکل مشخص شده در (۹) معرفی می کنیم.

الگوریتم DDPG دارای دو مرحله اصلی است: مقداردهی اولیه و یادگیری. در مرحله مقداردهی اولیه، شبکه‌های عصبی اولیه برای نقاد $Q(s, a|\theta^Q)$ و عملگر $Q(s, a|\theta^{\mu})$ ایجاد می‌شوند و پارامترهای θ^Q و θ^{μ} به‌عنوان پارامترهای شبکه نقاد-عملگر به‌صورت تصادفی مقداردهی اولیه می‌شوند. همچنین، به‌منظور افزایش پایداری در طول فرآیند یادگیری، دو کپی از شبکه‌های اولیه به نام شبکه‌های هدف نیز ایجاد می‌شود و پارامترهای آن‌ها با پارامترهای شبکه‌های اولیه تنظیم می‌شوند. به‌این‌ترتیب، پارامترهای $\theta^{Q'}$ و $\theta^{\mu'}$ با مقادیر θ^Q و θ^{μ} هماهنگ می‌شوند.

در طول مرحله یادگیری، DDPG در شکاف‌های زمانی T تکرار می‌شود و یک فرآیند تصادفی را برای کشف اقدام اولیه آغاز می‌کند. در هر شکاف زمانی t، یک اقدام a_t بر اساس سیاست تعریف‌شده توسط $Q^{\mu}(s_t|\theta^{\mu})$ انتخاب می‌شود. سپس عملکرد این اقدام با تعداد بسته‌های موفق دریافت شده ارزیابی و طبق عملکرد، پاداش آنی r_t محاسبه می‌شوند. پس از به‌روزرسانی پاداش آنی r_t، حالت بعدی s_{t+1} مشاهده و سپس یک چندتایی انتقالی (s_t, a_t, r_t, s_{t+1}) را تشکیل می‌شود. این چندتایی‌ها در یک بافر تجربه ذخیره می‌شوند و DDPG از آن به‌عنوان تکنیکی برای تسهیل یادگیری آینده استفاده می‌کند [۲۲-۲۱]. علاوه بر این‌ها، DDPG برای یادگیری سیاست از صعود گرادیان استفاده می‌کند که یک تکنیک بهینه‌سازی است که پارامترهای شبکه عملگر را در جهت یافتن سریع سیاست بهینه به‌روز می‌کند. به این صورت که، یک دسته کوچک به‌اندازه B چندتایی از بافر تجربه نمونه‌برداری می‌کند، سپس طبق رابطه (۱۳) گرادیان تابع Q^{μ} را با توجه به اقدام

اتخاذشده محاسبه و پارامتر θ^{μ} را به‌روز می‌کند.

$$\frac{1}{B}\sum_{i}\{\nabla_a Q(s,a|\theta^Q)|_{s=s_t,a=Q^{\mu}(s_t)}\nabla_{\theta^{\mu}}Q^{\mu}(s|\theta^{\mu})|_{s_t}\}, \quad (۱۳)$$

پارامترهای شبکه نقاد θ^Q نیز با به حداقل رساندن میانگین مربعات خطای تابع اتلاف به‌صورت رابطه (۱۴) به‌روز می‌شوند:

$$\frac{1}{B}\sum_{t}\left(v_t - Q(s_t,a_t|\theta^Q)\right)^2. \quad (۱۴)$$

درنهایت، پارامترهای شبکه‌های هدف، $\theta^{Q'}$ و $\theta^{\mu'}$، با استفاده از فرا پارامتر τ به‌روز می‌شوند.

با توجه به توضیحات ارائه‌شده، شبه کد روش پیشنهادی در الگوریتم ۱ نشان داده‌شده است. شکل ۲، تعامل بین عامل تصمیم‌گیری با محیط شبکه بی‌سیم را نشان می‌دهد.

۴-۲- پیچیدگی محاسباتی

در ادامه، به بررسی پیچیدگی زمانی الگوریتم پرداخته خواهد شد. برای شبکه عصبی کاملا متصل با x_i نورون به ازای i-امین لایه ($i \in \{2,\dots,I\}$)، پیچیدگی محاسباتی از رابطه (۱۵) محاسبه می‌شود:

