# Resource Allocation in Open Radio Access Networks (O-RANs): A Correlated Equilibrium Game Theoretic Approach

Mahbod Hamzeh, Vesal Hakami*

School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran.
E-mails: m_hamze@cmps2.iust.ac.ir; vhakami@iust.ac.ir

**Short Abstract**
Radio access network (RAN) provides integrated wireless communication between user devices and the core network as an interface. Open Radio Access Network (O-RAN) is an updated model of RAN design, similar to cloud RAN (C-RAN) and virtualized RAN (vRAN). In addition to finer-grained disaggregation features in network functions and the use of general-purpose server architecture, O-RAN benefits from the use of smart controllers, making it more suitable for 5G and 6G networks. Consequently, resource allocation in O-RAN requires different approaches. Controllers in O-RAN are categorized into Near-Real-Time (Near-RT) and Non-Real-Time (Non-RT) types, where the former utilizes applications called xApps and the latter uses applications called rApps to manage network resources. In this paper, we address the problem of joint allocation of open radio units (O-RUs) to users and sub-channels to O-RUs under uncertain and time-varying channel quality conditions. We formulate the problem using Markov modulated games and propose a multi-agent tracking equilibrium algorithm to achieve convergence and track the stable operating point of the network. Furthermore, through simulations, the efficiency of the proposed method is compared and evaluated against previous related solutions.

**Keywords**
O-RAN, resource allocation, game theory, correlated equilibria.

## 1- Short Introduction

This article presents a distributed and scalable resource allocation method for O-RAN (Open Radio Access Network) in a multi-tenant cellular network that is compatible with uncertainty and dynamics in channel quality. The centralized and model-based approaches are not suitable for O-RAN, because they do not account for the unknown radio environment, the intelligent controllers, and the time-varying channel quality. As such, model-free and AI-based resource allocation algorithms are needed for O-RAN.

## 2- Proposed Work and Methodology

Unlike traditional centralized and model-based solutions, we use a Markov-modulated game with one logical agent in the near-RT RIC (Near-Real-Time Radio Intelligent Controller) for each O-RU (Open Radio Unit). The agents use two xApps to select sub-channels and allocate O-RUs, and aim to achieve Correlated Equilibrium (CE) as an analytical guarantee. We introduce a model-free resource allocation algorithm called RTRA (Regret Tracking Resource Allocation) that tracks state-dependent CE behavior based on channel changes. Our experiments show that RTRA outperforms previous solutions in terms of network utility in time-varying channel conditions. Using a simulation, the proposed solution shows its superiority in terms of utility and cumulative distribution of utility compared to several other single-agent and multi-agent learning algorithms.

## 3- Conclusion

Despite O-RAN's flexibility offering advantages, uncertain channel conditions create complex resource allocation challenges. We address this issue by proposing a multi-agent learning algorithm for joint resource allocation in O-RAN. Leveraging a Markov modulated game model, our algorithm converges to a correlated equilibrium, dynamically adapting to channel changes. Simulations demonstrate its superiority in utility compared to various single- and multi-agent learning approaches.

# تخصیص منابع در شبکه‌های دسترسی رادیویی باز: یک رویکرد نظریه بازی مبتنی بر تعادل همبسته

**مهبد حمزه**
دانشجوی دکتری، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت، تهران، ایران

**وصال حکمی**
استاد، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت، تهران، ایران

**چکیده**
شبکه دسترسی رادیویی به عنوان واسط میان دستگاه‌های کاربر و شبکه هسته، ارتباط بی‌سیم یکپارچه را میان دستگاه‌ها فرآهم می‌آورد. شبکه دسترسی رادیویی باز (O-RAN)، الگوی به روز شده‌ای از الگوهای طراحی RAN به مانند RAN آبری و RAN مجازی‌سازی شده است که علاوه بر ویژگی‌های جداسازی ریزدانه‌تر در کارکردهای شبکه و استفاده از معماری تحت سرویس‌دهنده‌های همه‌منظوره، از مزیت استفاده از کنترل‌کننده‌های هوشمند بهره می‌برد و مناسبت بیشتری با شبکه‌های نسل پنجم و ششم دارد. از این رو تخصیص منابع در O-RAN نیاز به رویکردهای متفاوتی دارد. کنترل‌کننده‌ها در O-RAN به دو صورت نزدیک به بلادرنگ و غیر بلادرنگ هستند که اولی از برنامه‌های کاربردی به نام xApp و دومی از برنامه‌های کاربردی به نام rApp به منظور کنترل منابع شبکه بهره می‌برد. ما در این مقاله مسئله تخصیص توأم واحدهای رادیویی باز (O-RU) به کاربران و زیرکانال به O-RUها را در شرایط کیفیت کانال غیرقطعی و متغیر با زمان در نظر می‌گیریم، مسئله را با استفاده از بازی‌های مُدوله شده مارکُفی فرمول‌بندی کرده و یک الگوریتم رهگیر تعادل چند عامله برای حصول همگرایی و رهگیری نقطه کاری پایدار شبکه پیشنهاد می‌دهیم. همچنین از طریق شبیه‌سازی، کارایی روش پیشنهادی با راهکارهای پیشین مرتبط مقایسه و ارزیابی می‌شود.

**کلمات کلیدی**
O-RAN، تخصیص منابع، بازی مُدوله شده مارکُفی، تعادل همبسته.

نام نویسنده مسئول: دکتر وصال حکمی
ایمیل نویسنده مسئول: vhakami@iust.ac.ir

تاریخ ارسال مقاله: ۱۴۰۲/۱۲/۰۷
تاریخ(های) اصلاح مقاله: ۱۴۰۳/۰۵/۲۲
تاریخ پذیرش مقاله: ۱۴۰۳/۱۲/۱۵

## ۱- مقدمه

شبکه دسترسی رادیویی به عنوان بخش کلیدی زیرساخت شبکه، ارتباط ایمن و یکپارچه‌ای را میان دستگاه‌های بی‌سیم مانند تلفن‌های همراه و رایانه‌ها فراهم می‌کند. این شبکه، تجهیزات کاربر (UE) را به شبکه هسته متصل کرده و از ارتباطات سرتاسری از طریق رابط‌های خاص پشتیبانی می‌کند.

با ظهور فناوری ابری، مدل رایج شبکه دسترسی رادیویی از حالت توزیع‌شده به سمت مدل متمرکز تغییر کرده است. با توسعه مفهوم مجازی‌سازی، برای ایجاد یک شبکه انتها به انتها با استفاده از مجازی‌سازی کارکردهای شبکه، RANمجازی‌سازی شده معرفی شد .مجازی‌سازی شده شامل نرم‌افزاری کردن کارکردهای سخت‌افزاری، مجزاسازی و مجازی‌سازی است. ترکیب فناوری ابری و مجازی‌سازی در O-RAN متجلی شده است [۱] که مزایای بسیاری نسبت به RAN مجازی‌سازی شده ارائه می‌دهد؛ از جمله رابط‌های باز و استاندارد، گزینه‌های متنوع تقسیم کارکرد و کنترل‌کننده‌های هوشمند. O-RAN قادر است از روش‌های هوش مصنوعی و کنترل‌کننده‌های هوشمند برای بهینه‌سازی عملکرد شبکه استفاده کند. بنابراین، O-RANرا می‌توان تکاملی بر الگوهای پیشین طراحی RAN دانست که انعطاف‌پذیری، قابلیت همکاری و هوشمندی بیشتری را به همراه دارد.-O RAN یک چارچوب معماری نوآورانه پیشنهاد کرده است که نقش محوری رابط‌های باز در آن به دلیل تنوع ذاتی تجهیزات شبکه است که نیازمند سازگاری بین محصولات تولیدکنندگان مختلف است. تعهد به دسترسی باز و انعطاف‌پذیری بیشتر در سیاست O-RAN به ویژه در مزیت تفکیک-پذیری/مجزاشدگی آن مشهود است که بر این اساس، گره عمومی gNB به سه واحد مجزا تقسیم می‌شود: واحد مرکزی (O-CU)، واحد توزیع شده (O-DU) و واحد رادیویی(O-RU) .

شبکه O-RAN شامل دو کنترل‌کننده هوشمند رادیویی به نام‌های نزدیک به بلادرنگ (Near-RT RIC) و بلادرنگ (Non-RT RIC) است. در Near-RT RIC، مجموعه‌ای از برنامه‌های کاربردی سفارشی به نامxApp ، همراه با خدمات پشتیبانی ضروری، نقشی کلیدی در کارکردهایی که تاخیر حداکثر ۱ ثانیه‌ای دارند، ایفا می‌کنند. این برنامه‌ها مدیریت منابع رادیویی، تجزیه و تحلیل، و کنترل RAN را ممکن می‌سازند. Non-RT RIC از برنامه‌هایی به نام rApp پشتیبانی می‌کند که برای ارائه خدمات ارزش افزوده با تاخیر بیش از ۱ ثانیه مناسب هستند. همچنین به عنوان مکملی بر xApp و rApp، برنامه‌های کاربردی توزیع شده و سفارشی‌سازی شده‌ای به نام dApp، به عملگران شبکه اجازه مدیریت و کنترل ریزدانه و بلادرنگ در CU و DUها می‌دهند.

اگرچه ویژگی‌هایی مانند باز بودن، جداسازی، و کنترل‌کننده‌های هوشمند در O-RAN می‌توانند سازگاری، کاهش هزینه، و پویایی خدمات را فراهم کنند، اما انعطاف‌پذیری O-RAN منجر به ایجاد مشخصات و معماری پیچیده‌ای می‌شود که با رویکردهای سنتی نمی‌توان به درستی با آن مواجه شد. همچنین،

استفاده از باندهای رادیویی فرکانس بالا مانند sub6، mmWave، و THz در معماری‌های اخیر RAN در نسل پنج و شش، مانندO-RAN ، به دلیل محدودیت‌های ذاتی پوشش، نیاز به مدیریت دقیق منابع دارد. پرداختن به چالش عدم قطعیت کیفیت کانال در تخصیص منابع، که توسط عواملی مانند انسداد و محوشدگی ایجاد می‌شود، نیازمند راه‌حلی است که در چارچوب قابلیت‌های فنی O-RAN عمل کند. تعدادی از واژگان اختصاری پر تکرار در جدول ۱ ذکر شده است.

**جدول ۱- واژگان اختصاری**

| | |
|---|---|
| RAN | Radio Access Network |
| RTRA | Regret Tracking Resource Allocation |
| O-RAN | Open RAN |
| CE | Correlated Equilibrium |
| NE | Nach Equilibrium |
| MNO | Mobile Network Operator |
| LA | Logical Agent |

### ۲-۱- انگیزه تحقیق

به منظور افزایش بهره‌وری، معماری‌های زیرساخت جدید عمدتاً به صورت چند مستاجره طراحی شده‌اند و عملگرهای متنوع شبکه تلفن همراه را در نظر می‌گیرند. این مقاله به چالش تخصیص کارآمد O-RU و انتخاب زیرکانال در یک شبکه سلولی چندعملگره (اصطلاحاً چندمستاجره) می‌پردازد. در این شبکه، زیرساخت O-RAN شامل زیرکانال‌ها و واحدهای رادیویی به صورت اشتراکی استفاده می‌شود و هر مستاجر به کاربران متعددی خدمات ارائه می‌دهد.

