علوم رایانش و فناوری اطلاعات
نشریه علمی انجمن کامپیوتر ایران
مجلد ۱۷، شماره ۲، ۱۳۹۸
صفحات ۸۲- ۷۲
مقاله عادی

کنترل توأم فشرده‌سازی و ارسال داده‌ها در تجهیزات اینترنت اشیاء با انرژی تجدیدپذیر

فرنوش نامجونیا[1*]، وصال حکمی[2]
*نویسنده مسئول، دریافت: ۹۸/۱۲/۲۹، بازنگری: ۹۹/۰۲/۱۴، پذیرش: ۹۹/۰۳/۰۲
[1] کارشناسی ارشد، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران
[2] استادیار، گروه شبکه‌های کامپیوتری، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران

چکیده

یکی از مهم‌ترین چالش‌های توسعهٔ اینترنت اشیاء، محدودیت انرژی تجهیزات است. در راستای کاهش مصرف انرژی، در این مقاله، ما مسئله کنترل توأم نرخ فشرده‌سازی (با اتلاف) و تعداد بسته‌های ارسالی در واحد زمان را برای یک گره اینترنت اشیاء مجهز به منبع انرژی تجدیدپذیر مطرح می‌کنیم. نوآوری راهکار پیشنهادی در توجه هم‌زمان به دو هدف بهینه‌سازی یعنی: «سطح تطابق» داده‌های دریافتی با داده‌های اصلی و نیز رعایت قید تأخیر ارسال داده‌هاست. برای این منظور، با استفاده از چارچوب ریاضی فرآیند تصمیم مارکُفی مقیّد، مسئله را در قالب یک بهینه‌سازی تصادفی طرح می‌کنیم با هدف بیشینه کردن متوسط «سطح تطابق» داده‌ها در بلندمدت، ضمن ایجاد محدودیت در متوسط تأخیرِ گزارش رویدادهای حسگری. نامقیّدسازی مسئله با روش استاندارد «لاگرانژین» انجام می‌شود. الگوریتم پیشنهادی ما برای محاسبهٔ سیاست بهینهٔ تطبیق‌پذیر نیز بر مبنای دو تکنیک یادگیری تقویتی سریع به نام PDS و VE است که می‌تواند با جداسازی پویایی سیستم به دو بخش قطعی و تصادفی، صرفاً با اتخاذ تصمیمات حریصانه و بدون نیاز به دانش آماری فرآیندهای تصادفیِ کانالِ بی‌سیم، شارژ انرژی و وقوع رویدادهای حسگری، همگرایی به سیاست بهینه را تضمین نماید. کارایی سیاست‌های پیشنهادی با الگوریتم استاندارد Q-learning مورد مقایسه قرارگرفته و به لحاظ مصرف انرژی،میزان هدر رفت بسته‌های داده و هم‌چنین «سطح تطابق» داده‌های گزارش‌شده ارزیابی می‌شوند. نتایج نشان می‌دهند که سطح تطابق داده‌های گزارش‌شده در روش VE نسبت به روش استاندارد Q-learning به میزان ۶۳٫۷۴۱ درصد و روش PDS نسبت به روش استاندارد Q-learning میزان ۶۱٫۸۴۵ درصد بهبود یافته است.

کلمات کلیدی: اینترنت اشیاء، بهینه‌سازی انرژی، برداشت انرژی، تطابق داده‌ها، فرآیند تصمیم‌گیری مارکُف مقیّد، فشرده‌سازی، محدودیت تأخیر، یادگیری تقویتی PDS، یادگیری تقویتی VE.

۱- مقدمه

در سال‌های اخیر، اینترنت با اتصال میلیون‌ها شیء (با اندازه، قابلیت، قدرت پردازشی و محاسباتی متفاوت) در سطح جهانی به‌طور شگرفی تکامل یافته است و در ادامه این روند، اینترنت سنتی در شبکه‌ای فراگیرتر به نام اینترنت اشیاء ادغام می‌شود. در اینترنت اشیاء، اتصال تجهیزات حسگر توسط یک زیرساخت ارتباطی فراهم‌شده و داده‌های آن‌ها توسط یک واحد پردازش (معمولاً روی ابر)، مورد تحلیل، تصمیم‌گیری و اقدام قرار می‌گیرد [۱].

در توسعه اینترنت اشیاء چالش‌های متعددی وجود دارد: تجهیزات اینترنت اشیاء معمولاً متکی به باتری‌های کوچک با ظرفیت محدود هستند و تنها برخی از آن‌ها به شبکه برق متصل می‌باشند. به‌عنوان‌مثال، در کاربردهایی که از گره‌های اینترنت اشیاء برای جمع‌آوری داده‌های زیست‌محیطی استفاده می‌شود، بسیاری از گره‌ها در مکان‌های سخت یا دوردست مستقرشده‌اند. ازاین‌رو، هزینه تعویض باتری یا جایگزینی گره‌ها ممکن است سنگین‌تر از مزیت جمع‌آوری داده توسط آن‌ها باشد. از طرفی از بین رفتن گره به علت اتمام انرژی ممکن است منجر به فروپاشی کل معماری شود. ازآنجایی‌که انتظار می‌رود گره‌ها برای مدتی طولانی بدون جایگزینی باتری و به‌طور مستقل سرویس دهند، مسئله صرفه‌جویی و بهره‌وری انرژی یک نیاز کلیدی برای اینترنت اشیاء می‌باشد [۲،۳].

اخیراً، برای غلبه بر محدودیت باتری یا کاهش میزان مصرف توان از شبکهٔ برق، حسگرهایی با قابلیت «برداشت انرژی» از محیط در قالب انرژی خورشیدی، حرارتی، ارتعاشی یا رادیویی پیشنهاد شده‌اند که در حال حاضر به‌صورت تجاری در دسترس می‌باشند. ازاین‌رو پژوهش‌های متعددی برای گره‌ها علاوه بر باتری، قابلیت برداشت انرژی از محیط را در نظر گرفته‌اند [۴،۵]؛ بنابراین، دیگر نیازی به صرفه‌جویی انرژی در تمام طول عمر تجهیزات وجود ندارد اما باید سیاست‌های بهینه به‌صورت مدبرانه

انرژی محدودِ برداشت‌شده از محیط را در راستای تأمین نیازهای گره هزینه کنند [۶].

همچنین، استفاده از کُدگذاری تطبیق‌پذیر منبع با کاهش تعداد بیت‌هایی که باید منتقل شوند هزینه انرژی ارتباطی را به‌طور مؤثر کاهش و عمر شبکه را افزایش می‌دهد. در اغلب موارد، صرفه‌جویی انرژی حاصل از فشرده‌سازی قابل‌توجه است، چون با کاهش ازدحام سطح لینک تعداد «برخورد» و تلاش مجدد در شبکه نیز کاهش می‌یابد [۷].

تاکنون پژوهش‌های متعددی از فشرده‌سازی با اتلاف یا بدون اتلاف برای صرفه‌جویی در مصرف انرژی استفاده کرده‌اند. برخی از پژوهش‌ها از حداکثر سطح فشرده‌سازی برای کاهش حجم داده‌ها پیش از ارسال استفاده نموده‌اند که معمولاً برای کاربردهای ساده مانند حس کردن دمای محیط و گزارش آن کاربرد دارد [۳،۴]. اما در کاربردهای پیشرفته‌تر با جریان داده سنگین‌تر، عملیات فشرده‌سازی خود مستلزم انجام محاسبات پرهزینه برای پردازش داده‌ها خواهد بود و درنتیجه هزینه انرژیِ مصرفیِ فشرده‌سازی می‌تواند با هزینه‌های ارتباطی برابری کند. بنابراین ایجاد یک موازنه کارآمد میان پردازش داده‌ها (با هدف کاهش حجم) و توان مصرفی برای ارسال می‌تواند منجر به صرفه‌جویی قابل‌توجهی در مصرف انرژی شود[۳،۶،۸].