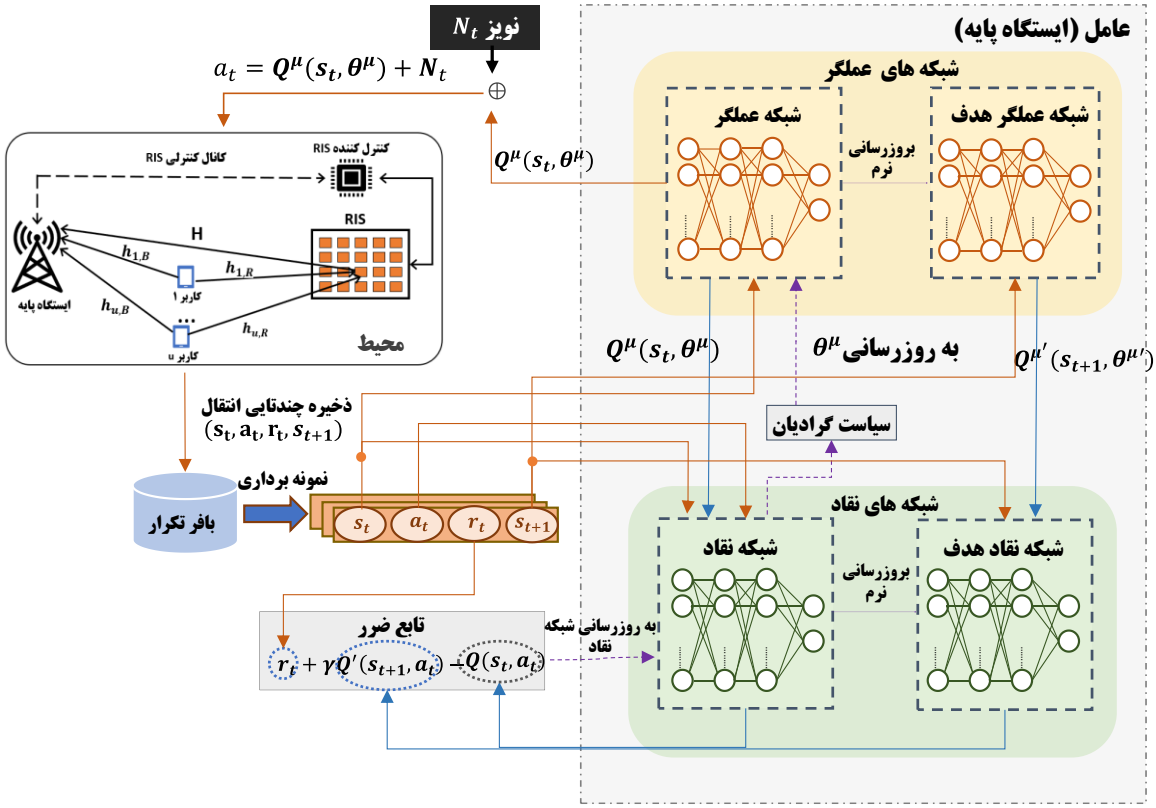
$$\mathcal{O}\left(\sum_{i=2}^{I-1}(x_{i-1}x_i + x_i x_{i+1})\right) \quad (۱۵)$$

در ادامه با در نظر گرفتن اندازه دسته کوچک نمونه‌ها، B، تعداد اپیزودها، E و طول یک اپیزود، T پیچیدگی زمانی الگوریتم DDPG از رابطه (۱۶) محاسبه می‌گردد. در این رابطه x_i^a و I^a تعداد نورون های لایه i ام شبکه عملگر و تعداد لایه های آن را نشان می‌دهد. به طور مشابه، نمادهای x_i^c و I^c برای شبکه نقاد در نظر گرفته شده است.

$$\mathcal{O}\left(B \times E \times T \times \left[\left(\sum_{i=2}^{I^a-1}(x_{i-1}^a x_i^a + x_i^a x_{i+1}^a)\right) + \left(\sum_{i=2}^{I^c-1}(x_{i-1}^c x_i^c + x_i^c x_{i+1}^c)\right)\right]\right) \quad (۱۶)$$

۵- پیاده‌سازی و نتایج شبیه‌سازی

در این بخش، سناریوی تشریح شده در بخش ۲ را شبیه‌سازی می‌کنیم. ابتدا به معرفی پارامترها و مقادیر آن‌ها می‌پردازیم و سپس ملاک‌های موردنظر برای ارزیابی روش پیشنهادی را شرح خواهیم داد و در آخر نمودارها و نتایج شبیه‌سازی را ارائه خواهیم کرد.



شکل ۲: نمودار جریان DDPG.

الگوریتم ۱. شبه کد روش یادگیری DDPG

// Initial Setup:
Set up the critic network $Q(s,a|\theta^Q)$.
Set up the actor network $Q^\mu(s|\theta^\mu)$.
Initialize the target critic network Q' with $\theta^{Q'} \leftarrow \theta^Q$.
Initialize the target actor network $Q^{\mu'}$ with $\theta^{\mu'} \leftarrow \theta^\mu$.
Set up the replay buffer, R, to store experiences.
// Learning:
for episode=1, E **do**
  Set up a randomized process N for exploration.
  Perceive the initial state s_1.
  Assign $r_a \leftarrow 0$.
  for t=1, T **do**
    Choose an action $a_t = Q^\mu(s_t|\theta^\mu) + N_t$.
    Execute the action a_t and receive reward r_t.
    Collect the reward $r_a \leftarrow r_a + r_t$.
    Take note of the new state s_{t+1}.
    Record (s_t, a_t, r_t, s_{t+1}) in R.
    Extract a minibatch of B-sized transitions from R.
    Set $v_t \leftarrow r_t + \gamma Q\left(s_{t+1}, Q^{\mu'}\left(s_{t+1}|\theta^{\mu'}\right)|\theta^{Q'}\right)$.
    Update θ^Q by minimizing the loss:
      $\frac{1}{B}\sum_t \left(v_t - \mathrm{Q}(s_t, a_t|\theta^Q)\right)^2$.
    Update θ^μ using the policy gradient:
      $\frac{1}{B}\sum_i \{\nabla_a Q(s,a|\theta^Q)|_{s=s_t, a=Q^\mu(s_t)} \nabla_{\theta^\mu} Q^\mu(s|\theta^\mu)|_{s_t}$
    Adjust the parameters of the objective networks:
      $\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$.
      $\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$.
  end for
  Assign the value of $r_{avg}[episode] = r_a$.
  if episode $\geq$ 50 **then**
    Provide the mean of $r_{avg}[episode]$.
  end if
end for