- **تخصیص منابع به صورت توزیع شده با تضمین بهینگی**

سطح بالای درجه آزادی در شبکه توزیع‌شده O-RAN ما را مجبور به استفاده از راه‌حل‌های تخصیص منابع به صورت توزیع‌شده و با بهره‌گیری از اطلاعات محلی می‌کند. راه‌حل‌های تخصیص منابع در شبکه‌های بی‌سیم به‌طور سنتی از رویکردهای متمرکز استفاده می‌کنند که منجر به سربار زیاد و عدم مقیاس‌پذیری می‌شود [۲، ۳]. بیشتر رویکردهای تخصیص منابع غیرمتمرکز نیز هیچ ضمانت تحلیلی برای الگوریتم‌های خود ارائه نمی‌دهند [۴-۶]، در حالی که تضمین‌های تحلیلی (مانند بهینه سراسری یا محلی و تعادل استراتژیک) در الگوریتم‌های تخصیص منابع می‌توانند تضمینی ریاضی برای دستیابی به یک راه‌حل بهینه فراهم کنند.

- **بهینه سازی سودمندی شبکه فارغ از مدل**

با توجه به پویایی محیط رادیویی و نیاز به کنترل هوشمند درO-RAN ، استفاده از الگوریتم‌های تخصیص منابع مبتنی بر هوش مصنوعی و فارغ از مدل ضروری است. با این حال، بسیاری از راه‌حل‌های موجود از رویکردهای مبتنی بر مدل استفاده می‌کنند که با واقعیت محیط O-RAN همخوانی ندارند [۲، ۳، ۶].

- **سازگاری در برابر کانال‌های متغیر با زمان**

در O-RAN، استفاده از فناوری‌های شبکه ارتباطی مبتنی بر باندهای mmWave، sub6 و THz حساسیت بالایی از کیفیت کانال را نسبت به عوامل تضعیف ایجاد می‌کند. این مهم مسئله عدم قطعیت را در شبکه ایجاد می‌کند که بر پویایی کیفیت کانال تأثیر گذاشته و با زمان متغیر است. با این حال، رویکردهای تخصیص منابع موجود این جنبه را نادیده گرفته و یک محیط شبکه ایستا یا قطعی را فرض گرفته‌اند [۲، ۳، ۶]. در نتیجه، نیاز به سازگاری و رهگیری نقطه عملیاتی بهینه شبکه وجود دارد.

### ۳-۱- نوآوری اصلی مقاله

- **فرمول‌بندی مسئله تخصیص منابع O-RAN به عنوان یک بازی مُدوله شده مارکُفی (Markov-modulated game)**

به منظور ارائه یک روش توزیع‌شده و مقیاس‌پذیر برای تخصیص منابع O-RAN که می‌تواند با شرایط کانال سازگار شود، از بازی مُدوله‌شده مارکُفی استفاده می‌کنیم که در آن به ازای هر O-RU یک عامل منطقی (Logical Agent (LA)) در near-RT RIC در نظر گرفته می‌شود. در این نوع از بازی‌ها، توابع سودمندی بازیگران و در نتیجه تعادل استراتژیک بازی به وضعیت شبکه وابسته است. در بازی طراحی شده، وضعیت بازی در هر زمان با توجه به کیفیت زیرکانال‌ها گسسته شده و در طول زمان متغیر خواهد بود. در روش پیشنهادی، واحد Near-RT RIC شامل دو xApp می‌شود: یکی برای انتخاب زیرکانال و دیگری برای تخصیص O-RUها. با بهره‌گیری از مزایای ذاتی تعادل همبسته[1] نسبت به تعادل نش[2]، مانند افزایش درجه همکاری و کارایی بالاتر، هدف ما ارائه CE به عنوان یک تضمین تحلیلی است. با توجه به اینکه در CE، عمل هر عامل نشان‌دهنده پاسخ بهینه به شرایط محیطی غالب و اقدامات پیش بینی شده سایر عامل‌ها است، دستیابی به CE را می‌توان به عنوان ایجاد یک اجماع نزدیک به بهینه در نظر گرفت.

- **الگوریتم یادگیری مبتنی بر پشیمانی به منظور رهگیری تعادل در بازی تخصیص منابع**

ما یک الگوریتم تخصیص منبع فارغ از مدل به نام RTRA[3] معرفی می‌کنیم. RTRA یک الگوریتم یادگیری توزیع شده بر اساس بازی، چند عامله و مبتنی بر رهگیری پشیمانی است که به طور موثر استراتژی‌های تصمیم‌گیری عامل‌ها را بر اساس تاریخچه مشاهدات خودشان شکل می‌دهد که منجر به تضمین حداقل سربار می‌شود. RTRA عامل‌ها را قادر می‌سازد تا رفتار CE وابسته به حالت را یاد گرفته و بر اساس تغییرات کانال‌های مابین O-RU ها و UEها رهگیری کنند. این قابلیت رهگیری بدون درک صریح از فرآیند محوشدگی کانال به دست می‌آید.

- **آزمایش های ارزیابی عملکرد روش معرفی شده**

ما شبیه‌سازی‌هایی را به منظور ارزیابی عملکرد RTRA در مقایسه با راهکارهای پیشین انجام داده‌ایم. نتایج، عملکرد برتر RTRA را از نظر سودمندی به دست آمده در شرایط کیفیت کانال متغیر با زمان را نشان می‌دهد.

### ۴-۱- بخش‌بندی مقاله

در ادامه، بخش دوم به بررسی جامع تلاش‌های قبلی تخصیص منابع می‌پردازد. بخش سوم با توصیف مدل‌های پیوند، شبکه و سودمندی، توضیحی از مدل شبکه ما ارائه می‌کند. بخش چهارم بازی تخصیص منابع و CE وابسته به وضعیت را تعریف و الگوریتم تخصیص منبع رهگیر پشیمانی را تشریح می‌کند. بخش پنجم به ارزیابی کارایی الگوریتم RTRA اختصاص دارد. در نهایت، بخش ششم حاوی جمع‌بندی و نتیجه‌گیری است.

[2] Correlated Equilibrium (CE)
[3] Nash Equilibrium (NE)
[4] Regret-tracking resource allocation (RTRA)

## ۲- مرور کارهای مرتبط پیشین

در این بخش به بررسی تعدادی از پژوهش‌های صورت گرفته در حوزه تخصیص منابع RAN می‌پردازیم و در ادامه ضمن ارزیابی رویکردهای فوق، نقاط ضعف و چالش‌های مورد غفلت واقع شده آن‌ها را عنوان می‌کنیم. در [۶] یک O-RAN در نظر گرفته است که در آن به منظور دستیابی به تعادل بار و بهبود عملکرد نرخ موثر شبکه، روشی چند-عاملی پیشنهاد کرده است. در نهایت نشان داده شده است که روش فوق توانسته است در مقایسه با تعدادی از روش‌ها به نرخ موثر بالاتری دست یابد و در عین حال تعادل بار بهتری بین O-RUها ایجاد کند. در [۷]، یک RAN ابری در نظر گرفته شده و به منظور تخصیص واحدهای رادیویی راه دور به MNOها، یک الگوریتم نظریه بازی را در قالب یک معماری شبکه مبتنی بر شبکه‌های نرم‌افزار محور ارائه کردند. آنها سناریوهای مختلفی را بر اساس تنوع تعداد کاربران در میان MNOها در نظر گرفتند. در [۲] با هدف تخصیص عادلانه منابع محاسباتی میان دو گروه مختلف از کاربران و اطمینان از کیفیت مطلوب خدمات آنها، مسئله تخصیص منابع محاسباتی را با استفاده از الگوریتم یادگیری تقویتی حل کردند و نتایجی نزدیک به بهینه به دست آوردند.

در [۳]، یک شبکه مبتنی بر معماری O-RAN در نظر گرفته شده است که با توجه به توافق سطح سرویس کاربران، هر عملکرد شبکه به DU یا CU اختصاص داده شده و منابع رادیویی نیز به RU اختصاص داده شده‌اند. هر دو مسئله تخصیص منابع فوق با استفاده از الگوریتم‌های مبتنی بر یادگیری تقویتی انجام شده‌اند. نتایج شبیه سازی تحقیق فوق نشان داده است که روش معرفی شده در مقایسه با حالتی که همیشه کارکردها تنها به CU یا DU اختصاص داده می‌شوند، از نظر معیارهایی مانند تاخیر و نسبت تحویل بسته و گذردهی بهبود یافته است.

در [۸]، برای بهینه سازی و به حداقل رساندن تاخیر تحت خدمات مختلف، یک روش برش منابع چند عاملی مبتنی بر یادگیری پیشنهاد شده است که در آن هر برش نقش یک عامل را ایفا کرده و سپس بر سر منابع با عامل‌های دیگر رقابت می‌کند. نتایج روش فوق نشان داده است که در مقایسه با برخی از روش-های مبتنی بر یادگیری، تاخیر و نرخ افت بسته کمتری دارد.

قابل ذکر است که برخی از محققان در تخصیص منابع RAN از الگوهای قدیمی طراحی RAN مانند RAN ابری استفاده کرده اند [۷] ، همچنین چندین مقاله مانند [۲، ۳] یک راه حل متمرکز ارائه کرده‌اند، راه حل‌های توزیع نشده به دلیل عدم امکان استفاده از اطلاعات محلی، سربار بالایی دارند و مقیاس-پذیری آنها نیز بسیار پایین است. اگرچه برخی از کارهای انجام شده در زمینه تخصیص منابع RAN از یک رویکرد توزیع شده در فرآیند تصمیم‌گیری استفاده کرده‌اند، اما همه آنها بر اساس یک فرض غیر واقعی، یعنی قطعیت تنظیمات شبکه، از جمله وضعیت کانال، استوار بوده‌اند [۶-۸]. عدم توجه به ویژگی پویایی شبکه باعث می‌شود که مقادیر سودمندی و در نتیجه نقطه تعادل پیکربندی شبکه به طور غیر واقعی تحت تأثیر قرار گیرد. لازم به ذکر است که از نظریه بازی در کاربردهای مختلف شبکه استفاده شده است، به عنوان مثال [۹] به منظور شناسایی گره‌های مخرب از نظریه بازی بهره برده است.

در [۱۰] یک راهکار دو سطحی برای برش‌بندی منابع O-RAN، میان عملگرهای مجازی (در سطح اول) و سپس از عملگرهای مجازی به کاربران (سطح دوم) ارائه شده است. برش‌بندی سطح اول به عنوان یک rApp با استفاده از بازی انتساب انجام شده و تخصیص منابع در سطح دوم با استفاده از یادگیری تقویتی عمیق صورت گرفته است. با این وجود مدل تغییرات کانال ارتباطی میان gNB با کاربران، مدل محوشوندگی سایه (با انحراف معیار سایه گوسی) در نظر گرفته شده که نوعی ساده‌سازی از سناریو واقعی است. همچنین [۱۱] برای بهبود گذردهی D2D سلولی از یادگیری تک عاملی استفاده کرده است. روشی که ما به منظور بهبود گذردهی ارائه می‌دهیم مبتنی بر یادگیری چندعاملی است.