در یک شبکه متکی به جریان انرژی تجدیدپذیر با ماهیت تصادفی، ایجاد موازنه میان عملیات فشرده‌سازی و ارسال داده‌ها، تحت تأثیر فرآیند تصادفی شارژ انرژی از یک‌سو و کیفیتِ متغیر با زمانِ کانالِ بی‌سیم و وقوع غیرقطعی رویدادهای محیطی از سوی دیگر، دچار چالش می‌شود. درواقع، برای بهینه‌سازی متوسط کارایی شبکه در درازمدت نیازمند محاسبه یک سیاست کنترلی تطبیق‌پذیر با وضعیت جاری سیستم خواهیم بود. برای محاسبه چنین سیاستی، روش‌های مطرح در پژوهش‌های گذشته را می‌توان برحسب فرض اطلاعاتی به کار گرفته‌شده در آن‌ها به سه دسته تقسیم کرد که در زیر بخش بعدی به مرور آن‌ها می‌پردازیم.

۱-۱- مرور کارهای پیشین

در این بخش، بر اساس نوع فرض اطلاعاتی، کارهای پیشین را در حوزه کنترل فشرده‌سازی و ارسال داده‌ها در شبکه‌های حسگر بی‌سیم و اینترنت اشیاء دسته‌بندی و مرور می‌نماییم:

فرض اطلاعاتی برون خط: در این دسته از پژوهش‌ها، بهینه‌سازی با این فرض انجام می‌گیرد که اطلاعات دقیق و کامل مربوط به خط سیر فرآیندهای تصادفی وضعیت کانال، وقوع رویدادها و شارژ انرژی از پیش در اختیار است.

برای مثال، در [۹]، یک طرح برای مدیریت سیاست‌های انرژی برای گره‌های حسگر چندرسانه‌ای بی‌سیم در نظرگرفته شده است که کاملاً به برداشت انرژی از محیط برای جمع‌آوری و انتقال داده متکی است. یک گره در شبکه چندرسانه‌ای حسگر بی‌سیم که از یک منبع توزیع‌شده به‌صورت تصادفی نمونه‌برداری، اندازه‌گیری و فشرده‌سازی می‌کند و بیت‌های داده را بر روی یک کانال بی‌سیم به سمت مقصد ارسال می‌کند، در نظر گرفته می‌شود. زمان با فاصله زمان‌بندی برهه‌های T ثانیه است، فرض می‌شود که واریانس منبع و ورودی‌های انرژی در حافظه زمانی بدون تغییر باقی می‌ماند و در میان برهه‌های زمانی مختلف تقریباً متفاوت است. سیستم در ابتدای هر برهه زمانی، انرژی را در بسته‌هایی با مقدار تصادفی جمع‌آوری می‌کند. داده‌های حس شده از محیط دارای یک محدودیت تأخیر سخت به‌اندازه یک برهه است هم‌چنین در این کار فرض شده است که کدگذاری تطبیق‌پذیر منبع و انتقال داده‌ها به‌صورت متوالی انجام می‌شود به این معنا که در این کار بافر برای ورود داده لحاظ نشده است. مسئله به حداقل رساندن اعوجاج در یک محیط برون خط قرار داده‌شده که در آن فرض شده است که واریانس منبع، ورودی‌های انرژی و دستاورد-های کانال شناخته‌شده است و بر اساس ماهیت راه‌حل بهینه سه الگوریتم ارائه می‌دهد. هدف این کار تشخیص نرخ فشرده‌سازی بهینه و انرژی تخصیص داده‌شده به فشرده‌سازی و انتقال در تنظیمات برون خط که بستگی به انرژی و توان فشرده-سازی دارد که میانگین اعوجاج را حداقل کند.

به‌عنوان نمونه دیگر، در [۱۰]، یک سیستم زمان‌بند در نظر گرفته‌شده که در آن داده‌های جدید و بسته‌های انرژی در ابتدای هر برهه زمانی (TS) وارد می‌شوند و انتظار می‌رود که بهره کانال در طول هر برهه زمانی ثابت باشد. بر اساس این فرض که ورود انرژی و داده‌ها و دستاوردهای کانال از پیش به‌صورت برون‌خط معلوم است، مسئله بهینه‌سازی برنامه‌ریزی فشرده‌سازی و انتقال به‌عنوان یک مسئله بهینه‌سازی محدّب فرمول‌بندی شده است.

فرض اطلاعاتی برخط: درصورتی‌که دانش آماری (توزیع‌های احتمال) مربوط به فرآیند شارژ انرژی، تغییرات کانال بی‌سیم و وقوع رویدادهای محیطی از پیش در اختیار باشد و اجرای الگوریتم تنها نیازمند وقوع رویدادهای تصادفی (به‌عنوان ورودی سیاست از پیش محاسبه‌شده) باشد، از آن به‌عنوان برخط یاد می‌شود.

مثلاً در [۸]، یک استراتژی زمان‌بندی (مبتنی بر تقسیم زمان دسترسی چندگانه) برای تجهیزات اینترنت اشیاء متکی به باتری ارائه‌شده است که در آن به‌طور توأم، نرخ بهینهٔ فشرده‌سازی داده‌ها و تخصیص انرژی برای هر گره در هر برهه زمانی تعیین می‌گردد. هم‌چنین در این کار فرض شده است که دانش آماریِ فرآیند تصادفیِ کانال در اختیار می‌باشد و به‌این‌ترتیب، جواب بهینه با استفاده از برنامه‌ریزی ریاضی محاسبه‌شده است و بافری برای ورود داده‌ها در آن در نظر گرفته نشده است. با توجه به اینکه در [۸]، فرض بر استفاده از فشرده‌سازیِ «با اتلاف» می‌باشد، هدف آن حداقل سازی «اعوجاج» داده‌ها ضمن تأمین یک قید معیّن برای طول عمر تجهیزات بوده است.

در [۱۱]، یک استراتژی بهینه فشرده‌سازی-انتقال پویا برای شبکه‌های حسگر چندگامه[۱] با قابلیت «برداشت انرژی» پیشنهاد شده است. این استراتژی بر اساس انرژیِ برداشت‌شده، وضعیت کانال، صف داده، بافر انرژی و هم‌چنین مدل آماری از هم‌بستگی منبع، تصمیم بهینه را انتخاب می‌کند و فرض بر آن است که به دانش آماری وضعیت‌ها وابسته می‌باشد. در این کار، داده‌های جمع‌آوری‌شده از محیط در هر برهه قبل از انتقال، با یک کدگذاری منبع تطبیق‌پذیر،تحت همبستگی مکانی داده اندازه‌گیری شده، فشرده می‌شوند و هم‌چنین در این کار بافر داده لحاظ نشده است. هدف از بهینه‌سازی، به حداکثر رساندن قابلیتِ بازسازی سیگنال با تضمین قید کفایت انرژی شبکه و اطمینان از پایداری صف می‌باشد. هم‌چنین، اثبات می‌شود که سیاست پیشنهادشده با یک موازنه قابل‌کنترل میان اندازه‌های صف‌ها و باتری‌ها متوسط هزینه شبکه را نزدیک به بهینگی به دست می‌آورد.