۵-۱- بستر آزمایش و تنظیم مقادیر شبیه‌سازی

در این پژوهش عملکرد روش پیشنهادی را از نظر ویژگی همگرایی و تغییرات پارامترهای مختلف شبکه بررسی می‌کنیم و نتایج اصل را با نتایج الگوریتم‌های زمان‌بندی تصادفی و زمان‌بندی نوبت چرخشی مورد مقایسه قرار می‌دهیم. ما همچنین تأثیر بهره بازتاب نسبی RIS، تأثیر وضعیت اطلاعات کانال و تأثیر حذف RIS بر متوسط تخفیف یافته دیرینگی اطلاعات را ارزیابی می‌کنیم.

فرا پارامترها و پارامترهای شبیه‌سازی برای هر سه الگوریتم طبق جدول ۱ و ۲ تنظیم و مقداردهی شده‌اند. در این شبیه‌سازی با توجه به [۱۹] و [۲۸]، بهره کانال طبق رابطه (۱۷)، از یک توزیع گاوسی مختلط پیروی می‌کند که در آن PL(d) به معنای از دست دادن مسیر (برحسب dB) در فاصله d (برحسب متر) می‌باشد.

$$\rho \sim \mathbb{CN}\left(0, 10^{-0.1PL(d)}\right) \qquad (۱۷)$$

طبق رابطه (۱۸) و $\mu \sim \mathcal{N}\left(0, \sigma_\mu^2\right)$ ضریب بهره کانال mmWave در دو حالت به دست می‌آید. در حالت LoS مقادیر پارامترهای $x_a =$ 61.4 و $x_b = 2$ دسی‌بل هستند و در حالت NLoS مقادیر $x_a =$ 72 و $x_b = 2.92$ برحسب dB حاصل می‌شود. طبق مقادیر به‌دست‌آمده، مسیر NLoS نسبت به مسیر LoS دارای تلفات مسیر بیشتر است که در این حالت بهره کانال واریانس کمتر و میانگین کمتری نسبت به بهره کانال LoS دارد.

$$PL(d) = X_a + 10X_b \log_{10}(d) + \mu\ [dB] \qquad (۱۸)$$

۵-۲- الگوریتم‌های مقایسه‌ای

در این مقاله، روش پیشنهادی خود را با دو الگوریتم زمان‌بندی پایه مقایسه و ارزیابی می‌کنیم. همچنین عملکرد الگوریتم پیشنهادی را در حالتی که شبکه RIS وجود نداشته باشد نیز بررسی می‌کنیم. این الگوریتم‌ها عبارت‌اند از:

- الگوریتم زمان‌بندی تصادفی (Random Scheduling): در این الگوریتم، زمان‌بند به‌صورت تصادفی یک کاربر را از بین کاربران انتخاب می‌کند و به آن اجازه ارسال بسته را می‌دهد.
- الگوریتم نوبت چرخشی (Round Robin Scheduling): در این الگوریتم، زمان‌بند به‌صورت نوبت چرخشی کاربران را انتخاب می‌کند. به این صورت که کاربران به ترتیب در هر شکاف زمانی دسترسی به کانال را خواهند داشت.

همچنین، به‌منظور ارزیابی تأثیر وجود RIS در شبکه، کارایی الگوریتم پایه DDPG در صورت عدم حضور RIS نیز مقایسه می‌شود. این مقایسه بینش‌های مفیدی را درباره کارایی الگوریتم بدون مؤلفه شکل‌دهی پرتو غیرفعال فراهم میاورد و چشم‌اندازی روشن از کارایی آن در شرایط عملیاتی مختلف ارائه می‌دهد.

جدول ۱: مقادیر فرا پارامترها در الگوریتم DDPG.

پارامتر	مقادیر	توضیحات
Optimizer	ADAM	بهینه‌ساز
Number of neural network layers	۲ لایه	تعداد لایه‌های شبکه عصبی
Activation Function	ReLU	تابع فعال‌ساز
Number of nerurons	(۴۰۰ و ۳۰۰) نورون	تعداد نورون‌ها در هر لایه شبکه عصبی (لایه اول، لایه دوم)
Mini-Batch Size	۶۴ عدد	اندازه دسته کوچک بافر تکرار
γ	۰.۹۵	ضریب تخفیف
Learning rtae	۰.۰۰۱	نرخ یادگیری
τ	۰.۰۰۵	ضریب به‌روزرسانی نرم

جدول ۲: پارامترهای شبیه‌سازی و مقادیر.