## ۳- مدل شبکه و فرضیات

در این بخش مدل شبکه، پیوندهای ارتباطی و سودمندی خود را شرح می‌دهیم. در جدول ۲ نمادهای مهم مدل سیستم به همراه توضیح مقتضی ارائه شده است.

**جدول ۲- شرح نمادهای مهم**

| | |
|---|---|
| $\mathcal{K}$ | مجموعه O-RUها |
| $\mathcal{M}$ | مجموعه MNOها |
| $\mathcal{N}_m$ | مجموعه UEهایی که به $m$-اُمین MNO متصل هستند |
| $v_{k,i}^m$ | تحت پوشش $k$-اُمین O-RU قرار داشتن $i$-اُم UE (متعلق به عملگر $m$) |
| $\boldsymbol{v}_i^m$ | امکان اتصال $i$-اُم UE از $m$-اُمین MNO به O-RUها |
| $\mathcal{F}$ | مجموعه فرکانس‌های ارتباطی |
| $c_{(k)}^t(f)$ | انتخاب زیرکانال $f$ برای ارسال سیگنال در زمان $t$ توسط $k$-اُمین O-RU |
| $s_{k,i}^{m,t}(f)$ | کیفیت زیرکانال $f$ میان $k$-اُمین O-RU و $i$-اُمین UE (متعلق به $m$-اُمین MNO) در زمان $t$ |
| $\Lambda_{k,m}^t$ | خدمت‌رسانی $k$-اُمین O-RU به $m$-اُمین MNO |

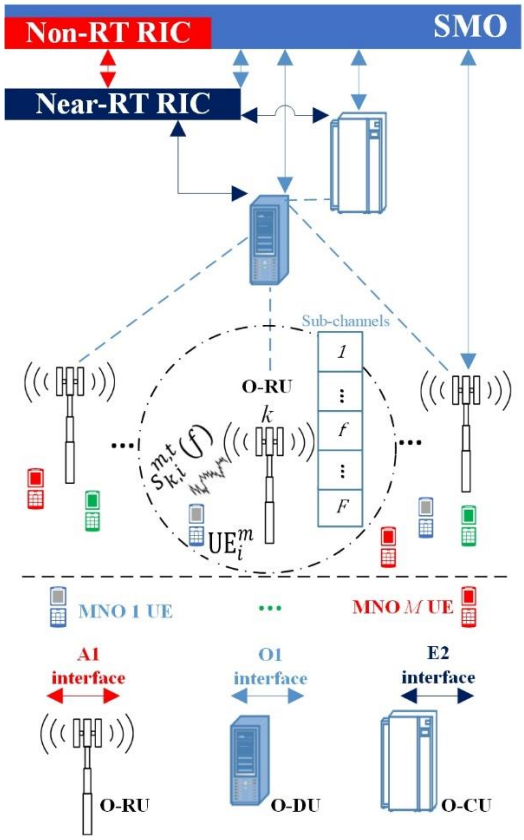
### ۳-۱- مدل شبکه

مطابق شکل ۱ ما یک معماری شبکه سلسله مراتبی بر اساس O-RAN در نظر گرفته‌ایم که از کارکردهایی مانند مدیریت خدمات و ساماندهی[4]، کنترل-کننده بلادرنگ، O-CU، O-DU و O-RU تشکیل شده است. این کارکردها از رابط‌های باز برای برقراری ارتباط با یکدیگر استفاده می‌کنند. به عنوان مثال، کنترل‌کننده بلادرنگ از رابط E2 برای تعامل با O-CU و O-DU استفاده کرده و بهینه‌سازی شبکه و تخصیص منابع را انجام می‌دهد. همچنین کنترل‌کننده غیر بلادرنگ از رابط A1 برای ارائه خدمات مدیریتی به کنترل‌کننده بلادرنگ استفاده می‌کند. شبکه ما شامل یک O-CU و O-DU می‌باشد و از نماد $\mathcal{K} = \{1, \dots, k, \dots, K\}$ برای نمایش مجموعه O-RUها استفاده می‌کنیم. همچنین زیرساخت O-RAN در نظر گرفته شده میان مجموعه MNOها که با $\mathcal{M} = \{1,2, \dots, M\}$ نشان داده می‌شود مشترک است. فرض می‌کنیم که روی اجرای xAPPهای مستقر در Near-RT RIC توسط ارائه دهنده زیرساخت، کنترل متمرکز وجود دارد و بنابراین برخلاف یک سناریوی نامتمرکز که در آن ممکن است xAppها از سوی MNOهای مختلف مستقر شده باشند، در مدل در نظر گرفته شده در این مقاله به لحاظ عملی امکان تضاد میان xAppها به

[5] Service management and orchestration (SMO)

حداقل می‌رسد و استحکام معماری پیشنهادی در مقابل تعارضات احتمالی به‌طور مؤثری تضمین می‌شود. همچنین می‌توان فرض نمود که مکانیزم‌های ایزوله‌سازی (جداسازی) منابع تعبیه شده‌اند به گونه‌ای که اطمینان می‌دهند هر xApp در محدوده منابع تخصیص یافته خود عمل می‌کند. این مکانیزم‌ها، مانند تکنیک‌های مجازی‌سازی و کانتینریزاسیون، به کاهش ریسک تعارض و تضاد بین xAPPها کمک می‌کنند.

UEها می‌توانند به شبکه یکی از MNOها متصل شده و از خدمات شبکه بهره‌مند شوند، هچنین از $\mathcal{N}_m = \{1,2,\dots,N_m\}$ به منظور نشان دادن مجموعه UE که به $m$-اُمین MNO متصل هستند استفاده می‌کنیم.




**شکل ۱- معماری شبکه سلولی مبتنی بر O-RAN**

### ۲-۳- مدل پیوندهای ارتباطی

نماد $v_{k,i}^m$ بیانگر این است که UE $i$-اُم (متعلق به عملگر $m$) آیا تحت پوشش $k$-اُمین O-RU قرار دارد یا خیر، همچنین هر UE می‌تواند از تعداد متناهی O-RU سیگنال دریافت کند.

$$v_{k,i}^m = \begin{cases} 1 & O-RU\ k \text{ is in range } i \in N_m \\ 0 & Otherwise \end{cases} \quad (۱)$$

نماد $q_{m,k}$ به عنوان یک متغیر دودویی زمانی مقدار ۱ دارد که $m$-اُمین MNO تحت سرویس $k$-اُمین O-RU باشد و در غیر این صورت مقدار ۰ خواهد داشت. همچنین فرض می‌کنیم که هر O-RU در شبکه زیرساخت تنها می‌تواند به یک MNO در هر واحد زمانی تخصیص داده شود، یعنی خواهیم داشت: $\forall\, k \in \mathcal{K}\ \sum_{m\in\mathcal{M}} q_{m,k} = 1$.

امکان اتصال UE $i$-اُم از $m$-اُمین MNO به O-RUها با $\boldsymbol{v}_i^m = [\,v_{1,i}^m, v_{2,i}^m, \dots, v_{k,i}^m\,]^T$ نشان داده شده است. همچنین فرض می‌کنیم که مجموعاً تعداد $F$ زیرکانال فرکانسی برای ارتباط بین هر جفت O-RU-UE وجود دارد که با $f\epsilon\mathcal{F} \stackrel{\text{def}}{=} \{1,2,\dots,F\}$ نشان داده می‌شوند. از نماد $c_{(k)}^t(f) \in \{0,1\}$ به منظور نشان دادن اینکه $k$-اُمین O-RU زیرکانال $f$ را برای ارسال سیگنال در زمان $t$ انتخاب می‌کند استفاده می‌کنیم. همچنین بردار $\boldsymbol{c}^t = \left(c_k^t(f)\right)_{k\in K, f\epsilon\mathcal{F}}$ بیانگر زیرکانال‌های انتخاب شده توسط تمامی O-RUها در زمان $t$ است. کیفیت کانال را می‌توان با گسسته‌سازی یک مدل مقدار پیوسته کانال (مبتنی بر متغیرهای تصادفی گوسی مختلط متقارن دورانی) به $Q$ سطح مختلف محاسبه نمود، نماد $s_{k,i}^{m,t}(f)$ بیانگر کیفیت زیرکانال $f$ میان $k$-اُمین O-RU و $i$-اُمین UE (متعلق به $m$-اُمین MNO) در زمان $t$ است:

$$s_{k,i}^{m,t}(f) \in \mathcal{S} \stackrel{\text{def}}{=} \{\mathcal{s}_1, \dots, \mathcal{s}_q, \dots. \mathcal{s}_Q\} \quad (۲)$$

اتصال O-RUها و MNOها در زمان $t$ از طریق ماتریس $\boldsymbol{\Lambda}_{K*M}^t$ نمایش داده شده است، درایه $\Lambda_{k,m}^t \in \{0,1\}$ مشخص می‌کند که آیا $k$-اُمین O-RU به $m$-اُمین MNO خدمت‌رسانی می‌کند یا خیر. وضعیت امکان اتصال $k$-اُم O-RU به MNOها در سطر $k$ و به صورت $\boldsymbol{\Lambda}_{\boldsymbol{k}}^t = [\Lambda_{k,1}^t, \dots, \Lambda_{k,M}^t]$ قابل مشاهده می‌باشد. همچنین O-RUهایی که آمادگی ارائه خدمت به $m$-اُمین MNO را دارند با ستون m و به صورت $\boldsymbol{\Lambda}_m^t = [\Lambda_{1,m}^t, \dots, \Lambda_{k,m}^t, \dots, \Lambda_{K,m}^t]^T$ مشخص می‌شوند.

$$\boldsymbol{\Lambda}_{K*M}^t = \begin{bmatrix} \Lambda_{1,1}^t & \cdots & \Lambda_{1,m}^t & \cdots & \Lambda_{1,M}^t \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ \Lambda_{k,1}^t & \cdots & \Lambda_{k,m}^t & \cdots & \Lambda_{k,M}^t \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ \Lambda_{K,1}^t & \cdots & \Lambda_{K,m}^t & \cdots & \Lambda_{K,M}^t \end{bmatrix} \quad (۳)$$

بر همین اساس بردار کیفیت زیرکانال کاربر $i$ از $m$-اُمین MNO با هر یک از O-RUها به صورت رابطه (۴) قابل تعریف است که البته تعداد محدودی مقدار غیر صفر در آن وجود دارد (به دلیل قید محدودیت پوشش). $s_{k,i}^{m,t}(f)$ نشان دهنده کیفیت کانال مابین کاربر $i$ (متعلق به $m$-اُمین MNO) و $k$-اُمین O-RU در لحظه $t$ هست:

$$s_i^{m,t}(f) = \left[s_{1,i}^{m,t}(f), \dots, s_{k,i}^{m,t}(f), \dots, s_{K,i}^{m,t}(f)\right]^T \quad (۴)$$

**فرض۱** (*مدل محوشوندگی کُند کانال*): به طور کلی، تکامل کانال‌های محو شوندگی (Fading) را می‌توان به عنوان یک زنجیره مارکُف حالت محدود (Finite State Markov Chain (FSMC)) با فضای حالت $\mathcal{S}$ (رابطه (۲) را ببینید) مدل کرد [۱۲-۱۴]. در این مدل، فرض می‌کنیم که مجموعه $\Gamma = \{\Gamma_1 = 0, \Gamma_2, \dots, \Gamma_{Q+1} = \infty\}$ از Q+1 محدوده نسبت سیگنال به نویز و تداخل (SINR) تشکیل شده است که هر محدوده معرف وضعیت متفاوتی از شرایط کانال است. از این رو اگر SNR زیرکانال $f$ برای پیوند $ik$ برابر با $\gamma$ باشد و $\Gamma_Q < \gamma < \Gamma_{Q+1}$، خواهیم داشت: $s_{(g,l),i}^{m,t}(f) = \mathcal{s}_Q$.