در [۱۲] ، یک سیاست مدیریت انرژی برای شبکه حسگر بی‌سیم با قابلیت «برداشت انرژی» مطرح شده است، که درآن‌واحدِ مدیریت انرژی باید بر اساس آمار روند برداشت انرژی، کیفیت داده اندازه‌گیری شده، نرخ سیگنال به نویز کانال و اندازه صف داده، انرژی را به واحد کسب منبع (اندازه‌گیری، نمونه‌برداری، فشرده‌سازی) و انتقال داده اختصاص دهد. ازآنجایی‌که فشرده‌سازی «با اتلاف» فرض شده است، هدف، ایجاد یک توازن بهینه میان سطح اعوجاج داده و تأخیر انتقال می‌باشد. در این کار ابتدا بر روی یک سیستم با یک گیرنده و یک حسگر را در نظر می‌گیرد به دلیل اینکه تمرکز روی جنبه اصلی مسئله باشد. حسگر به یک باتری مجهز شده که این قابلیت را دارد که انرژی برداشت‌شده از محیط را ذخیره کند و در هر برهه زمانی حسگر توالی زمان را برای پدیده موردعلاقه به دست می‌آورد که توسط یک نسبت سیگنال به نویز اندازه‌گیری و ارتباط خودکار برقرار می‌شود و بیت‌های حاصل را بعد از فشرده‌سازی احتمالی ذخیره می‌کند و به صف داده می‌رساند . بر اساس آمار فرآیند برداشت انرژی و بر اساس وضعیت فعلی کیفیت داده‌های اندازه‌گیری شده، SNR کانال و صف داده، واحد مدیریت انرژی باید تخصیص انرژی را بین دریافت داده‌ها و انتقال آن‌ها انجام دهد تا به‌طور مطلوب بین الزامات رقابتی اعوجاج در بازسازی داده‌ها در گیرنده، ثبات صف و تأخیر تعادل برقرار شود. این پژوهش مسئله

بهینه‌سازی را با موازنه مطلوب میان متوسط اندازه داده جمع‌آوری‌شده برای فشرده‌سازی و متوسط اعوجاج سیگنال فرموله و آن را با استفاده از برنامه‌نویسی پویا حل کرده است.

در [۱۳]، مسئله طراحی سیاست‌های مؤثر برای پردازش و انتقال داده‌ها را به‌طور توأم بررسی می‌کند. در این کار N منبع همگون و متجانس در نظر گرفته‌شده است که به‌صورت بی‌سیم داده‌ها رو به یک ایستگاه مرکزی ارسال می‌کنند کاربران به‌صورت تسهیم زمان پویا به کانال دسترسی دارند و زمان به برهه‌هایی تقسیم می‌شود. در این سناریو، هر گره به‌صورت دوره‌ای داده‌ها را تولید می‌کند و تصمیم می‌گیرد که چه مقدار بر روی آن‌ها فشرده‌سازی انجام شود و درنهایت به یک فرستنده معمولی انتقال دهند و بافر داده نیز در این کار در نظر گرفته نشده است. هدف در این کار این است که یک نقطه عملیاتی بهینه تعیین شود که در آن بین افزایش طول عمر شبکه و تضمین اعوجاج کم در طول انتقال داده به‌منظور ایجاد یک استراتژی زمان‌بند انتقال مبتنی بر تسهیم زمان پویا که به‌صورت بهینه منابع را تخصیص می‌دهد، یک موازنه برقرار شود. هم‌چنین در این کار از یک منحنی «نرخ-اعوجاج» تصادفی استفاده می‌شود تا تأثیر فشرده‌سازی را بر روی کیفیت داده‌های ارسالی اندازه‌گیری شود. در این کار، فرض شده تجهیزات «IoT» مبتنی بر باتری هستند و قابلیت برداشت انرژی از محیط را ندارند و بنابراین با محدودیت انرژی مواجه می‌شوند. سطح اولیه از انرژی باتری که در دسترس است بر روی عملکرد سیستم تأثیر به سزایی دارد.

در [۱۴] ، به مسئله طراحی سیاست‌های کارآمد برای انجام فرایند پردازش و انتقال داده‌ها به‌صورت توأم پرداخته می‌شود. به‌طور خاص، هدف این کار تعریف استراتژی برنامه‌ریزی با هدف دوگانه گسترش طول عمر شبکه و تضمین یک اعوجاج کلی در مورد داده‌های منتقل‌شده است. هم‌چنین، یک الگوریتم مبتنی بر دسترسی زمانی پیشنهاد شده که به‌طور مؤثر منابع را به گره‌های ناهمگن اختصاص می‌دهد. از منحنی‌های نرخ واقعی اعوجاج برای کم کردن تأثیر فشرده‌سازی بر کیفیت داده استفاده می‌کند و یک مدل کامل انرژی ارائه می‌دهد که شامل انرژی صرف شده برای پردازش و انتقال داده‌ها می‌شود. هر دو دانش کامل و آماری از کانال‌های بی‌سیم در نظر گرفته می‌شود.

فرض اطلاعاتی مبتنی بر یادگیری: در شرایطی که دانش آماری از شرایط تصادفی محیط عملیاتی در زمان طراحی در اختیار نیست، ناگزیر از ارائه راهکارهای مبتنی بر یادگیری ماشین برای محاسبه سیاست موازنه بهینه هستیم. استفاده از تکنیک‌های یادگیری از اهمیت ویژه‌ای برخوردار است چراکه در بسیاری از سناریوهای کاربردی، امکان مدل‌سازی دقیق ساختار احتمالاتی سیستم میسّر 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 نمی‌باشد.

در [۱۵]، در مورد چگونگی ترکیب بهره‌وری مؤثر از انرژی با الزامات کیفیت سرویس و یا در اصطلاح کیفیت داده‌های گزارش‌شده بررسی شده است. در این کار یک سیستم نظارت در نظر گرفته‌شده که گره‌های حسگر در توپولوژی چندگامه، داده‌های خوانده‌شده را به یک گیرنده مشترک در شبکه گزارش می‌کنند. هدف هر دستگاه استفاده از انرژی موجود است که با استفاده از این انرژی اعتبار بازسازی داده‌ها را بیشینه کند که بستگی به اعوجاج ارائه‌شده توسط الگوریتم فشرده‌سازی با اتلاف و ناپایداری‌های کانال دارد. برای این منظور، مکانیسم کدگذاری کانال منبع که شامل انتقال داده‌های جدید و انتقال مجدد داده‌هایی که به‌صورت موفق دریافت نشده‌اند، توسعه داده شده است.

در [۱۶]، یک سناریو به این صورت در نظر گرفته شده است که یک حسگر بصری برداشت انرژی، بدون داشتن دانش قبل از ورود انرژی و ترافیک‌های ورودی و پویایی کانال می‌خواهد که انرژی ذخیره‌شده‌اش را به حداکثر برساند و محدودیت تأخیر صفر داشته باشد. به‌منظور صرفه‌جویی در انرژی به‌طور مشترک لایه فیزیکی و مصرف انرژی در سطح سیستم را با استفاده از کنترل توان هم تأخیر را در بهینه-سازی در نظر می‌گیرد در این کار نه‌تنها مدل وضعیت باتری، وضعیت ترافیک صف که مشکل را به‌طور قابل‌ملاحظه‌ای پیچیده‌تر می‌کند، منظور شده است.

در این کار چالش انرژی برای انتقال اطلاعات تصویری حساس به تأخیر یک حسگر از راه دور طی یک کانال متغیر بازمان در نظر گرفته شده است. حسگر در این کار انرژی را از محیط برداشت می‌کند و ازاین‌رو مصرف کارآمد انرژی از اهمیت زیادی برخوردار است. هدف دست‌یابی به زمان‌بندی انتقال و سیاست‌های مدیریت توان به این منظور که انرژی برای انتقال‌های آینده موجود باشد. این در حالی است که با محدودیت تأخیر صف نیز مواجه است. برای حل این مشکل پیچیده برخط که دارای فضای حالت بسیار بزرگی است، حاوی تمام ترکیبات، وضعیت صف ترافیک، وضعیت باتری، وضعیت مدیریت توان و وضعیت کانال است. این مسئله را با استفاده از فرایند تصمیم مارکُف MDP مدل‌سازی می‌کند و نیز یک الگوریتم یادگیری تقویتی برای حل برخط آن ارائه می‌دهد.