پارامتر	مقادیر	توضیحات
Iteration Number	هشتاد هزار تکرار	تعداد تکرار الگوریتم
τ	۰.۰۰۱ میلی‌ثانیه	طول شکاف زمانی
M	۲۰ عدد [23]	تعداد عناصر RIS
N	۱۶ عدد	تعداد آنتن‌های ایستگاه پایه
f_u	۵۰۰۰ بایت [23]	حجم بسته‌های داده
W	۲۰۰ گیگاهرتز [24]	پهنای باند کانال
U	۴ کاربر	تعداد کاربران
P_{MAX}	۰.۲ وات [23]	حداکثر توان ارسالی
$\Delta_{MAX}\ (AoI)$	۱۰۰	حداکثر AoI کاربران
$\beta_{u,l}$	$\mathcal{CN}(0, 10^{-0.1PL(d)})$ [28] , [19]	ضریب بهره مختلط مسیر l-ام بین کاربر و RIS
$\alpha_{u,l}$	$\mathcal{CN}(0, 10^{-0.1PL(d)})$ [28] , [19]	ضریب بهره مختلط مسیر l-ام بین کاربر و ایستگاه پایه
β_l	$\mathcal{CN}(0, 10^{-0.1PL(d)})$ [28] , [19]	ضریب بهره مختلط مسیر l-ام بین ایستگاه پایه و RIS
ρ_B	$\mathcal{CN}(0, 10^{-0.1PL(d)})$ [28] , [19]	بهره دریافت در ایستگاه پایه
ρ_u	$\mathcal{CN}(0, 10^{-0.1PL(d)})$ [28] , [19]	بهره ارسالی کاربر
ρ_{RIS}	$10\sqrt{\rho_u * \rho_B}$ [26]	بهره دریافت در RIS
σ^2	$-174 - 30 + (10\log_{10} W)$	σ^2
$\Delta(AoA)$	$\mathcal{U}(0, 2\pi)$	زاویه رسیدن برای مسیر l-ام کاربر
$\Delta(AoD)$	$\mathcal{U}(0, 2\pi)$	زاویه خروج برای مسیر l-ام کاربر
$\Delta(u, B)$	$\mathcal{U}(0, 2\pi)$	زاویه رسیدن برای مسیر l-ام کاربر
$\Delta(u, R)$	$\mathcal{U}(0, 2\pi)$	زاویه خروج برای مسیر l-ام کاربر
$L_{u,R}$	۳ مسیر	تعداد مسیرهای کانال mmWave بین کاربر و RIS
$L_{u,B}$	۳ مسیر	تعداد مسیرهای کانال mmWave بین کاربر و ایستگاه پایه
$L_{B,R}$	۳ مسیر	تعداد مسیرهای کانال mmWave بین ایستگاه پایه و RIS
(X_{RIS}, Y_{RIS})	$(0, 100)$	موقعیت قرارگیری RIS
(X_{BS}, Y_{BS})	$(0, 0)$	موقعیت قرارگیری ایستگاه پایه

۵-۳- ملاک‌های ارزیابی

برای ارزیابی الگوریتم روش پیشنهادی DDPG و مقایسه با الگوریتم‌های دیگر، متوسط تخفیف یافته را ملاک قرار می‌دهیم. درواقع الگوریتمی عملکرد بهتری خواهد داشت که در بلندمدت متوسط تخفیف یافته کمتری به دست آورد.

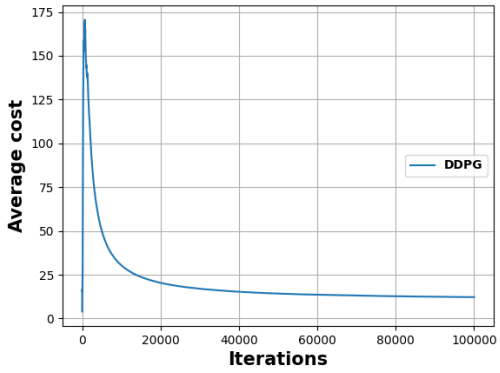
۵-۴- نتایج شبیه‌سازی

ابتدا ویژگی همگرایی الگوریتم پیشنهادی DDPG را بررسی می‌کنیم. همان‌طور که در شکل ۳، مشاهده می‌شود متوسط دیرینگی اطلاعات به‌عنوان هزینه بعد از تقریباً ۸۰ هزار تکرار به مقدار بهینه همگرا می‌شود. همگرایی به این معنی است که روش موردنظر توانسته به سیاست بهینه دست یابد.

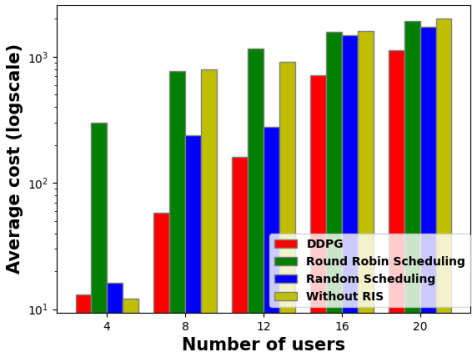
در این بخش پارامترهای شبیه‌سازی تغییر داده‌شده و تأثیر آن‌ها بر روی مقدار متوسط هزینه (دیرینگی اطلاعات) در هر سه الگوریتم پیشنهادی DDPG، الگوریتم زمان‌بندی تصادفی و الگوریتم نوبت چرخشی و همچنین در حالت بدون RIS در شبکه، مقایسه شده‌اند. آزمایش‌ها برای هر نقطه ۱۰ بار با تعداد تکرار ۸۰ هزار بار انجام‌شده است.