یک FSMC به طور کلی می‌تواند تکامل کانال‌های محو شده را مدل کند [۱۵]. ما از یک مدل FSMC آهسته به منظور توصیف محوشوندگی آهسته کانال‌ها در طول زمان استفاده می‌کنیم. این بدان معناست که بسیار محتمل است $\gamma$ در همان ناحیه‌ای که در زمان $t$ بود، در زمان $t+1$ باشد و احتمال کمی نیز وجود دارد تا به ناحیه‌متفاوتی تغییر پیدا کند. بنابراین احتمال حالت ماندگار بودن در حالت $Q$ به صورت رابطه (۵) داده می‌شود:

$$v_q = \int_{\Gamma_q}^{\Gamma_{q+1}} g(\gamma)d\gamma\,, \qquad q = 1,2,\dots,Q \quad (۵)$$

که در آن $g(\gamma)$ تابع چگالی احتمال $\gamma$ است. در عمل ما متکی بر هیچ گونه توزیع بر روی کیفیت کانال نیستیم و برای هر پیوند، مدل کانال‌های Slow Rayleigh را مد نظر داریم. در این مدل نسبت سیگنال به نویز $\gamma$ یک متغیر تصادفی توزیع شده نمایی با تابع چگالی احتمال $\frac{1}{\bar{\gamma}} e^{\frac{-\gamma}{\bar{\gamma}}}$ است که $\bar{\gamma} = \mathbb{E}[\gamma]$ میانگین نسبت سیگنال به مجموع نویز می‌باشد.

**فرض ۲** (*ساختار ماتریس گذار محوشوندگی کند کانال*): در مدل FSMC برای تغییرات کُند کانال، ماتریس حالت گذار به فرم رابطه (۶) در نظر گرفته می‌شود [۱۶] که در آن $\theta = 1 - 2\sigma$ و $\sigma = 0(f_d\tau)$، که $\tau$ طول مدت زمان

ارسال بسته روی کانال می‌باشد، همچنین $f_d\tau$ نشان دهنده تغییر فرکانس داپلر (Doppler frequency shift) است و سرعت محو شدن کانال را نسبت به طول بسته مشخص می‌کند، هر چه قدر این مقدار کوپکتر باشد، سرعت محوشوندگی کانال کمتر خواهد بود.

$$\mathbb{P}=\begin{bmatrix}\theta & \sigma & 0 & 0 & 0 & \dots & 0 & \sigma\\ \sigma & \theta & \sigma & 0 & 0 & \dots & 0 & 0\\ 0 & \sigma & \theta & \sigma & 0 & \dots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ \sigma & 0 & 0 & 0 & 0 & \dots & \sigma & \theta\end{bmatrix} \quad (۶)$$

**توضیح ۱:** مدل زیر کانال FSMC برای ارتباطات OFDMA و تحت هر دو نوع محوشوندگی Nakagami-m [۱۲] و Rayleigh [۱۳] ارائه شده است و مطابق آزمایشات اعتبار سنجی گسترده صورت گرفته در این مراجع، ویژگی‌های آماری کانال انتخابگر فرکانسی، در این مدل‌های مارکفی حفظ می‌شود. در بسیاری از سیستم‌های OFDM عملیاتی، فرض مدلسازی هر زیرکانال به عنوان یک FSMC، فرضی نزدیک به واقعیت است، چراکه پهنای باند کلی خیلی بزرگتر از پهنای باند همدوسی (coherence) کانال محوشونده است [۱۴].

**توضیح ۲:** علی‌رغم **توضیح ۱**، الگوریتم پیشنهادی (RTRA) در بخش ۴-۳، نوعی الگوریتم تقریب تصادفی است که در آن عامل‌ها هیچ فرض صریحی روی ساختار مارکفی برای تحول بازی در زمان ندارند. در نظریه بازی‌های مدوله شده مارکفی [۱۹] فرض مارکفی روی تغییرات وضعیت بازی فقط در "تحلیل" همگرایی ضعیف الگوریتم و جهت اطمینان از چابکی رفتار جمعی عامل‌ها در رهگیری تعادل همبسته متغیر با زمان استفاده می‌شود. درواقع خود الگوریتم یادگیری چندعاملی می تواند در محیط‌های متغیر با زمان عمومی (تحت تاثیر هر نوع فرآیند تصادفی بیرونی اعم از ایستان و غیر ایستان ) اجرا گردد.

### ۳-۳- مدل تابع سودمندی

به طور واقع‌بینانه فرض می‌کنیم که هر O-RU ($k\in\mathcal{K}$) در هر واحد زمانی $t$ هیچ یک از اقلام اطلاعاتی مرتبط با موارد زیر را در اختیار ندارد:

- سایر O-RUها و پروفایل تصمیمات اتخاذ شده توسط آنها یعنی: $\boldsymbol{\Lambda}_{-k}^{t},\boldsymbol{C}_{-k}^{t}$
- وضعیت کانال ارتباطی تمام UEهای شبکه یعنی: $\boldsymbol{s}\stackrel{\text{def}}{=}\left(s_i^{m,t}(f)\right)_{m\in\mathcal{M},i\in\mathcal{N}_m,f\in\mathcal{F}}$

فرض می‌کنیم $A_k^T$ بهره آنتن فرستنده $k$-اُمین O-RU باشد و $A_{m,i}^R$ بهره آنتن گیرنده $i$-اُمین UE تحت پوشش آن باشد. همچنین با در نظر گرفتن توان ارسالی $P_k$ برای $k$-اُمین O-RU، میزان SINR برای UE $i$-اُم عبارت است از:

$$SINR_{k,i}^{m,t}(f)=\frac{P_K\ A_k^T\ s_{k,i}^{m,t}(f)\ A_{m,i}^R}{\sigma_k^2(f)+\ I_k^t(f)} \quad (۷)$$

که $\sigma_k^2(f)$ واریانس نویز گوسی سفید جمع‌شونده است. همچنین $I_k^t(f)$ عبارت است از تداخل هم-کانالی دریافتی از سوی سایر O-RUها.

$$I_k^t(f)=\sum_{k'\neq k}^{k'\in\mathcal{K}} p_{k'}\ A_{k'}^T\ s_{k',i}^{m,t}(f)\ A_{m,i}^R\ \times\ c_{k'}(f)\ \ \forall f\in F \quad (۸)$$

با این تفاسیر نرخ شانون قابل دستیابی برای $i$-اُمین UE متعلق به $m$-اُمین MNO که توسط $k$-اُمین O-RU پوشش داده شود عبارت است از :

$$\mathcal{R}_{k,i}^{m,t}(f)=\mathcal{B}\ log_2[1+\ SINR_{k,i}^{mt}(f)] \quad (۹)$$

در رابطه (۹)، $\mathcal{B}$ میزان پهنای باند اختصاص داده شده به زیرکانال است. برای هر O-RU میزان پاداش را می‌توانیم بر اساس مجموع نرخ کاربران تحت پوشش تعریف نماییم:

$$\mathcal{R}_k(\Lambda_k,\boldsymbol{c}_k,\boldsymbol{\Lambda}_{-k},\boldsymbol{c}_{-k},\boldsymbol{s})=\sum_{f\in F}\sum_{m\in\mathcal{M}}\sum_{i=1}^{\mathcal{N}_m}\mathcal{R}_{k,i}^{m,t}(f)\times\Lambda_{k,m}\times c_k(f) \quad (۱۰)$$

با توجه به رابطه (۱۰) پاداش دریافتی هر O-RU به واسطه تداخل هم کانال با سایر O-RUها دچار افت خواهد شد. همچنین در صورتی که طی تصمیمات O-RUها پوشش کامل برای کلیه کاربران شبکه فراهم نگردد، پاداش برای O-RUها لحاظ نخواهد شد؛ به عبارت نمادین:

$$U_k(\Lambda_k,\boldsymbol{\Lambda}_{-k},c_k,\boldsymbol{c}_{-k},s^t)=\begin{cases}\mathcal{R}_k(\Lambda_k,\boldsymbol{\Lambda}_{-k},s^t) & If\ (\boldsymbol{\mathcal{V}_m^i})^T.\boldsymbol{\Lambda}_m>0\ \forall m\in M\ ,\forall i\in\mathcal{N}_m\\ 0 & \text{Otherwise}\end{cases} \quad (۱۱)$$

## ۴- فرمول‌بندی مسئله

### ۴-۱- نگاشت مسئله تخصیص منابع بازی مُدوله شده مارکُفی

به دلیل ماهیت تصادفی و پویای توابع سودمندی O-RU، نیاز به رهگیری نقطه بهینه سراسری در طول زمان وجود دارد، از طرفی یک رویکرد بهینه سازی متمرکز برای این مسئله در ابعاد بالا غیرممکن خواهد بود. ما مسئله تخصیص منابع را به عنوان یک بازی $\mathbb{G}=(\mathcal{K},\boldsymbol{\Lambda},C,\boldsymbol{S},(u_k(\cdot)),(\sigma_k)_{k\in\mathcal{K}})$ مدل می‌کنیم این بازی حائز اجزای زیر است:

- **مجموعه عامل‌ها:** به ازای هر O-RU، داخل Near-RT RIC یک عامل منطقی در نظر می‌گیریم. به تبعیت از نمادهای به کار رفته در مدل شبکه، ما این مجموعه را با $k$ نشان می‌دهیم.
- **مجموعه اعمال (تصمیمات):** مجموعه اعمال $k$-اُمین عامل بوسیله بردار $\boldsymbol{\Lambda}_k\in\{0,1\}^{\mathcal{M}}$ و $C_k\in\{0,1\}^{\mathcal{F}}$ نشان داده می‌شود.
- **وضعیت بازی:** در واحد زمانی $t$ پروفایل وضعیت بازی بر اساس $\boldsymbol{s}=\left(s_i^{m,t}(f)\right)_{f\in\mathcal{F},m\in\mathcal{M},i\in\mathcal{N}_m}$ تعریف می‌شود.
- **توابع سودمندی:** به دلیل وابستگی سودمندی به وضعیت کانال‌ها، مقادیر این توابع با زمان به صورت تصادفی عوض می‌شوند. همچنین امکان محاسبه مقدار تابع $u_k(\cdot)$ توسط عامل‌ها پیش از اتخاذ تصمیم بر اساس یک فرمول ریاضی وجود ندارد. در عوض هر عامل بایستی یک اقدام را اتخاذ نماید و یک مقدار سودمندی را به عنوان بازخورد دریافت نماید، این بازخورد شامل مقدار نرخ واقعی $\mathcal{R}_{k,i}^{m,t}(t)$ است که توسط هر کاربر متعلق به $m$-اُمین MNO و تحت پوشش $k$-اُمین O-RU گزارش شده است. همچنین Near-RT RIC تصمیم می‌گیرد که با ارسال یک سیگنال دودویی مشروط، پاداش فوق را به حساب عامل فوق اعتبار دهد یا خیر.
- **استراتژی انتساب:** در هر واحد زمانی از اجرای بازی $\mathbb{G}$، هر عامل مانند $k$، یک استراتژی مختلط (Mixed) مثل $\sigma_k^t=[\sigma_k^t(a)]_{a\in A_k}$ برای انتخاب یک تصمیم انتساب استفاده می‌کند. سپس یک مقدار عددی سودمندی $u_k^t$ بابت این واحد زمانی دریافت می‌کند و مقدار استراتژی $\sigma_k^{t+1}$ خودش را بر اساس تاریخچه مشاهداتش بروزرسانی می‌نماید. تاریخچه بازی از منظر هر O-RU یک تاریخچه اختصاصی به فرم $h_k^t$ است و با طول t یعنی:

$$h_k^t=(a_k^0,u_k^0,a_k^1,u_k^1,\dots,a_k^{t-1},u_k^{t-1})\in\mathcal{H}_k^t\stackrel{\text{def}}{=}(\{0,1\}^{\mathcal{M}}\times\{0,1\}^f\times\mathbb{R})^t \quad (۱۲)$$

بنابراین، هر عامل $k$ عمل خود را به صورت خودمختار طبق نگاشت زیر برمی‌گزیند که $\Delta(A_k)$ بیانگر همه توزیع احتمالات ممکن روی فضای اعمال عامل $k$ است.

$$\sigma_k^{t+1}:\cup_t H_k^t \rightarrow \Delta(A_k) \quad (۱۳)$$

### ۲-۴- هدف بهینه‌سازی

به منظور دستیابی به رضایت سراسری در تصمیم‌گیری‌های تخصیص منابع، ما بر تعمیم مهم NE، که به عنوان CE شناخته می‌شود، تمرکز می‌کنیم. برخلاف NE که در آن هر عامل فقط استراتژی خود را در نظر می‌گیرد، CE با اجازه دادن به هر عامل برای در نظر گرفتن توزیع مشترک اقدامات عامل‌ها، عملکرد بهتری را به دست می‌آورد. با این حال، در بازی تخصیص منابع $\mathbb{G}$ ، شرایط کانال و همچنین سودمندی عامل‌ها مطابق با روند محوشدگی آهسته دگرگون می‌شود، از این رو مجموعه CE باید به صورت وابسته به وضعیت تعریف شود. برای تعریف CE، ابتدا لازم است که فضای توزیع احتمالات روی اعمال جمعی عامل‌ها را (که با نماد $\Delta(A)$ نمایش داده می‌شود) بیان کنیم:

$$\Delta(A) \stackrel{\text{def}}{=} \{p \in \mathbb{R}^{|A|}; p(a) \geq 0, \sum_{a\in A} p(a) = 1 \quad (۱۴)$$

حال یک CE وابسته به وضعیت $s$ مثل $\pi_s$، یک توزیع احتمالاتی متعلق به فضای $\Delta(A)$ است (یعنی $\pi_s \in \Delta(A)$) و مجموعه کلیه CEهای وابسته به وضعیت بازی $\mathbb{G}$ با نماد $\mathbb{C}(s)$ تعریف می‌شوند:

$$\mathbb{C}(s) = \{\pi_s: \sum_{a_{-k}\in A_{-k}} \pi_s(a, a_{-k}) \cdot [u_k(b, a_{-k}; s) - u_k(a, a_{-k}; s)] \leq 0 \quad (۱۵)$$
$$, \forall a, b \in A_k, k \in K\}$$

مجموعه تعادل $\mathbb{C}$ در بازی ما با زمان متغیر است. همچنین، چون وضعیت کانال به دلیل محوشدگی آهسته تغییر می‌کند، باید این مجموعه را در طول زمان رهگیری کنیم و فقط به مفهوم همگرایی تکیه نکنیم. به عبارت دیگر یک تعادل همبسته $\pi_s \in \mathbb{C}(s)$ یک توزیع احتمال در فضای اقدام مشترک عامل‌ها است که دارای ویژگی تعادلی است. هدف بازی این است که عامل‌های Near-RT RIC به یک اجماع (تعادل) برسند به نحوی که در هر دوره زمانی بهترین انتساب منابع (زیرکانال و واحدهای رادیویی) صورت بگیرد. ما از مزایای CE [۱۷] نسبت به NE مانند درجه بالاتر همکاری و کارایی بالاتر بهره می‌بریم که ذاتا مفهومی غیر همکارانه است. در CE عمل هر عامل بهترین پاسخ به وضعیت محیط و به اعمال تخمینی سایر عامل‌هاست، بنابراین حصول یک CE به منزله شکل‌گیری یک اجماع زیربهینه در میان تصمیمات عامل‌های بازی تخصیص منابع است.

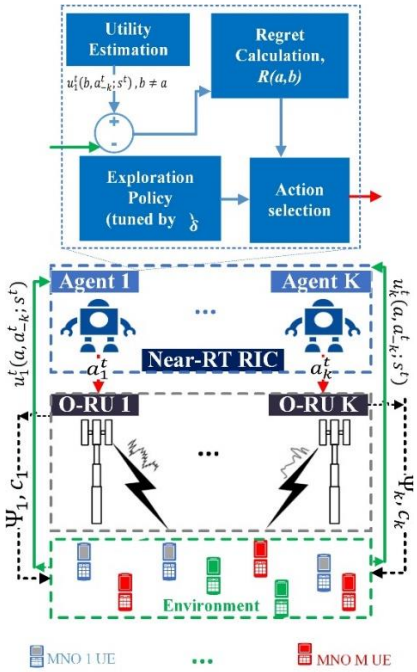
### ۳-۴- الگوریتم یادگیر چندعامله جهت اجرا در Near-RT RIC

مطابق معماری ارده شده در شکل ۲ به منظور یادگیری و رهگیری رفتار تعادلی وابسته به وضعیت توسط عامل‌های موجود در Near-RT RIC در بازی تخصیص منابع، ما یک الگوریتم توزیع‌شده به نام RTRA ارائه می‌دهیم که بر مبنای روال مبتنی بر پشیمانی است [۱۸]. در RTRA هر عامل به صورت پویا استراتژی تخصیص منابع خود را به منظور تقویت اعمالی که بابت عدم استفاده کافی از آن‌ها در گذشته پشیمان است، به‌روزرسانی می‌کند. بر اساس نتیجه نظری ارائه شده در [۱۸، ۱۹] ما استدلال می‌کنیم که اگر همه عامل‌ها از الگوریتم پیروی کنند، رفتار جمعی تخصیص منابع آن‌ها اجماع مبتنی بر CE مطلوب را رهگیری خواهد کرد.

مزیت اصلی الگوریتم پیشنهادی این است که هر عامل تنها متکی به مشاهدات تاریخچه‌ای اختصاصی خود (مجموع نرخ محقق شده کاربران تحت پوشش، اعمال انتخابی خود و سیگنال کنترلی برای تامین قید پوشش) به منظور تنظیم استراتژی تخصیص منابع خود بدون نیاز به مذاکره صریح با سایر عامل‌های موجود در شبکه است. بنابراین روش ما نیازمندی‌های عدم تمرکز (توزیع‌شدگی) و سربار پایین تخصیص منابع در شبکه‌های RAN باز را تامین می‌کند. در RTRA اقدام تخصیص منابع عامل‌ها با پشیمانی از عدم اتخاذ آن در گذشته تقویت می‌شود. به طور مشخص، پشیمانی میانگین $k$-اُمین عامل از بابت انتخاب $a \in A_k$ بجای عمل $b \in A_k$ در زمان $t$ به شکل رابطه (۱۶) تعریف می‌شود:

$$R_k^t(a,b) = \left[\sum_{\eta\leq t: a_k^\eta = a} \varepsilon(1-\varepsilon)^{t-\eta}[u_k^\eta(b, a_{-k}^t; s^t) - u_i^\eta]\right]^+ \quad (۱۶)$$

رابطه فوق تفسیری ساده از اندازه‌گیری پشیمانی میانگین $k$-اُمین عامل در بازه زمانی $t$ به دلیل عدم انتخاب اقدام $b$، هر زمان که اقدام $a$ در گذشته انتخاب شده بوده است را ارائه می‌دهد. توجه داشته باشید که این میانگین‌گیری به شیوه‌ای تخفیفی انجام می‌شود تا سودمندی‌های تازه‌تر را بالاتر از سودمندی‌های قدیمی‌تر ارزش‌گذاری کنند. به شکل شهودی عامل تخفیف ثابت (گام زمانی) $0 < \varepsilon \ll 1$ مشخص کننده فراموشی نمایی گذشته است.




**شکل ۲- معماری شبکه مبتنی بر O-RAN**

نکته ظریف دیگر این است که در بازی تخصیص منابع ما $k$-اُمین عامل نمی‌تواند تابع سودمندی $u_k^\eta(b, a_{-k}^t; s^t)$ خود را برای زمان‌های $\{\eta \leq t : a_k^\eta = a\}$ ارزیابی نماید. این به آن دلیل است که عامل $k$ به شکلی واقع‌بینانه تنها مقادیر نمونه $u_k^\eta$ را از اقدام حقیقی $a$ اتخاذ شده خود درک می‌کند. در عوض ما می‌توانیم پشیمانی میانگین را به صورت رابطه (۱۷) تخمین بزنیم:

$$\hat{R}_k^t(a,b) = \left[\sum_{\eta\leq t: a_k^\eta = b} \varepsilon(1-\varepsilon)^{t-\eta}\left[\frac{\sigma_k^\eta(a)}{\sigma_k^\eta(b)}\right]u_k^\eta - \sum_{\eta\leq t: a_k^\eta = a} \varepsilon(1-\varepsilon)^{t-\eta}u_k^\eta\right]^+ \quad (۱۷)$$

که $\sigma_k^t(\cdot)$ احتمال اینکه $k$-اُمین عامل اقدام $a \in A_k$ را در زمان $t$ انتخاب کند می‌باشد و $\hat{R}_k^t(a,b)$ بیانگر پشیمانی میانگین متناظر است. درواقع پشیمانی تخمینی، تفاوت میانگین سودمندی بر بازه‌هایی که عمل $b$ استفاده شده بوده و بازه‌هایی که عمل $a$ استفاده شده بوده را اندازه می‌گیرد. بر پایه $\hat{R}_k^\eta(a,b)$ هر عامل استراتژی تخصیص منابع خود را به این ترتیب روزرسانی می‌کند: اگر