در[۱۷]، یک کاربرد نظارت محیطی در نظر گرفته‌شده که در آن گره‌های حسگر توسط منابع انرژی تجدید پذیر به‌صورت دوره‌ای اندازه‌گیری‌های خود را به نقطه جمع‌آوری‌کننده داده‌ها می‌رسانند. قبل از انتقال، عملیات پردازش را انجام می‌دهند (مثلاً فشرده‌سازی با اتلاف همراه با کدگذاری کانال منبع) که اعتبار داده‌های بازسازی‌شده در گیرنده را تحت محدودیت فرایند برداشت انرژی به حداکثر می‌رساند. به‌طور ویژه اعتبار داده‌های بازسازی‌شده به موازنه بین دقت فشرده‌سازی منبع و مقاومت برعلیه ناپایداری‌های کانال بستگی دارد،که این مفهوم در اصطلاح با اعوجاج داده‌ها بیان می‌شود. در این کار مسئله بیشینه کردن کیفیت داده‌های گزارش‌شده تحت محدودیت برداشت انرژی و انرژی مصرفی و ناپایداری کانال ارتباطی بررسی می‌شود. نقطه اعوجاج مطلوب بر اساس یک سیاست کدگذاری کانال- منبع انتخاب می‌شود: گره باید تعداد بیت‌ها را برای انتقال انتخاب کند که به این معنی است که برای تعیین میزان فشرده‌سازی با اتلاف در منبع و نرخ کدگذاری اصلاح خطا تصمیم-گیری شود. مسئله مطرح‌شده از طریق تجزیه، به دو فرایند بهینه‌سازی توزیع‌شده تقسیم شده است. مسئله بیرونی به‌عنوان یک مسئله MDP مدل می‌شود. مسئله درونی آن‌هم طرح‌های کدگذاری کانال و منبع را به‌صورت هم‌زمان بهینه می‌کند تا مطمئن شود که کیفیت داده‌های دریافتی در گیرنده بیشینه است.

در [۶]، یک الگوریتم طراحی شده تا سیاست‌های فشرده‌سازی- انتقالِ بهینه را از طریق یک روش تخفیف لاگرانژی همراه با جستجوی دودویی برای ضریب لاگرانژ، محاسبه نماید. در این کار، پویایی انرژی و انتقال داده و انجام فشرده‌سازیِ با اتلاف به‌صورت مسئله تصمیم‌گیری مارکُف محدود اجرا می‌شود. هدف بهینه‌سازی، به حداکثر رساندن دقت بازسازی در مقصد ضمن تأمین پایداری بافر انرژی می‌باشد و بدین منظور سطح فشرده‌سازی به‌صورت پویا بر اساس فرآیند تصادفی ورود انرژی، وضعیت بافر انرژی و کانال، انتخاب می‌شود. برای فشرده‌سازی، از الگوریتم فشرده-سازی زمانیِ با اتلاف LTC^{2} استفاده کرده است که نشان داده شده موازنهٔ خوبی را ازلحاظ اعوجاج و مصرف انرژی ارائه می‌دهد.

۱-۲- انگیزه ارائه راهکار جدید و نوآوری‌ها

در پژوهش‌های پیشین از میان این سه رویکرد شرح داده شده، رویکرد مبتنی بر یادگیری کاربردی‌تر و واقع‌بینانه‌تر است. ما نیز در این مقاله، از رویکرد مبتنی بر یادگیری بهره می‌گیریم. به بیان مشخص‌تر، ابتدا یک فرمول‌بندی رسمی از مسئله کنترل توأم فشرده‌سازی ارسال در قالب یک مسئله بهینه‌سازی تصادفی ارائه می-دهیم. فرمول‌بندی ارائه شده و راهکار پیشنهادی برای حل مسئله نسبت به کارهای پیشین از چند جهت دارای نوآوری است:

- اول اینکه، برخلاف روش‌های پیشین که تنها متمرکز بر محاسبه سیاست بهینه کُدگذاری منبع بوده‌اند، فرمول‌بندی پیشنهادی در این مقاله به‌طور توأم، کنترل ارسال‌ها را نیز در برمی‌گیرد. این امر با اعطای درجه آزادی بالاتر به گره باعث

تطبیق‌پذیری بیشتر با محیط تصادفی شده و امکان بهره‌وری انرژی و دستیابی به «سطح تطابق» بالاتر را ارتقاء می‌دهد.

- دوم این‌که، فرمول‌بندی ما تنها معطوف به بهینه‌سازی یک تابع هدف (مثل: انرژی مصرفی، اعوجاج، «سطح تطابق» و غیره) نیست بلکه علاوه بر بهینه-سازی «سطح تطابق» با تأمین یک قید متوسط تأخیر از معطل ماندن بیش از حدّ داده‌های حسگری حساس به بهانه دستیابی به فرصت‌های فشرده-سازی و ارسال بهتر جلوگیری می‌کند. بعلاوه، میزان هدر رفت بسته‌ها بابت سرریز فضای بافر داده نیز به‌عنوان معیار کارایی دیگر در تابع هدف سیستم دخالت داده می‌شود تا عملکرد گره به‌صورت یک‌جانبه‌گرایانه به افزایش «سطح تطابق» سوق داده نشود.
- سومین نوآوری ما مربوط به ماهیت الگوریتم یادگیری پیشنهادی برای محاسبه سیاست بهینه کنترل توأم فشرده‌سازی/ ارسال است. درواقع، با توجه به اهمیت سرعت همگرایی و محاسبه جواب بهینه در کاربردهای حساس مبتنی بر اینترنت اشیاء، الگوریتم‌های پیشنهادی در این مقاله بر مبنای دو تکنیک مهم یادگیری تقویتی سریع به نام‌های PDS[۳] و VE[۴] برای تقریب سیاست کنترلی بهینه طراحی می‌شود [۱۸]. الگوریتم‌های پیشنهادی به دلیل بهره‌برداری از قابلیت جداسازی پویایی سیستم به دو بخش معلوم و نامعلوم و همچنین انجام به‌روزرسانی‌های فرصت‌طلبانه «دسته‌ای» قادرند در تعداد تکرار کمتر و با پیچیدگی نمونه‌برداری پایین‌تر نسبت به روش‌های استاندارد یادگیری تقویتی (مثلاً: Q-learning) همگرایی به سیاست بهینه کنترلی را حاصل نمایند [۱۹].
- عملکرد روش‌های پیشنهادی با انجام آزمایش‌های عددی تحت سناریوهای مختلف و بر مبنای معیارهای کارایی متفاوت مورد ارزیابی قرار می‌گیرد.

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 فرایند مارکُف 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 انجام شده خواهیم پرداخت.

۲- مدل سیستم و فرضیات

مطابق با شکل ۱ سناریویی در نظر گرفته شده که در آن یک گره IoT مجهّز به چند حسگر، در هر برهه زمانی، داده‌ها را از محیط دریافت و آن‌ها را از طریق کانال بی‌سیم برای گره مقصد ارسال می‌نماید. به‌منظور مدیریت انرژیِ در دسترس مطابق با شرایط محیطی، گره باید در خصوص مقدار انرژی تخصیص داده شده به واحدهای فشرده‌سازی و ارسال تصمیم‌گیری نماید. ازاین‌رو، گره در هر برهه زمانی با توجه به وضعیت انرژی موجود در باتری، حجم داده‌های دریافتی توسط حسگرها، وضعیت بافر داده و کیفیت کانال باید تصمیم بگیرد که داده‌های ورودی را با چه سطحی فشرده کند و چه تعداد بستهٔ داده ارسال نماید به‌گونه‌ای که ضمن کاهش مصرف انرژی، میانگین تأخیر ارسال بسته‌ها از یک مقدار آستانه‌ای فراتر نرود. بااین‌حال، ما یک سیاست رفتاری ثابت برای گره تعریف نمی‌کنیم چراکه برای داشتن عملکرد بهینه در طولانی‌مدت، لازم است گره رفتاری تطبیق‌پذیر نسبت به وضعیت جاری سیستم داشته باشد. در این قسمت ابتدا به معرفی مدل برای سیستم موردنظر می‌پردازیم.