برای بررسی تأثیر تعداد کاربران بر متوسط هزینه، مطابق با مقادیر موجود جدول ۲، برای تعداد کاربران متفاوت [۲, ۴, ۸, ۱۶, ۲۰] شبیه‌سازی در هر سه الگوریتم اجرا شده است.

همان‌طور که شکل ۴، نشان می‌دهد با افزایش تعداد کاربران، مجموع متوسط هزینه به‌طور قابل‌توجهی افزایش می‌یابد. این افزایش ناشی از کاهش فرصت‌های ارسال برای هر کاربر است. به‌این‌ترتیب، با افزایش تعداد کاربران، نوبت تخصیص‌یافته به هر کاربر کاهش می‌یابد و محدودیت در تعداد بسته‌های ارسالی، منجر به افزایش میزان دیرینگی اطلاعات می‌شود. علاوه بر این، نتایج بیانگر آن است که استفاده از RIS به دلیل انعکاس سیگنال‌ها، به بهینه‌سازی پوشش سیگنال و کاهش تداخل کمک کرده و درنهایت، به بهبود کیفیت کانال و کاهش متوسط هزینه منجر می‌شود.



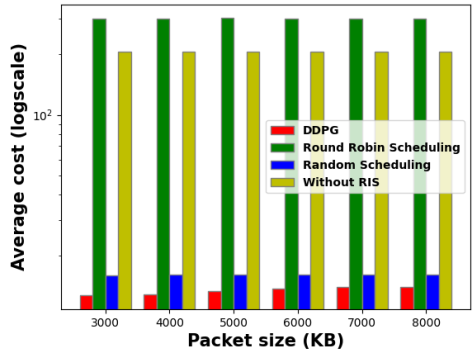
شکل ۳: همگرایی در روش پیشنهادی DDPG.



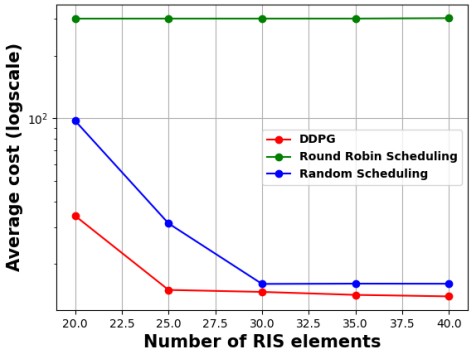
شکل ۴: تأثیر افزایش تعداد کاربران در الگوریتم‌ها.

شکل ۵، تأثیر اندازه بسته بر متوسط هزینه را با استفاده از مقادیر متفاوت [۳۰۰۰, ۴۰۰۰, ۵۰۰۰, ۶۰۰۰, ۷۰۰۰, ۸۰۰۰] کیلوبایت [۲۳] برای هر سه الگوریتم مختلف نشان می‌دهد. مشاهده می‌کنیم که با افزایش اندازه بسته، متوسط هزینه نیز به نسبت افزایش می‌یابد، زیرا بسته‌های بزرگ‌تر زمان انتقال طولانی‌تری نیاز دارند، که این امر به افزایش تأخیر در ارسال بسته‌ها و کاهش احتمال ارسال آن‌ها در شکاف زمانی معین منجر می‌شود. علاوه بر این، مشاهده می‌کنیم که حضور RIS در شبکه به کاهش متوسط هزینه کمک می‌کند.

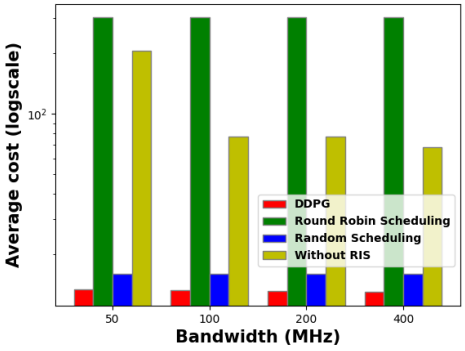
در شکل ۶، تأثیر تعداد عناصر RIS با مقادیر متفاوت [۲۰, ۲۵, ۳۰, ۳۵, ۴۰] بر متوسط هزینه برای هر سه الگوریتم مورد بررسی قرارگرفته است [۲۳]. مطابق این شکل، با افزایش تعداد عناصر مقدار متوسط کاهش می‌یابد؛ زیرا افزایش عناصر RIS می‌تواند پرتو متمرکزتری ایجاد کند و به بهره بالاتر دست یابد که همین امر می‌تواند باعث کاهش متوسط هزینه شود.



شکل ۵: تأثیر افزایش اندازه بسته در الگوریتم‌ها.



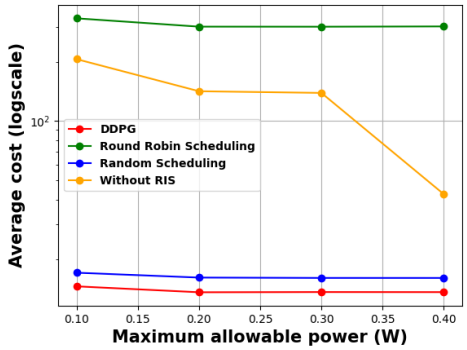
شکل ۶: تأثیر تعداد عناصر RIS در الگوریتم‌ها.