اقدام $a$ را در زمان $t$ انتخاب کند، احتمال انتخاب اقدام $b$ در زمان $(t+1)$ متناسب با پشیمانی میانگین از $a$ به $b$ است؛ و با باقی‌مانده این احتمال، اقدام $a$ دوباره انتخاب می‌شود. از این رو اقدام تخصیص منابع با پشیمانی بزرگتر در بازه زمانی حاضر، با احتمال بیشتری در بازه زمانی بعدی انتخاب می‌شود، به طور خاص در حالتی که اقدام $a$ در زمان $t$ انتخاب شده باشد، احتمالات بازی $k$-اُمین عامل در زمان $(t+1)$ برای $b \in A_k, b \neq a$ به شکل زیر مشخص می‌شود:

$$\sigma_k^{t+1}(b) = \left(1-\frac{\delta}{t^\rho}\right) min\left\{\frac{\hat{R}_k^t(a,b)}{\mu}, 1\right\} + \frac{\delta}{2t^\rho} \quad (۱۸)$$

$$\sigma_k^{t+1}(a) = 1 - \sum_{b \in A_k} \sigma_k^{t+1}(b) \quad (۱۹)$$

که $\mu > \hat{R}_k^t(a,b)$ ضریب نرمالسازی است که به عنوان لَختی (Inertia) به‌روزرسانی می‌تواند دیده شود و به منظور جلوگیری از بزرگتر شدن جمع احتمالات بازی از "یک" استفاده می‌شود. $0 < \delta < 1$ یک فاکتور اکتشاف است که برای یادگیری مداوم عامل‌ها از سودمندی‌شان ضروری است. چنین اکتشافی منجر به این می‌شود که تمامی اقدامات با کمترین فرکانس نیز همچنان انتخاب شوند و عدم انتخاب شدن برخی از اقدامات با احتمال انتخاب کم، نفی شود. بر اساس روابط (۱۷ و ۱۸) هر عامل اساسا نیازی به در اختیار داشتن صریح پروفایل وضعیت کاربران ندارد و تنها با دریافت بازخوردهای ضمنی از گذردهی کاربران تحت پوشش هر O-RU، استراتژی خود را به روز رسانی می‌نماید. در ادامه یک شبه کد کامل از RTRA به ازای هر عامل و در هر برش زمانی تحت نام الگوریتم ۱ شرح داده شده است.

**تعریف ۱** (*متوسط رفتار جمعی تخصیص منابع عامل‌ها*): متوسط رفتار جمعی تخصیص منابع عامل‌ها تا لحظه $t$، با نماد $z_\varepsilon^t \in \Delta(A)$ نشان داده می‌شود که بیانگر فرکانس تجربی جمعی بازی تا لحظه $t$ است و در آن همه عامل‌ها به طور همزمان الگوریتم RTRA را اجرا می‌کنند. یعنی برای یک پروفایل اقدام مشترک $a = (a_1, \dots, a_I) \in A$ داریم: $z_e^t(a) = \sum_{\eta \le t} \varepsilon(1-\varepsilon)^{t-\eta} \mathbb{I}_{\{a^\eta = a\}}$. در [۱۹] نشان داده شده است که $z_\varepsilon^t$ به شکل حدی مجموعه متغیر با زمان $\mathbb{C}(s)$ را دنبال می‌کند. در یک محیط مارکُفی، شرط نظری که قابلیت رهگیری فوق را تضمین می‌دهد این است که زنجیره مارکُف زیربنایی در فواصل زمانی نسبتا طولانی تغییر می‌کند که این شرط در کار ما نیز برآورده می‌شود (رجوع به **فرض ۱ و۲**). در نتیجه در اجرای طولانی بازی تخصیص منابع ما در هر وضعیت کانال سراسری $s \in S$، هیچ عاملی از استراتژی تخصیص منابع $\sigma_k$ خود پشیمان نمی‌شود یعنی هر کدام به شکلی بهینه به محیط و اقدامات سایرین پاسخ می‌دهند (از آن جهت که به بیشترین نرخ گذردهی دست می‌یابد.). این عملکردهای بهینه محلی در تمام عامل‌ها منجر به عملکرد نزدیک به بهینه قابل توجهی در شبکه می‌شود. در واقع، در سطح اجتماعی، رفتار تخصیص منابع جمعی $z_\varepsilon^t$ در وضعیت $s$ با اجماع مطلوب $\mathbb{C}(s)$ مطابقت دارد.■

**الگوریتم ۱: RTRA**

```
//Initialization
Begin
 Choose σ_k^0 arbitrarily
 Choose a = a_k^0
 Use (11) to calculate u_k^t
 while 1 do
  Use (17)to calculate R̂_k^t(a,b) for all b ∈ A_k
  Use (18)to calculate σ_k^{t+1}(b) for all b ∈ A_k, b ≠ a
  Use (19) to calculate σ_k^{t+1}(a)
 end while
End
```

RTRA یک الگوریتم تخصیص منابع کم هزینه است. تنها قسمتی که قابلیت بحث بیشتری دارد محاسبه‌ی $\hat{R}_k^t(a, 1-a)$ است. به مانند [۱۹] ما فرمولی بازگشتی موثرتری مد نظر داریم که این خود از محاسبه $\hat{R}_k^t(a,b)$ از ابتدا در هر دوره جلوگیری می‌کند که منجر به کاهش پیچیدگی می‌شود. در نتیجه در هر تکرار، RTRA تنها به تعداد کمی محاسبات و مقایسات معمول به همراه تولید عدد تصادفی به منظور انتخاب عمل بعدی $a_k^{t+1}$ نیاز دارد:

$$\hat{R}_k^t(a,b) = \hat{R}_k^{t-1}(a,b) + \varepsilon\left(\left[\frac{\sigma_k^t(a)}{\sigma_k^t(b)} u_i^t(a_k^t).\, \mathbb{I}_{\{a_k^t=b\}} - u_k^t(a_k^t).\, \mathbb{I}_{\{a_k^t=a\}}\right]^+ - \hat{R}_k^{t-1}(a,b)\right) \quad (۲۰)$$

## ۵- شبیه‌سازی راهکار پیشنهادی

در ادامه، شرح تنظیمات، روش‌های مورد مقایسه و نتایج آزمایش‌ها به تفصیل پرداخته شده است.

### ۵-۱- تنظیمات شبیه‌سازی

شبکه ما یک شبکه سلولی چند مستاجره شش ضلعی با فاصله بین سایتی ۵۰۰ متری ما بین O-RUها است. تحت فرکانس حامل به میزان ۵/۹ GHz، با در نظر گرفتن یک زیر-کانال متشکل از ۱۰ بلاک منبع فیزیکی[۵] متوالی که هر کدام دارای پهنای باند ۳۶۰ کیلوهرتز بوده و فاصله بین زیر-حامل (Sub-carrier spacing) نیز ۱۵ کیلوهرتز و کل پهنای باند نیز۱۰ مگاهرتز بوده است، در نتیجه مدت زمان هر برش زمانی ۰/۵ میلی‌ثانیه و تعداد زیر-کانال‌ها ۳ محاسبه می شود، زیرا بر اساس اطلاعات [۲۰] حاصل ضرب بازه برش زمانی و پهنای باند PRB همیشه ۱۸۰ می‌باشد. گام زمانی ($\varepsilon$) به صورت ثابت و برابر با ۰/۰۰۱ می‌باشد، فاکتور اکتشاف در بازه $0 < \delta < 1$ در نظر گرفته شده و پارامتر $\rho$ مقداری کوچکتر از ۰/۲۵ خواهد داشت،جدول ۳ سایر تنظیمات شبکه را نشان می‌دهد.

**جدول ۳- برخی از تنظیمات شبیه‌سازی**

| پارامتر شبکه | مقدار |
|---|---|
| تعداد O-RUها | ۱۰ |
| تعداد MNO | ۴ |
| تعداد اجرا | ۱۰ |
| تعداد برش زمانی هر اجرا | ۳۰۰۰ |
| تعداد کاربر هر MNO | ۱۵ |
| نویز | -۱۷۴ dBm/Hz |
| توان O-RU | ۴۰ dBm |

### ۵-۲- روش‌های مورد مقایسه

در این بخش ۳ روش مختلف معرفی شده است که در بخش شبیه‌سازی

[6] Physical Resource Block (PRBs)

روش معرفی شده خود را با آنها مقایسه می‌کنیم. یک شبه کد نیز برای هر یک از این روش‌ها ارائه شده است.

۱) یادگیری چندعامله تکاملی، مبتنی بر پویایی تکثیری[۶] [۲۲]: این روش یک روش یادگیر چند عامله می‌باشد که در ابتدا هر عامل یک عمل تصادفی را انتخاب می‌کند، پس از آن در هر زمان هر عامل، سودمندی به دست آمده خود را مشاهده نموده و با مقدار متوسط سودمندی جمعیت، که از رابطه (۲۱) محاسبه می‌شود، مقایسه می‌نماید. اگر میزان سودمندی مشاهده شده کمتر از میزان فوق بوده باشد، در زمان بعدی یک عمل را به شکل تصادفی انتخاب می‌کند و در غیر این صورت همان عمل قبل را مجددا برای زمان بعدی انتخاب خواهد کرد. در رابطه (۲۱) $\vartheta_a$ متوسط سودمندی حاصل از انتخاب عمل $a$ و $x_a$ نسبت تعداد O-RUهایی که عمل $a$ را انتخاب نموده‌اند به کل تعداد O-RUها است. شبه کد روش EGTRD با الگوریتم ۲ مشخص شده است.

$$\bar{\vartheta} = \sum_{a \in A} \vartheta_a x_a \qquad (۲۱)$$

**الگوریتم ۲: EGTRD**

```
//Initialization
Choose a = a_k^0
Use (11) to calculate u_k^t
While (1):
  Use (21) to calculate ϑ̄
  Compare ϑ̄ and U_k for all k ∈ K
  Choose a = a_k^{t+1}
  End while()
```

۲) یادگیری مبتنی بر پشیمانی[۷] غیر تطبیقی برای RAN ابری [۷]: روش فوق مشابه الگوریتم ۱ ارائه شده است، اما گام زمانی آن با زمان متغیر خواهد بود که موجب تفاوت در عملکرد الگوریتم می‌شود و به همین دلیل، تنها همگرایی به تعادل را تضمین می‌کند. بنابراین در مسائلی که ویژگی‌های آماری محیط یادگیری، با زمان متغیر هستند قابلیت رهگیری تعادل را در پی نخواهد داشت. در واقع، در این روش، پارامتر ε در رابطه (۲۰) به جای مقدار ثابت، مقدار $1/t$ را به منظور دستیابی به میانگین‌گیری یکنواخت خواهد داشت.