مطابق با آخرین تکنولوژی در تجهیزات اینترنت اشیاء، مشابه [۶]، فرض کردیم گره IoT متّکی به یک منبع انرژی تجدیدپذیر است که قابلیت برداشت انرژی از محیط را دارد و از یک پنل خورشیدی برای تأمین انرژی استفاده می‌کند. پویایی منبع انرژی را به‌صورت زنجیرهٔ مارکُف با تعداد حالات محدود FSMC[۵] مدل می-کنیم. $x_t = 0$ به این معنا است که در حالت کم‌نور یا بی‌نور قرار داریم و مقدار انرژی برداشت‌شده در این برهه زمانی $e_t^{in} = 0$ است و هنگامی‌که در حالت پرنور باشد، $x_t = 1$ یعنی مقدار انرژی برداشت‌شده از یک بازه $e_t^{in} \in E = \{1, \dots, E\}$ به‌صورت تصادفی با توزیع یکسان محقق می‌گردد.

انرژی برداشت‌شده از محیط در یک باتری با ظرفیت محدود ذخیره می‌شود. فضای حالت باتری را به‌صورت $E = \{1, \dots, E\}$ در نظر می‌گیریم که در آن صفر به معنای خالی بودن و E به معنای پر بودن باتری است. مشابه[۶]، موقعیت بعدی باتری به‌صورت رابطه (۱) بیان می‌شود.

$$e_{t+1} = min\big(max(e_t - e_t^{out}, 0) + e_t^{in}, E\big) \quad (۱)$$

حسگرهای موجود در دستگاه IoT وظیفه دارند که داده‌های محیطی را در هر برهه زمانی T_{sens} جمع‌آوری کنند و داده‌های اندازه‌گیری شده را به گره چاهک ارسال کنند. در این پژوهش کاربردهایی در نظر گرفته شده که نرخ نمونه‌برداری ثابتی دارند. داده‌هایی که در هر برهه زمانی جمع‌آوری می‌شوند فرض می‌شود که با یکدیگر هم‌بستگی دارند و سایز آن‌ها N_b بیت در نظر گرفته شده است در این کار از الگوریتم فشرده‌سازی با اتلاف LTC برای فشرده‌سازی داده‌ها استفاده می‌شود [۲۰]. داده‌ها ابتدا به یک یا چند بسته با طول ثابت L بیت فشرده می‌شوند. داده‌های فشرده برای انتقال روی لینک بی‌سیم، به یک بافر با اندازهٔ محدود با فضای حالت گسسته و متناهی به‌صورت $\mathcal{B} = \{0,1,2, \dots, B\}$ وارد می‌شوند. مشابه باتری، حالت بعدی بافر داده به‌صورت رابطه (۲) بیان می‌شود.

$$b_{t+1} = min(max(b_t - b_t^{out}, 0) + b_t^{in}, B) \quad (۲)$$

که در آن b_t^{in} تعداد بسته‌های ورودی به بافر، b_t تعداد بسته‌های موجود در بافر و b_t^{out} تعداد بسته‌های خروجی از بافر که توسط واحد تصمیم‌گیرنده متناسب با شرایط محیطی تعیین و ارسال می‌گردد.

سطح فشرده‌سازی انتخابی در لحظه t را با k نشان می‌دهیم که مقدار یک به این معنا است که فشرده‌سازی بر روی داده‌ها انجام نشده است و k برابر با یک یعنی بالاترین سطح فشرده‌سازی صورت پذیرفته است و وقتی این مقدار برابر با n باشد یعنی هیچ فشرده‌سازی بر روی داده‌ها انجام نمی‌شود. ما در این کار مقدار انرژی که برای جمع‌آوری داده‌ها در دستگاه IoT مصرف می‌شود در مقابل مقدار انرژی موردنیاز برای فشرده‌سازی و انتقال داده، نادیده می‌گیریم. مطابق کار [۲۱] مقدار انرژی مصرفی برای فشرده‌سازی داده‌ها از رابطه (۳) قابل محاسبه است.

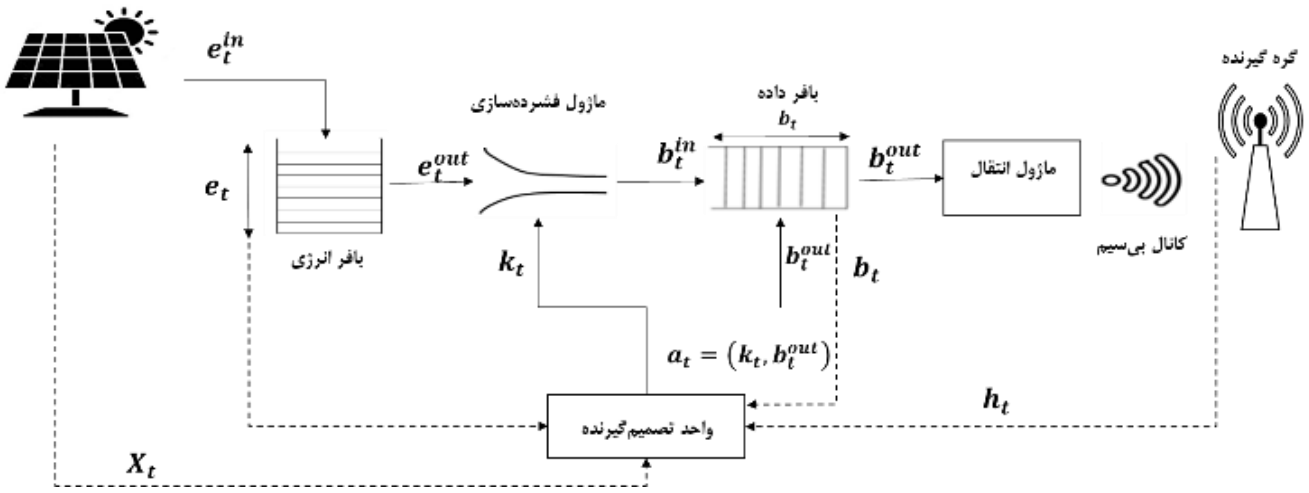
$$e_c(k) \begin{cases} \left(\gamma \frac{k}{n} + \ell\right) N_b E_0 & 0 < k < n \\ 0 & k \epsilon \{0, n\} \end{cases} \quad (۳)$$

داده‌هایی که در بافر قرار می‌گیرند باید روی یک کانال با شرایط متغیر ارسال شوند و درنتیجه توان دریافتی در هر گیرنده را می‌توان به‌صورت ضرب بهره کانال در توان ارسالی در نظر گرفت. بهره کانال g_t بین گره فرستنده و گیرنده را به‌صورت رابطه (۴) می‌توان بیان کرد که در آن h_t ضریب کانال بین فرستنده و گیرنده و d فاصله بین آن‌ها است. هم‌چنین α توان افت مسیر است.

$$g_t = |h_t|^2 * d^{-\alpha} \quad (۴)$$

همان‌طور که در رابطه نشان داده شده است، بهره کانال از دو بخش قطعی و غیرقطعی که به ترتیب d و h_t هستند تشکیل شده است. h_t یک متغیر تصادفی مستقل با توزیع یکسان در نظر گرفته شده است که مدلی برای پارامترهای تصادفی کانال مانند محوشدگی، پراکندگی است.

ضریب کانال رادیویی بین گره‌های شبکه به‌صورت مدل مارکُف حالت محدود در نظر گرفته شده است [۲۲].



شکل ۱- مدل سیستم برای گره اینترنت اشیا

مشابه[۲۳]، توان لازم برای ارسال b_t^{out} بستهٔ داده به طول L بیت در کانال g_t با پهنای باند W برای یک ارتباط قابل‌اطمینان و بدون خطا به‌صورت رابطهٔ (۵)، محاسبه می‌گردد.

$$p(g_t, b_t^{out}) = \frac{WN_0}{g_t}\left(2^{\frac{b_t^{out}{}_l}{W}} - 1\right) \qquad (۵)$$

N_0 بیان‌گر نویز گوسی سفید با میانگین صفر و واریانس N_0^2 می‌باشد و حاصل‌ضرب WN_0 به ۱ نرمال شده است. همان‌طور که از رابطهٔ فوق برمی‌آید، در یک کانال، توانِ ارسال تابعی اکیداً صعودی از b_t^{out} می‌باشد.