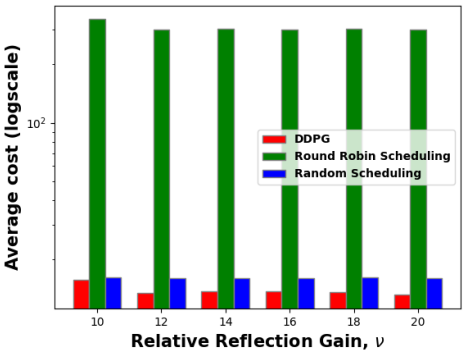
شکل ۷: تأثیر افزایش پهنای باند در الگوریتم‌ها.

شکل ۷، تأثیر میزان پهنای باند بر متوسط هزینه را نشان می‌دهد. همان‌طور که مشاهده می‌شود با افزایش پهنای باند مقدار متوسط هزینه در هر سه الگوریتم کاهش می‌یابد. به این دلیل که با زیادشدن پهنای باند، امکان ارسال داده بیشتر در هر شکاف زمانی فراهم می‌شود که همین موضوع می‌تواند بر کاهش تأخیر ارسال و متوسط تخفیف یافته تأثیر گذارد. همچنین، طبق نتایج به‌دست‌آمده تأثیر وجود RIS نشان می‌دهد هزینه متوسط بدون استفاده از RIS افزایش می‌یابد. شبیه‌سازی با مقادیر متفاوت [۵۰, ۱۰۰, ۲۰۰, ۴۰۰] مگاهرتز [۲۴] در هر سه الگوریتم اجراشده است.

شکل ۸، تأثیر توان ارسالی بر مقدار متوسط هزینه را نشان می‌دهد. شبیه‌سازی‌ها با توان‌های متفاوت [۰.۱, ۰.۲, ۰.۳, ۰.۴] وات [۲۵] برای سه الگوریتم انجام شده‌اند. مشخص است که با افزایش توان ارسالی، به دلیل افزایش نرخ داده، احتمال دریافت موفقیت‌آمیز بسته در گیرنده افزایش می‌یابد؛ درنتیجه، متوسط هزینه در هر سه الگوریتم کاهش می‌یابد. علاوه بر این، مشاهده می‌شود که الگوریتم پیشنهادی DDPG در مقایسه با دو الگوریتم دیگر، از متوسط هزینه پایین‌تری برخوردار است.



شکل ۸: تأثیر افزایش توان ارسالی بسته در الگوریتم‌ها.

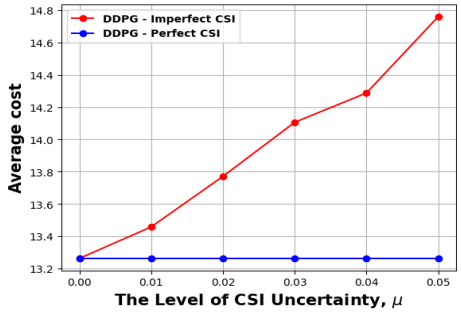


شکل ۹: تأثیر افزایش بهره بازتاب نسبی RIS در الگوریتم‌ها.

در شکل ۹، تأثیر بهره بازتاب نسبی[۳۵] RIS بر متوسط هزینه با مقادیر متفاوت [۱۰, ۱۲, ۱۶, ۱۸, ۲۰] برای هر سه الگوریتم موردبررسی قرارگرفته است [۲۶]. در شبیه‌سازی‌ها، مقدار بهره بازتاب نسبی با در نظر گرفتن بهره آنتن کاربر و ایستگاه پایه به ترتیب U = 0 dBi و B = 9.82 dBi [۲۸,۲۷]، با استفاده از رابطه (۱۹) محاسبه شده است. نتایج نشان می‌دهند که با افزایش بهره بازتاب نسبی، متوسط هزینه کاهش می‌یابد. این کاهش به دلیل بهبود کیفیت کانال است؛ زیرا با افزایش بهره بازتاب، انعکاس سیگنال‌ها بهبودیافته و تداخل کمتری ایجاد می‌شود که درنتیجه انتقال داده‌ها کارآمدتر صورت می‌گیرد. علاوه بر این، همان‌طور که مشاهده می‌شود الگوریتم پیشنهادی DDPG در مقایسه با دیگر الگوریتم‌ها، بهره‌وری بیشتری در کاهش متوسط هزینه نشان می‌دهد.