۳) یادگیری Q تک عامله[۸] [۲۱]: یک روش یادگیری تک عامله هست که در آن هر O-RU، در هر زمان یک تخمینی از هر عمل در نظر می‌گیرد که با رابطه (۲۲) به روز رسانی می‌شود، پس از آن انتخاب عمل بعدی به این صورت خواهد بود که به احتمال بیشتری عملی با تخمین بهتر را انتخاب می‌نماید و با احتمال کمی نیز یکی از سایر اعمال را بر می‌گزیند، چنین تخمینی از عمل $a$ در زمان $t$ به صورت $Q(a)_k^t$ تعریف می‌شود و خواهیم داشت:

$$Q(a)_k^{t+1} = Q(a)_k^t + (\varepsilon * [U_k(\Psi_k, \boldsymbol{\Psi}_{-k}, c_k, \boldsymbol{c}_{-k}, s^t) - Q(a)_k^t]) \qquad (۲۲)$$

**الگوریتم ۳: SQL**

```
//Initialization
Q_k^0 arbitrarily
Choose a = a_k^0
While (1):
  Use (11) to calculate u_k^t
  Use (22) to calculate Q(a)_k^{t+1}
  Choose a = a_k^{t+1}
End while()
```

درواقع، روش‌های مورد مقایسه تنوعی از الگوریتم‌های یادگیری پرکاربرد از هر دو نوع تک_عاملی و چند_عاملی را در برمی‌گیرد هدف ما از انتخاب این الگوریتم‌ها جهت مقایسه این بوده است که روی نیاز به الگوریتمی با هر دو قابلیت یادگیری چندعامله و رهگیری نقطه بهینه در طول زمان تاکید شود و نشان دهیم الگوریتم‌های یادگیری موجود که عمدتا فاقد یکی از این قابلیت هستند، مناسبت کمتری برای سناریوی توزیع شده و متغیر با زمان در O-RAN دارند.

### ۳-۵- ارزیابی راهکار پیشنهادی

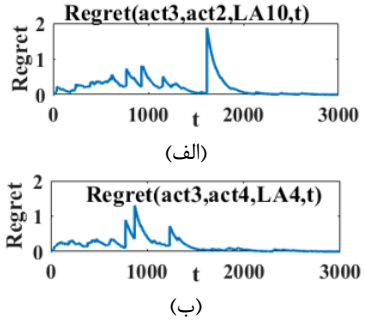
در این بخش، عملکرد الگوریتم پیشنهادی خود را از دو منظر ارزیابی کرده‌ایم. منظر نخست بررسی همگرایی الگوریتم پیشنهادی است و منظر دوم شامل مقایسه روش پیشنهادی با روش‌های معرفی شده در بخش ۵-۲ بر اساس تغییر تنظیمات مختلف می‌باشد.

### نتایج همگرایی روش پیشنهادی

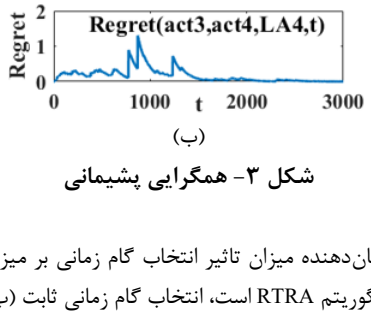
با توجه به اینکه شاخص همگرایی الگوریتم RTRA نزدیک شدن تخمین پشیمانی $\hat{R}_k^t(a,b)$ به مقدار صفر است، باید نشان دهیم که به ازای مقادیر مختلف $k$ (عامل‌های منطقی مختلف) و نیز مقادیر مختلف a و b (زوج اعمال مختلف)، درایه‌های ماتریس $\hat{R}_k^t(a,b)$ به سمت صفر همگرا می‌شوند.

شکل ۳ (الف)، همگرایی الگوریتم پیشنهادی را بر اساس نمودار پشیمانی از عدم انتخاب عمل ۲ به جای عمل ۳ (از میان کلیه مجموعه اعمال ممکن) در طول زمان از دید عامل منطقی ۱۰ بیان می‌کند و شکل ۳ (ب) نمودار مشابهی را از دید عامل ۴ و میان زوج اعمال ۴ و ۳ آن نمایش می‌دهد.

در نمودارهای شکل ۳ تقریباً ۲۰۰۰ برش زمانی برای رسیدن الگوریتم به همگرایی مورد نیاز بود که برابر با ۱ ثانیه می‌باشد (طول هر برش زمانی ۰/۵ ثانیه است). این مقدار به نسبت مدت زمان برقراری یک تماس صوتی نمونه میان کاربران (که می‌تواند حداقل چندین ثانیه طول بکشد)، مقدار مناسبی است.




(الف)




(ب)

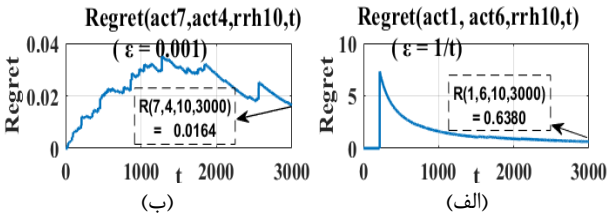
**شکل ۳- همگرایی پشیمانی**

شکل ۴ نشان‌دهنده میزان تاثیر انتخاب گام زمانی بر میزان پشیمانی در طول زمان در الگوریتم RTRA است، انتخاب گام زمانی ثابت (ب) در مقایسه با گام زمانی متغیر با زمان (الف) کاهش میزان پشیمانی در طول زمان را به دنبال

[7] Evolutionary game-theoric replicator dynamics (EGTRD)
[8] Non adaptive regret-based learning for C-RAN (RBRA)
[9] Single-agent Q-learning (SQL)

دارد.

با توجه به فاصله بیشتر پشیمانی با صفر در الگوریتم با گام زمانی متغیر، در شرایطی که بواسطه محوشوندگی کُند کانال‌ها، ویژگی‌های آماری محیط یادگیری با زمان تغییر می‌کنند افت کارایی در این الگوریتم از الگوریتم با گام زمانی ثابت بیشتر خواهد بود (چنانکه در نمودارهای ارزیابی کارایی بعدی قابل ملاحظه است).




**شکل ۴- تاثیر گام زمانی بر میزان پشیمانی در طول زمان (الف) گام زمانی متغیر با زمان و (ب) گام زمانی ثابت**

### نتایج مقایسه‌ای روش پیشنهادی

نمودارهای نمایش داده شده درشکل‌های ۵،۶و۷ به ترتیب تاثیر تغییرات تعداد O-RU، MNO و زیر-فرکانس را بر روی میانگین نرخ به دست آمده توسط O-RUها به ازای هر واحد از پهنای باند به صورت بازده طیفی نشان می‌دهند که معیار فوق معیاری است که موئد میزان بهینگی استفاده از پهنای باند در سامانه ارتباطی است.

با افزایش تعداد زیرکانال‌ها، میزان بازده طیفی سودمندی به صورت میانگین افزایش پیدا کرده است چراکه تعداد انتخاب‌های فرکانس برای هر LA بیشتر شده و این امر موجب کاهش تداخلات می‌گردد.

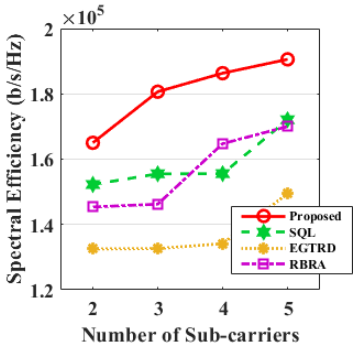
با افزایش تعداد MNO میانگین بازده طیفی سودمندی O-RUها کاهش می‌یابد چراکه با توجه به محدود بودن O-RUها تعداد MNOهایی که ممکن است از هیچ O-RUای خدمات نگیرند افزایش می‌یابد.

با افزایش تعداد O-RU نیز به دلیل افزایش تعداد ارسال بین کاربران و O-RUها، میزان تداخلات افزایش می‌یابد و موجب کاهش میانگین سودمندی و طبیعتا بازده طیفی سودمندی O-RUها می‌شود.

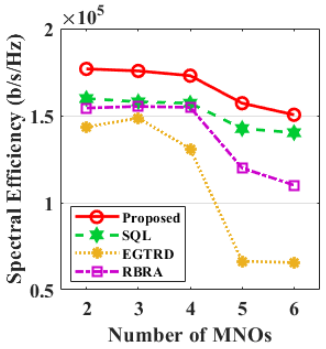
روش SQL به دلیل استفاده از اندازه گام ثابت توانسته خود را با شرایط کانال تطبیق دهد و در نتیجه بر روش EGTRD برتری دارد زیرا روش EGTRD قابلیت رهگیری تعادل را ندارد.

روش RBRA بر خلاف SQL توانایی انطباق با تغییرات رژیم احتمالی کانال را ندارد، اما به دلیل اینکه یک روش مبتنی بر CE است، در بسیاری از تست‌ها با روش SQL برابری کرده است.البته RBRA توانسته در تمامی تست ها نسبت به روش EGTRD که بر پایه NE است برتری داشته باشد.

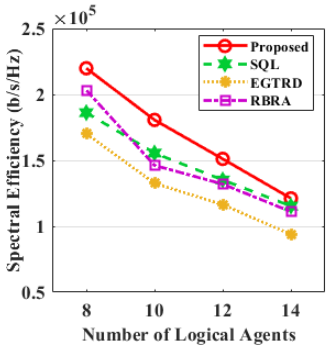
در تمام آزمایش‌ها، الگوریتم پیشنهادی به دلیل قابلیت رهگیری تغییرات رژیم احتمالاتی ناشی از عدم قطعیت کانال به عنوان یک روش مبتنی بر CE چند عاملی، توانسته است به میانگین نرخ بالاتری نسبت به سایر روش‌ها دست یابد.




**شکل ۵- اثر افزایش تعداد زیر-فرکانس‌ها بر سودمندی عامل‌ها به صورت بازده طیفی**




**شکل ۶- اثر افزایش تعداد MNOها بر سودمندی عامل‌ها به صورت بازده طیفی**

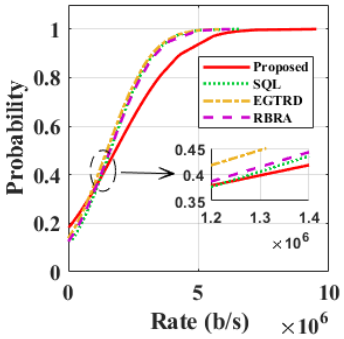



**شکل ۷- اثر افزایش تعداد عامل‌ها بر سودمندی عامل‌ها به صورت بازده طیفی**

شکل ۸ نمودار تابع توزیع تجمعی تجربی بر روی میزان سودمندی بدست‌آمده از عامل‌ها را نشان می‌دهد. از آنجا که روش RBRA مشابه روش پیشنهادی مبتنی بر روال پشیمانی است اما تقریب تصادفی بازگشتی آن با اندازه گام متغییر با زمان است، این امر موجب می‌شود که متوسط‌گیری پشیمانی در طول زمان به صورت یکنواخت صورت گیرد و قابلیت تطبیق با عدم قطعیت بیرونی با شرایط کانال را نداشته باشد. روش EGTRD نیز اگرچه یک روش یادگیر چند-عامله است، اما صرفا همگرایی به تعادل را در شرایط مانایی فرآیند تصادفی کانال را تضمین می‌کند و از قابلیت رهگیری بی‌بهره است، این در حالی است که روش RTRA رهگیری مجموعه تعادل‌ها را ممکن می‌سازد. روش SQL با توجه به به‌کارگیری اندازه گام به‌روزرسانی ثابت، قابلیت مواجهه با عدم قطعیت بیرونی ناشی از شرایط کانال را دارد اما به عنوان یک روش یادگیری تک‌عامله امکان تطبیق با اثر متقابل تصمیمات عامل‌های مستقر در xApp (عدم قطعیت درونی استراتژیک) را ندارد. درحالیکه روش پیشنهادی RTRA به عنوان یک روش یادگیری چند-عامله امکان تطبیق با عدم قطعیت درونی (استراتژیک و چندعامله) را دارا می‌باشد و یک متوسط‌گیری وزن‌دار

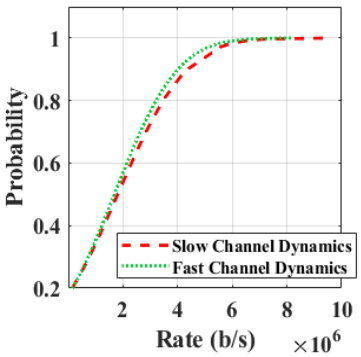
نمایی در روال تقریب تصادفی خود انجام می‌دهد و چابکی خوبی برای تطبیق با شرایط جدید را دارد.