انرژی لازم برای ارسال بسته‌های داده به‌صورت حاصل‌ضرب توان در واحد زمان طبق رابطهٔ (۶) محاسبه می‌شود.

$$E_{Tx}(g_t, b_t^{out}) = p(g_t, b_t^{out}).\tau \qquad (۶)$$

۳- فرمول‌بندی مسئله با استفاده از فرآیند تصمیم مارکُفِ مقیّد

در بسیاری از موارد در بهینه‌سازی سیستم‌های پویا، کنترل‌کننده باید یک هدف را با ایجاد محدودیت روی اهداف دیگر بهینه‌سازی کند تا یک موازنه میان اهداف مختلف حاصل شود. این کلاس از فرآیندهای تصمیم مارکُف (MDP) به‌عنوان فرآیندهای تصمیم مارکُف مقیّد (CMDP) نامیده می‌شوند [۳۰]. در مسئله موردبررسی ما نیز کنترل‌کننده باید متوسط تطابق داده‌ها را به حداکثر برساند ضمن اینکه قید متوسط تأخیر ارسال نیز تأمین شود ازاین‌رو مسئله ما در قالب CMDP مدل می‌شود. CMDP را می‌توان به‌صورت چندتایی $\{S, A, P, r(.), c(.)\}$ تعریف کرد که به ترتیب از چپ به راست شامل: فضای حالت موقعیت‌ها، فضای حالت اقدامات امکان‌پذیر، تابع احتمال گذار، پاداش آنی و قید آنی می‌باشد.

فضای موقعیت سیستم یک فضای گسسته و متناهی به‌صورت $s = \mathrm{E} \times \mathcal{B} \times \mathcal{H} \times \mathcal{X}$ می‌باشد. $s_t \in S$ موقعیت سیستم در لحظهٔ t، به‌صورت چندتایی $s_t = [e_t,\ b_t, h_t, x_t]$ تعریف می‌شود که در آن e_t وضعیت باتری، b_t وضعیت جاری بافر داده، h_t وضعیت جاری کانال بی‌سیم و x_t وضعیت منبع انرژی می‌باشد. در هر برهه زمانی، واحد تصمیم‌گیرنده چهار مؤلفه از وضعیت گره و محیط یعنی $s_t = [e_t,\ b_t, h_t, x_t]$ را به‌عنوان ورودی دریافت و اقدامات امکان‌پذیر در وضعیت جاری را شناسایی می‌کند. فضای حالت اقدامات امکان‌پذیر در s_t، یک فضای گسسته و متناهی به‌صورت $A_{s_t} = \mathcal{K}_{s_t} \times \mathcal{B}_{s_t}$ است که در آن $\mathcal{K}_{s_t}$ مجموعهٔ سطوح فشرده‌سازی امکان‌پذیر (با توجه به وضعیت باتری) و $\mathcal{B}_{s_t}$ مجموعهٔ تعداد بسته‌های دادهٔ قابل ارسال (با توجه به وضعیت باتری، بافر داده و کانال) می‌باشد. هر یک از سطوح فشرده‌سازی k و هر تعداد از بسته‌های داخل بافر b که در شرط رابطهٔ (۷) صدق نمایند عناصر مجموعهٔ $\mathcal{K}_{s_t}$ و $\mathcal{B}_{s_t}$ را تشکیل می‌دهند.

طبق رابطه (۷)، ملاک تشخیص امکان‌پذیر بودن اقدامات، مقایسه مجموع انرژی فشرده‌سازی و انرژی ارسال با سطح انرژی باتری است؛ چون تنها اقدامی توسط گره قابل انجام است که باتری قادر به تأمین انرژی آن‌ها باشد.

$$\mathbb{I}\left(k \in \mathcal{K}_{s_t}, b_t^{out} \in \mathcal{B}_{s_t}\right) = \begin{cases} 1 & if \quad e_c(k) + E_{TX}(g_t, b_t^{out}) \le e_t * E_0 \ and \ \ b_t^{out} \le b_t \\ 0 & otherwise \end{cases} \qquad (۷)$$

تصمیم کنترلیِ فعلی واحد تصمیم‌گیرنده با نماد $a_t \in A$ به‌صورت دوتایی $a_t = (k, b_t^{out})$ تعریف می‌شود.

مشابه[۱۵]، پاداش آنی ناشی از اجرای اقدام k_t در موقعیت s_t، را با نماد $R(k_t)$، به‌صورت رابطه (۸) نشان می دهیم.

$$R(k) = \begin{cases} 0 \\ 1 - \left(\dfrac{p_1\left(\frac{k}{n}\right)^2 + p_2\left(\frac{k}{n}\right) + p_3}{\left(\frac{k}{n}\right) + q_1}\right)\sigma_{noise}^2 + p_4 * \mathcal{R} \end{cases} \qquad (۸)$$

که در آن $p_1 = -1.5370$ ، $p_2 = 1.700305$ ، $p_3 = 0.17466$، $p_4 = -0.025$ و $q_1 = 0.00267$ که این ضرایب ثابت هستند و به‌صورت کلی به هم‌بستگی زمانی سیگنال‌های مشاهده شده، مرتبط هستند. σ_{noise} نیز واریانس نویز سفید که در سیگنال‌های جمع‌آوری شده، وجود دارد. مقدار $\mathcal{R}$ نیز تعداد بسته‌هایی است که در برهه زمانی قبل از بافر داده سرریز شده‌اند و باعث هدر رفت این بسته‌ها می‌شود.

برای تعریف قید تأخیر گزارش رویدادها از قانون Little پیروی می‌کنیم. طبق قانون Little در[۱۵]، مطابق رابطهٔ (۹)، متوسط طول بافر ارسال $\bar{C}_B$ برابر است با حاصل‌ضرب متوسط نرخ ورود داده‌های حسگری $\bar{a}$ در متوسط تأخیر تجربه‌شده توسط بسته‌های داخل بافر $\bar{D}$.

$$\bar{C}_B = \bar{a}\,\bar{D} \qquad (۹)$$

مشابه [۲۱]، متوسط تأخیر را مترادف با متوسط طول بافر در نظر گرفته و از ثابت $\bar{a}$ صرف‌نظر شده است.

ازآنجاکه داده‌ها پیش از ارسال داخل بافر قرار می‌گیرند و با ایجاد قید، تأخیری که بسته‌های داخل بافر تجربه می‌کنند از مقدار آستانهٔ δ فراتر نرود، بنابراین هزینهٔ بافر کردن را به‌عنوان یک قید آنی به‌صورت رابطه (۱۰) مشابه [۱۶]، در نظر گرفته شده است.

$$C_B(s_t, a_t, s_{t+1}) \triangleq b_{t+1} \quad (۱۰)$$

ازآنجاکه گره در یک محیط تصادفی عمل می‌کند هدف، به حداکثر رساندن متوسط تطابق داده‌ها در گیرنده در بلندمدت $\bar{R}_f$، تحت تأمین محدودیت δ روی متوسطِ طول بافر $\overline{C}_B$ به‌عنوان تابع قید به‌صورت مسئله (۱۱) می‌باشد.

$$\textbf{Maximize } \bar{R}_f \quad \textbf{s. to } \overline{C}_B \le \delta \quad (۱۱)$$

مشابه [۲۲]، گره اینترنت اشیاء باید سیاست بهینه را جهت کنترل سطح فشرده‌سازی و نرخ ارسال یاد بگیرد؛ که در آن مقدار δ توسط طراح بر اساس میزان تحمل‌پذیری تأخیر در کاربردهای عملی تعیین می‌گردد.

مسئله مقیّد بیان شده در رابطه (۱۱) را می‌توان با استفاده از روش استاندارد لاگرانژین، به فرم نامقیّد بازنویسی کرد. این روش، برای ترکیب تابع هدف با قید در رابطه (۱۱)، از یک ضریب لاگرانژ $\lambda \ge 0$ استفاده می‌کند و همانند [۲۴]، یک تابع هزینه جدید به نام لاگرانژین به‌صورت رابطه (۱۲) تعریف می‌کند.

$$l(s, a, \lambda) = R_f(s, a) - \lambda(C_B(s_t, a_t, s_{t+1}) - \delta) \quad (۱۲)$$

در رابطهٔ (۱۲) اگر متوسط طول بافر C_B از مقدار آستانه δ بیشتر شود با ضریب مثبت λ جریمه و هزینه بیشتری برای آن لحاظ می‌شود.