$$\nu = \frac{\rho_R}{\sqrt{\rho_u * \rho_B}} \qquad (۱۹)$$

در شکل ۱۰، برای بررسی تأثیر وضعیت کانال بر متوسط هزینه، شبیه‌سازی در دو حالت قطعیت[۳۶] و عدم قطعیت[۳۷] اجراشده است. به دلیل چالش‌های موجود در تخمین دقیق اطلاعات کانال از RIS به کاربر، فرض می‌کنیم که کانال بین RIS و کاربر به‌صورت $h_{u,R}$ = $\hat{h}_{u,R} + \Delta h_u$ فرموله شده است، که در آن Δh_u بردار خطای کانال و $\hat{h}_{u,R}$ تخمینی از کانال $h_{u,R}$ است [۳۰,۲۹]. بر اساس مدل کران خطای کانال Δh_u را مبتنی بر $\|\Delta h_u\|_2 \leq \mu\|\hat{h}_{u,R}\|_2$ [۲۹]، تعریف می‌کنیم، که در آن (μ) سطح عدم قطعیت CSI را نشان می‌دهد. طبق نتایج به‌دست‌آمده، دسترسی به اطلاعات دقیق وضعیت کانال به کاهش متوسط هزینه منجر می‌شود، زیرا این اطلاعات به بهینه‌سازی تصمیمات ارسال و دریافت کمک می‌کند و احتمال موفقیت در انتقال داده‌ها را افزایش می‌دهد. اما، در صورت وجود عدم قطعیت در اطلاعات، این بهینه‌سازی مختل شده و متوسط هزینه افزایش می‌یابد، زیرا تداخل و خطاهای انتقال بیشتری رخ می‌دهد. این شبیه‌سازی‌ها مطابق مقادیر جدول ۲ و ۳ با سطوح عدم قطعیت متفاوت [۰ ,۰.۰۱ ,۰.۰۲ ,۰.۰۳ ,۰.۰۴ , ۰.۰۵] اجرا شده‌اند [۳۱].



شکل ۱۰: تأثیر وضعیت اطلاعات کانال در روش پیشنهادی.

۶- نتیجه‌گیری

با توسعه شبکه‌های بی‌سیم نسل جدید مانند 5G و 6G و ظهور فناوری‌های نوین، اهمیت دسترسی به داده‌های به‌روز و حساس به زمان بیش‌ازپیش افزایش یافته است. به‌طوری‌که اتکا به داده‌های قدیمی می‌تواند پیامدهای جدی مالی و جانی در پی داشته باشد. در این راستا، پژوهش جاری به بررسی کاهش متوسط تجمعی تخفیفی دیرینگی اطلاعات در شبکه‌های بی‌سیم mmWave با استفاده از RIS پرداخته است. این مسئله به‌صورت یک مسئله بهینه‌سازی با فضای اقدامات پیوسته فرموله بندی شده که هدف آن حداقل کردن متوسط تخفیف یافته مبتنی بر AoI است. با توجه به ماهیت تصادفی مسئله و عدم دسترسی به اطلاعات کامل، از یک الگوریتم مبتنی بر یادگیری تقویتی عمیق بنام DDPG، برای حل مسئله استفاده شده است. این الگوریتم قادر است سیاست‌های بهینه برای زمان‌بندی کاربران، توان ارسالی، انتقال فاز در RIS و شکل‌دهی پرتو در BS را محاسبه کند. درنهایت عملکرد روش پیشنهادی ازنظر همگرایی و تأثیر پارامترهای مختلف شبکه مورد ارزیابی قرارگرفته و با دو الگوریتم زمان‌بندی تصادفی و نوبت چرخشی مقایسه شده است که نتایج نشان‌دهنده برتری عملکرد روش پیشنهادی ما است. به‌عنوان کارهای آینده، قصد داریم الگوریتم پیشنهادی را با در نظر

گرفتن نوع متفاوت دسترسی چندگانه و بهینه‌سازی دیرینگی اطلاعات همراه با بهینه‌سازی انرژی توسعه دهیم.

مراجع

[1] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," IEEE Journal on Selected Areas in Communications, vol.39, no.5, pp.1183–1210, 2021.

[2] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in 2012 Proceedings IEEE INFOCOM, pp.2731–2735, IEEE, 2012.

[3] A. Kosta, N. Pappas, V. Angelakis, et al., "Age of information: A new concept, metric, and tool," Foundations and Trends® in Networking, vol.12, no.3, pp.162–259, 2017.

[4] M. N. Soorki, W. Saad, M. Bennis, and C. S. Hong, "Ultra-reliable indoor millimeter wave communications using multiple artificial intelligence-powered intelligent surfaces," IEEE Transactions on Communications, vol.69, no.11, pp.7444–7457, 2021.

[5] A. Muhammad, M. Elhattab, M. A. Arfaoui, A. Al-Hilo, and C. Assi, "Age of information optimization in a ris-assisted wireless network," arXiv preprint arXiv:2103.06405, 2021.

[6] S. Gong, X. Lu, D. T. Hoang, D. Niyato, L. Shu, D. I. Kim, and Y.-C. Liang, "Toward smart wireless communications via intelligent reflecting surfaces: A contemporary survey," IEEE Communications Surveys Tutorials, vol.22, no.4, pp.2283–2314, 2020.

[7] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. de Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," IEEE Journal on Selected Areas in Communications, vol.38, no.11, pp.2450–2525, 2020.

[8] A. Mandal, R. Bhattacharjee, and A. Sinha, "Optimizing age-of-information in adversarial environments with channel state information," in 2022 14th International Conference on COMmunication Systems & NETworkS (COMSNETS), pp.522530, IEEE, 2022.

[9] A. Muhammad, M. Elhattab, M. Shokry, and C. Assi, "Leveraging reconfigurable intelligent surface to minimize age of information in wireless networks," in ICC 2022-IEEE International Conference on Communications, pp.2525–2530, IEEE, 2022.

[10] D. Sinha and R. Roy, "Optimal scheduling for maximizing information freshness and system performance in industrial cyber–physical systems," Computer communications, vol.169, pp.33–47, 2021.

[11] M. P. Abdollahi, H. Azarhava, A. Haghrah, and J. M. Niya, "On the rate and age of information for non-preemptive systems with prioritized arrivals and deterministic packet deadlines in iot networks," Ad Hoc Networks, vol.124, p.102717, 2022.