با توجه به شکل ۸ برای توزیع تجمعی سودمندی‌ها، اگر خطی به موازات محور افقی روی عدد ۰/۶-۱یعنی ۰/۴ ترسیم کنیم تا نمودار را قطع کرده و سپس نقطه تقاطع را روی محور افقی تصویر کنیم، مقدار ۱/۳ را نشان می‌دهد، که به این معناست که در روش پیشنهادی، ۶۰ درصد از عامل‌ها توانسته‌اند مقدار سودمندی بیشتر از ۱/۳ مگابیت در ثانیه را تجربه کنند، در حالی که سایر روش ها مقدار کمتری را تجربه کرده اند. با توجه به اینکه روش RTRA در مقایسه با سایر روش‌ها جابه‌جایی به راست (یعنی: تحقق مقادیر بالاتر سودمندی با احتمال بیشتر) قابل توجهی داشته و همچنین از نظر دستیابی به بالاترین سودمندی مشاهده شده نیز دست بالاتر را دارد، برتری محسوسی نسبت به روش‌های مقایسه شده داشته است.




**شکل ۸- نمودار توزیع تجمعی تجربی سودمندی عامل‌ها**

مشابه نمودار شکل ۸، شکل ۹ نمودار توزیع تجمعی را به گونه‌ای نمایش می‌دهد که روش پیشنهادی ما تحت محوشدگی کُند و تند مورد ارزیابی قرار گرفته باشد. رویکرد ما در محیطی تحت محوشوندگی کُند در مقایسه با محیطی تحت محو شوندگی تند، ضمن دستیابی به بالاترین سطح سودمندی حاصل شده، میزانی جابه‌جایی به راست را نیز نشان می‌دهد و ۵۵٪ از عامل‌ها توانسته‌اند میزان سودمندی بیشتر از ۱/۵ مگابیت در ثانیه را تجربه کنند که مقدار فوق مقدار بالاتری از محیط تحت محوشوندگی تند را نشان می‌دهند.




**شکل ۹- مقایسه نمودار ECDF از سودمندی عامل‌ها، تحت محوشوندگی تند و آهسته**

## ۶- نتیجه‌گیری

در مقایسه با معماری‌های سنتی شبکه‌های دسترسی رادیویی، معماری جدید مشهور به O-RAN ویژگی هایی مانند باز بودن، جداسازی اجزا و کنترل‌کننده های هوشمند را فرآهم می کند. با این حال انعطاف‌پذیری O-RAN باعث ایجاد مشخصات و معماری پیچیده شده است و با رویکردهای سنتی نمی‌توان به درستی با آن مواجه شد. یکی از چالش‌های مهم در حوزه تخصیص منابع O-RAN، عدم قطعیت کانال ارتباطی است.

در این مقاله ما مسئله تخصیص توأم O-RUها به کاربران مستأجرهای یک زیرساخت O-RAN و زیرکانال به O-RUها را در شرایط کیفیت کانال غیرقطعی و متغیر با زمان در نظر گرفتیم. با توجه به محوشوندگی کُند شرایط کانال، سودمندی عامل‌ها به صورت وابسته به وضعیت تغییر می‌کنند. از این رو، ما مسئله تخصیص منابع را به صورت یک بازی مُدوله شده مارکُفی مدل کرده‌ایم که در آن محاسبه تعادل بازی نیاز به رهگیری دارد. به همین منظور یک الگوریتم یادگیری چند-عامله پیشنهاد دادیم که با اجرای مکرر آن، رفتار جمعی تخصیص منابع O-RUها به طور مجانبی به مفهوم تعادل همبسته همگرا می‌شود. الگوریتم پیشنهادی قابلیت رهگیری وضعیت تعادل با تغییرات تصادفی کانال در طی زمان را دارا می‌باشد. با استفاده از شبیه‌سازی، برتری راهکار پیشنهادی به لحاظ میزان سودمندی و توزیع تجمعی سودمندی را در مقایسه با چندین الگوریتم یادگیری تک-عامله و چند عامله دیگر نشان می‌دهد. به عنوان یکی از محورهای گسترش روش پیشنهادی در آینده می‌توان مسئله تخصیص منابع در O-RAN را تحت محوشوندگی سریع کانال در نظر گرفت.

وقتی در شبکه سلولی چندین عملگر وجود دارند، که به صورت اشتراکی از زیرساخت شامل زیرکانال و واحدهای رادیویی، استفاده میکنند و هر عملگر به کاربران متعددی خدمات ارائه میدهد، ممکن است تضاد مناقع بین عملگرها در تخصیص منابع پیش آید. امکان گسترش کار پیشنهادی از طریق انجام آزمایش‌ها و مراحل اعتبارسنجی جهت ارزیابی مقاومت سیستم در برابر تعارضات احتمالی میان xAppهای مستقر شده توسط عملگرهای مختلف وجود دارد و باید به بررسی استراتژی‌های مواجهه با این تعارضات پرداخت.

## مراجع

[1] M. Polese, L. Bonati, S. D'Oro, S. Basagni and T. Melodia, "Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges," in IEEE Communications Surveys & Tutorials, vol. 25, no. 2, pp. 1376-1411, 2023.

[2] M. Sharara, T. Pamuklu, S. Hoteit, V. Vèque and M. Erol-Kantarci, "Policy-Gradient-Based Reinforcement Learning for Computing Resources Allocation in O-RAN," 2022 IEEE 11th International Conference on Cloud Networking (CloudNet), pp. 229-236, 2022.

[3] S. Mollahasani, M. Erol-Kantarci and R. Wilson, "Dynamic CU-DU Selection for Resource Allocation in O-RAN Using Actor-Critic Learning," 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, pp. 1-6, 2021.

[4] N. Hammami and K. K. Nguyen, "Multi-Agent Actor-Critic for Cooperative Resource Allocation in Vehicular Networks," 2022 14th IFIP Wireless and Mobile Networking Conference (WMNC), Sousse, Tunisia, pp. 93-100, 2022.

[5] N. Kazemifard and V. Shah-Mansouri, "Minimum delay function placement and resource allocation for Open RAN (O-RAN) 5G networks," Computer Networks, vol. 188. Elsevier BV, p. 107809, 2021.

[6] C.-H. Lai, L.-H. Shen, and K.-T. Feng, "Intelligent Load Balancing and Resource Allocation in O-RAN: A Multi-Agent Multi-Armed Bandit Approach." arXiv, 2023.

[7] O. Narmanlioglu and E. Zeydan, "Learning in SDN-based multi-tenant cellular networks: A game-theoretic perspective," 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM), pp. 929-934, 2017.

[8] H. Zhou, M. Elsayed and M. Erol-Kantarci, "RAN Resource Slicing in 5G Using Multi-Agent Correlated Q-Learning," 2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Helsinki, Finland, pp. 1179-1184, 2021.

[9] Bejani, R. , Kalantari, M. and Eftekhari Moghaddam, A. M., "A Game Theory Framework to Cooperate Nodes in Malicious Nodes

Detection Process in Wireless Sensor Network," TABRIZ JOURNAL OF ELECTRICAL ENGINEERING, vol.47, no.4, pp. 1329-1342, 2018.

[10] A. Filali, Z. Mlika and S. Cherkaoui, "Open RAN Slicing for MVNOs With Deep Reinforcement Learning," in IEEE Internet of Things Journal, vol. 11, no. 10, pp. 18711-18725, 2024.

[11] Hakami, V. , Mostafavi, S. A. and Arefinezhad, Z., "RL-based Resource Allocation for Improving Throughput in Cellular D2D Communications, " TABRIZ JOURNAL OF ELECTRICAL ENGINEERING, pp. 50, no. 3, pp. 1165-1177, 2020.

[12] R. Zhang and L. Cai, "Packet-Level Channel Model for Wireless OFDM Systems," IEEE GLOBECOM 2007 - IEEE Global Telecommunications Conference, Washington, DC, USA, pp. 5215-5219, 2007.

[13] D. Rastovac and D. Vukobratovic, "Modeling LTE-A channel using FSMC model for OFDM systems," 2012 20th Telecommunications Forum (TELFOR), Belgrade, Serbia, pp. 436-439, 2012.

[14] Y. Xu and Z. Cao, "A Packet Loss Reduction Scheduling Scheme With Cross-layer Design for OFDM Downlinks," MILCOM 2006 - 2006 IEEE Military Communications conference, Washington, DC, USA, pp. 1-7, 2006.

[15] Heng Wang and N. B. Mandayam, "A simple packet-transmission scheme for wireless data over fading channels," in IEEE Transactions on Communications, vol. 52, no. 7, pp. 1055-1059, 2004.

[16] Hong Shen Wang and N. Moayeri, "Finite-state Markov channel-a useful model for radio communication channels," in IEEE Transactions on Vehicular Technology, vol. 44, no. 1, pp. 163-171, Feb. 1995.

[17] R. J. Aumann, “Correlated Equilibrium as an Expression of Bayesian Rationality,” Econometrica, vol. 55, no. 1. JSTOR, p. 1, Jan. 1987.

[18] S. Hart and A. Mas-Colell, “A Reinforcement Procedure Leading to Correlated Equilibrium,” Economics Essays. Springer Berlin Heidelberg, pp. 181–200, 2001.

[19] O. Namvar Gharehshiran, V. Krishnamurthy and G. Yin, "Distributed Tracking of Correlated Equilibria in Regime Switching Noncooperative Games," in IEEE Transactions on Automatic Control, vol. 58, no. 10, pp. 2435-2450, 2013.

[20] V. Todisco, S. Bartoletti, C. Campolo, A. Molinaro, A. O. Berthet and A. Bazzi, "Performance Analysis of Sidelink 5G-V2X Mode 2 Through an Open-Source Simulator," in IEEE Access, vol. 9, pp. 145648-145661, 2021.

[21] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, MIT Press, 2018.

[22] W. H. Sandholm, "Population Games and Evolutionary Dynamics, ser, " Economic Learning and Social Evolution. Cambridge, MA, USA: MIT Press, 2011.