۴- روش پیشنهادی مبتنی بر یادگیری تقویتی

محاسبه سیاست رفتاری تطبیق‌پذیر برای گره IoT از طریق روش‌های «مبتنی بر مدل» میسّر نیست زیرا جواب نظیر یک مدل خاص با تغییر شرایط، اعتبار خود را از دست می‌دهد. هم‌چنین، در بسیاری از سناریوهای کاربردی، امکان مدل‌سازی دقیق ساختار احتمالاتی سیستم میسّر نیست و اغلب در پیاده‌سازی عملی، به دست آوردن دانش آماریِ فرآیندهای تصادفی محیط عملیاتی (نظیر: مدل ورود داده‌های حسگری، مدل احتمالاتی برداشت انرژی از محیط، توزیع احتمال تغییر کیفیت کانال) دشوار می‌باشد. با توجه به مقید بودن مسئله موردبحث ما، کل پروسه یادگیری به‌صورت یک رویه تقریباً تصادفی با دو مقیاس زمانی صورت می‌گیرد. در مقیاس زمانی سریع‌تر مقادیر تابع ارزش موقعیت-عمل در Q-Learning یا تابع ارزش حالت در PDS و VE به‌روزرسانی می‌شود.

۴-۱- روش یادگیری تقویتی Q

در این بخش ما از تکنیک‌های یادگیری تقویتی [۱۹]، برای یافتن سیاست بهینه برای کنترل سطح فشرده‌سازی و نرخ ارسال در گره اینترنت اشیاء استفاده می‌کنیم. انگیزهٔ استفاده از تکنیک‌های یادگیری تقویتی این است که ازآنجاکه در پیاده‌سازی عملی، به دست آوردن دانش آماری از فرآیندهای تصادفی محیط عملیاتی (نظیر: مدل ورود داده‌های حسگری، مدل احتمالاتی برداشت انرژی از محیط، توزیع احتمال تغییر کیفیت کانال) دشوار می‌باشد، ما نیاز به یک الگوریتم مستقل از مدل داریم که سیاست بهینه را صرفاً از طریق بازخورد آنی از پاداش آنی $R_{k_t}(s_t, a_t)$ و قید بافر $C_B(s_t, a_t, s_{t+1})$ یاد بگیرد.

با توجه به فرمول‌بندی مبتنی بر CMDP که در بخش۳ بیان شد، ما می‌توانیم از تکنیک‌های یادگیری تقویتی برای یادگیری سیاست بهینه استفاده کنیم. به‌طور خاص، ما از الگوریتم استاندارد Q-learning [۱۹] به‌عنوان مبنای یادگیریِ مستقل از مدل استفاده می‌کنیم که نمونه‌های واقعی گذار و هزینه‌ها را به‌عنوان تجربه آنلاین بکار می‌گیرد و هزینهٔ بهینه را با میانگین تصادفی تخمین می‌زند. برای یک λ ثابت، الگوریتم Q-learning به‌تدریج سیاست بهینه را بدون دانش احتمال‌های گذار P یاد می‌گیرد. در این الگوریتم، $Q^{*,\lambda}(s, a)$ بیان‌گر مجموع لاگرانژین آنی $l(s, a, \lambda)$ حاصل از اجرای عمل a در موقعیت s همراه با مقدار متوسط بلندمدت لاگرانژین با پیروی از سیاست بهینه $\pi^{*,\lambda}$ در تمام موقعیت‌های بعدی می‌باشد که به‌صورت معادله بلمن(۱۳) تعریف می‌شود:

$$Q^{*,\lambda}(s, a) = l(s, a, \lambda) - \rho^* + \sum_{\acute{s} \in \delta} \mathbb{P}(\acute{s}|s, a)\bar{\mathcal{L}}^{\lambda,\pi^*}(\acute{s}) \quad (۱۳)$$

در این رابطه ρ^* نشان‌دهنده متوسط پاداش بهینه است و علاوه بر این:

$$\bar{\mathcal{L}}^{\lambda,\pi^*}(s) = \max_{a \in \mathcal{A}(\acute{s})} Q^{*,\lambda}(s, a), \quad \forall s \in \mathcal{S} \quad (۱۴)$$

مقدار $Q(s, a)$ برای هر موقعیت-عمل در یک عدد خلاصه می‌شود و یادگیری $Q^{*,\lambda}(s, a)$ به یادگیری سیاست بهینه منجر می‌شود. با تعریف بازگشتی $Q(s, a)$ و استفاده از الگوریتم‌هایی که مکرراً $Q(s, a)$ را تخمین می‌زنند، می‌توان به مقدار واقعی $Q^{*,\lambda}(s, a)$ برای تمام (s, a)ها دست یافت. در ابتدای کار، یادگیرنده مقدار فرضی $\hat{Q}(s, a)$ را برای تمام (s, a) ها با صفر یا مقادیر دلخواه، مقداردهی می‌نماید. در هر تکرار الگوریتم، گره موقعیت فعلی s را مشاهده، از میان اقدامات امکان‌پذیر در موقعیت فعلی، عمل a را به‌صورت احتمالاتی انتخاب و آن را اجرا می‌کند. سپس هزینهٔ اجرای عمل a را محاسبه و موقعیت بعدی ś را مشاهده می‌کند. در موقعیت ś اقدامات امکان‌پذیر را شناسایی و طبق قانون آموزش (۳-۳۰) مقدار فرضی $\hat{Q}(s, a)$ را برای (s, a) فعلی به‌روزرسانی می کند.

$$\hat{Q}_n(s, a) \leftarrow (1 - f(n))\hat{Q}_{n-1}(s, a) + f(n)\left[l(s, a, \lambda) - \rho^* + \max_{\acute{a} \in \mathcal{A}(\acute{s})} \hat{Q}_{n-1}(\acute{s}, \acute{a})\right] \quad (۱۵)$$

$\widehat{Q}_n(s, a)$ بیانگر n اُمین تخمین از مقدار $Q^{*,\lambda}(\mathrm{s,a})$ می باشد.

۴-۲- روش یادگیری تقویتیPDS

در این بخش از روش PDS [۱۸] برای حل مسئله بیان شده استفاده می‌کنیم که یک روش تسریع شده نسبت به الگوریتم استاندارد Q-learning [۱۹] است و هم‌چنین سرعت همگرایی بالاتری نیز دارد. در این روش ابتدا یک موقعیت میانی را به‌عنوان موقعیت PDS تعریف می‌کنیم: $\widetilde{s_t} = (\widetilde{e_t}, \widetilde{b_t}, \widetilde{h_t}, \widetilde{x_t}) \in S$؛ این موقعیت از سیستم بعد از این‌که همه بخش‌های پویای شناخته شده سیستم رخ دهد اما قبل از اینکه بخش‌های پویای ناشناخته سیستم به وقوع بپیوندد به‌صورت رابطه (۱۶) در سیستم موردنظر ما قابل تعریف است:

$$\widetilde{s_t} = (\widetilde{e_t}, \widetilde{b_t}, \widetilde{h_t}, \widetilde{x_t}) = ((e_t - e_t^{out}), \min(b_t + b_t^{in} - b_t^{out}, N_b), h_t, x_t) \quad (۱۶)$$