[12] Q. Zhang, X. Wang, Z. Li, and Z. Wei, "Design and performance evaluation of joint sensing and communication integrated system for 5g mmwave enabled cavs," IEEE Journal of Selected Topics in Signal Processing, vol.15, no.6, pp.1500–1514, 2021.

[13] S. Gong, L. Cui, B. Gu, B. Lyu, D. T. Hoang, and D. Niyato, "Hierarchical deep reinforcement learning for age-of-information minimization in irs-aided and wireless-powered wireless networks," IEEE Transactions on Wireless Communications, 2023.

[14] M. Sherman, S. Shao, X. Sun, and J. Zheng, "Optimizing aoi in uav-ris assisted iot networks: Off policy vs. on policy," IEEE Internet of Things Journal, 2023.

[15] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," IEEE transactions on wireless communications, vol.18, no.11, pp.53945409, 2019.

[16] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," IEEE transactions on wireless communications, vol.18, no.8, pp.4157–4170, 2019.

[17] T. Bai, C. Pan, Y. Deng, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Latency minimization for intelligent reflecting surface aided mobile edge computing," IEEE Journal on Selected Areas in Communications, vol.38, no.11, pp.2666–2682, 2020.

[18] M. Cui, G. Zhang, and R. Zhang, "Secure wireless communication via intelligent reflecting surface," IEEE Wireless Communications Letters, vol.8, no.5, pp.1410–1414, 2019.

[19] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," IEEE journal on selected areas in communications, vol.32, no.6, pp.1164–1179, 2014.

[20] Z. Lin, T. Lv, and P. T. Mathiopoulos, "3-d indoor positioning for millimeter-wave massive mimo systems," IEEE Transactions on Communications, vol.66, no.6, pp.2472–2486, 2018.

[21] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," arXiv preprint arXiv:1509.02971, 2015.

[22] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in International conference on machine learning, pp.448–456, pmlr, 2015.

[23] Y. Cao, T. Lv, Z. Lin, and W. Ni, "Delay-constrained joint power control, user detection and passive beamforming in intelligent reflecting surface-assisted uplink mmwave system," IEEE Transactions on Cognitive Communications and Networking, vol.7, no.2, pp.482–495, 2021.

[24] ETSI. TR 121 915 - V15.0.0 - digital cellular telecommunications - ETSI. 3GPP, 2019.

[25] X. Yu, G. Wang, X. Huang, K. Wang, W. Xu, and Y. Rui, "Energy efficient resource allocation for uplink ris-aided millimeter-wave networks with noma," IEEE Transactions on Mobile Computing, 2022.

[26] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for intelligent reflecting surface enhanced wireless networks," in 2019 IEEE Global Communications Conference (GLOBECOM), pp.1–6, IEEE, 2019.

[27] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget," IEEE Communications Surveys & Tutorials, vol.20, no.2, pp.870–913, 2017.

[28] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," IEEE Transactions on Vehicular Technology, vol.69, no.12, pp.14960–14973, 2020.

[29] G. Zhou, C. Pan, H. Ren, K. Wang, M. Di Renzo, and A. Nallanathan, "Robust beamforming design for intelligent reflecting surface aided miso communication systems," IEEE Wireless Communications Letters, vol.9, no.10, pp.1658–1662, 2020.

[30] C. Jia, J. Cheng, H. Gao, and W. Xu, "High-resolution channel estimation for intelligent reflecting surface-assisted mmwave communications," in 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, pp.1–6, IEEE, 2020.

[31] F. O. Catak, U. Cali, M. Kuzlu, and S. Sarp, "Uncertainty aware deep learning model for secure and trustworthy channel estimation in 5g networks," in 2023 12th Mediterranean Conference on Embedded Computing (MECO), pp.1–4, IEEE, 2023.

پاورقی‌ها:

[1] Freshness
[2] Age of Information (AoI)
[3] Delay
[4] Throughput
[5] Utilization
[6] Millimeter Wave (mmWave)
[7] Terahertz Waves (THz)
[8] Scattering
[9] Blockage
[10] Reconfigurable Intelligent Surfaces (RIS)
[11] Phase Shift
[12] Channel State Information (CSI)
[13] Greedy
[14] Wireless Powered Sensor Networ (WPSN)
[15] Deadline-aware Highest Latency First (DHLF)
[16] Rate of Information (RoI)
[17] Age to Rate Ratio of Information (ARRoI)
[18] Model-free
[19] Connected and Automated Vehicle (CAV)
[20] Base Station (BS)
[21] Beamforming
[22] Hierarchical deep reinforcement learning (HDRL)
[23] Uplink
[24] Energy Harvesting
[25] Deep Q-Learning (DQN)
[26] Proximal Policy Optimization (PPO)
[27] Time Division Multiple Access (TDMA)
[28] Analog Beamforming (ABF)
[29] Radio Frequency Chain (RFC)
[30] Constant Modulus (CM)
[31] Cascade Channel
[32] Angle of Arrival (AoA)
[33] Signal to Interference plus Noise Ratio (SINR)
[34] Deep Deterministic Policy Gradient (DDPG)
[35] Relative Reflection Gain
[36] Perfect
[37] Imperfect