حالت بافر انرژی PDS، به این صورت تعریف می‌شود: قبل از اینکه بسته‌های جدید انرژی در این برهه زمانی به بافر برسند و بعد از این‌که مقدار انرژی که باید در این برهه از بافر خارج شود $\widetilde{e_t} = e_t - e_t^{out}$. حالت بافر PDS، به‌صورت قطعی از بین مقادیر ممکن در این رابطه $\widetilde{b_t} = \min(b_t + b_t^{in} - b_t^{out}, N_b)$ تعریف می‌شود. وضعیت کانال و انرژی منبع در حالت PDS مشابه با حالت معمول موقعیت-ها قابل تعریف است. به‌عبارت‌دیگر PDS تمام اطلاعات شناخته شده در مورد گذر از حالت s_t به حالت s_{t+1} را پس از اقدام a_t ثبت می‌کند. اگر در این حالت تعداد e_t^{in} بسته به بافر انرژی برسد وضعیت بافر انرژی در زمان $t+1$ برابر با $e_{t+1} = \min(\tilde{e}_t + e_t^{in}, N_e)$ می‌شود، همچنان که حالت انرژی منبع و کانال به ترتیب با h_{t+1}, x_{t+1} نمایش داده می‌شود.

$$s_{t+1} = (e_{t+1}, b_{t+1}, h_{t+1}, x_{t+1}) = (\min(\tilde{e}_t + e_t^{in}, N_e), \tilde{b}_t, h_{t+1}, x_{t+1}) \quad (۱۷)$$

حال وقتی‌که حالت میانیPDS را در نظر گرفته شده این می‌تواند تمام فضای حالت موقعیت‌ها را پوشش دهد، به دلیل این‌که مقدار ارزش حالت s_t با استفاده از

حالت بعد آن نیز می‌تواند تعیین شود، زیرا هر دو $\tilde{S}$ و S' از یک فضا مقدار می‌پذیرند، $\tilde{V}^*(\tilde{s})$ را می‌توان با استفاده از رابطه (۱۸) محاسبه کرد.

$$\tilde{V}^*(\tilde{s}) = \sum_{\acute{x}} \sum_{e_{in}} \sum_{\acute{h}} P(\acute{x}|\tilde{x}) P(e_{in}|\tilde{x}) P(\acute{h}|\tilde{h})\, V^*(\min(\tilde{e}_t + e_t^{in}, N_e), \tilde{b}_t\ , \acute{h}\ , \acute{x}) \quad (۱۸)$$

$E^S[V^*(s)]$ که میانگین روی همه موقعیت‌های بعد از PDS است. به رابطه(۱۹) که برای محاسبه ارزش موقعیت PDS به کار می‌رود معادله بلمن می‌گویند.

$$V^*(s) = \left(\max_{\mathbb{I}(k\in\mathcal{K}_{s_t}, b_t^{out}\in\mathcal{B}_{s_t})} \left[l(a,\tilde{s},\lambda) + \sum_{b_t^{in}} \mathcal{P}^{b_t^{in}}(b_t^{in}|k,N_{in})\, \tilde{v}^* \left(\left((e_t - e_t^{out}), \min(b_t + b_t^{in} - b_t^{out}, N_b), h_t, x_t \right) \right) \right] \right) - \beta \quad (۱۹)$$

در رابطه(۱۹)، β به‌عنوان یک ضریب بهینه به ازای هزینه هر مرحله معرفی می‌شود که باید این پارامتر تخمین زده شود، می‌توان به جای این پارامتر یک حالت دلخواه از وضعیت‌ها $\acute{s_t} = (\acute{e}_t,\ \acute{b}_t, \acute{h}_t, \acute{x}_t)$ را به‌عنوان موقعیت مرجع در نظر گرفت و ارزش آن حالت را در معادله بلمن قرار داد.

$$V^*(s) = \left(\max_{\mathbb{I}(k\in\mathcal{K}_{s_t}, b_t^{out}\in\mathcal{B}_{s_t})} \left[l(a,\tilde{s},\lambda) + \sum_{b_t^{in}} \mathcal{P}^{b_t^{in}}(b_t^{in}|k,N_{in})\, \tilde{v}^* \left(\left((e_t - e_t^{out}), \min(b_t + b_t^{in} - b_t^{out}, N_b), h_t, x_t \right) \right) \right] \right) - \tilde{v}^* \left((\acute{e}_t,\ \acute{b}_t, \acute{h}_t, \acute{x}_t) \right) \quad (۲۰)$$

در ادامه الگوریتمی برای ارزیابی V^* بیان می‌شود. با تعریف بازگشتی $V^*(s)$ و استفاده از الگوریتم‌هایی که مکرراً $\tilde{v}(\tilde{s})$ را تخمین می‌زنند، می‌توان به مقدار واقعی $\tilde{v}(\tilde{s})$ دست یافت. در ابتدای کار، یادگیرنده مقدار فرضی $\tilde{v}(\tilde{s})$ را تخمین می‌زنند، می‌نماید. در هر تکرار الگوریتم، گره موقعیت فعلی s را مشاهده، از میان اقدامات امکان‌پذیر در موقعیت فعلی، عمل a را به‌صورت حریصانه انتخاب می‌کند و آن را اجرا می‌کند. طبق قانون آموزش (۲۱) مقدار فرضی $\tilde{v}(\tilde{s})$ را تخمین می‌زنند، را برای s فعلی به‌روزرسانی می‌کند.

$$\tilde{V}(\tilde{e},\tilde{b},\tilde{h},\tilde{x}) \leftarrow (1-f(t))\tilde{V}(\tilde{e},\tilde{b},\tilde{h},\tilde{x}) + f(t)\left[\max_{\mathbb{I}(k\in\mathcal{K}_{s_t}, b_t^{out}\in\mathcal{B}_{s_t})} \left[l(a,\tilde{s},\lambda) + \sum_{b_t^{in}} \mathcal{P}^{b_t^{in}}(b_t^{in}|k,N_{in})\, \tilde{v}^* \left(\left((e_t - e_t^{out}), \min(b_t + b_t^{in} - b_t^{out}, N_b), h_t, x_t \right) \right) \right] \right. \quad (۲۱)$$

۴-۳- روش یادگیری تقویتیVE

در این روش از این واقعیت استفاده می‌کنیم که پویایی ناشناخته مستقل از اجزای خاصی از وضعیت سیستم است. ما از این خاصیت برای انجام به‌روزرسانی «دسته‌ای» در چندین حالت بعد از هر وضعیت در هر برهه زمانی بهره‌برداری می‌کنیم. ما به این به‌روزرسانی دسته‌ای به‌عنوان یادگیری تجربه مجازی (VE)[۱۸] اشاره می‌کنیم. یادگیری تجربه مجازی یک تکنیک تسریع همگرایی در حوزه یادگیری تقویتی است.

در مسئله ما ، یادگیری تجربه مجازی با این واقعیت امکان‌پذیر است که ورود بسته انرژی ناشناخته، پویایی انتقال کانال و پویایی انتقال منبع انرژی مستقل از بافر انرژی پس از تصمیم‌گیری و حالت‌های بافر داده‌ها است. این به ما این امکان را می‌دهد تا همه حالت‌های بعد را با $\tilde{h}_t$ و $\tilde{x}_t$ یکسان اما با $\acute{e}$ و $\acute{b}$ متفاوت به‌روزرسانی کنیم. به‌روزرسانی $|S_e \times S_b|$ حالت‌های پس‌ازآن در هر برهه زمانی به‌طور قابل‌توجهی سرعت همگرایی را با هزینه افزایش پیچیدگی محاسباتی بهبود می‌بخشد. به‌طور خاص، اگر به‌روزرسانی در هر برهه زمان T اعمال شود، میانگین تعداد حالت‌های پس از به‌روزرسانی در هر برهه زمانی $|S_e \times S_b|/T$ است.

۴-۴- یادگیری ضریب لاگرانژ

برای محاسبه ضریب لاگرانژ در هر برهه زمانی از تکنیک یادگیری استفاده می‌شود و مشابه [۲۲]، تخمین فعلی از ضریب لاگرانژ λ_t را بر اساس الگوریتم استاندارد «صعود زیرگرادیان تصادفی»طبق رابطه (۲۲) به‌روزرسانی می‌شود.

$$\lambda_{t+1} = \Lambda\left[\lambda_t + e(t)(C_B(s_t,a_t,s_{t+1}) - \delta)\right] \quad (۲۲)$$

رابطه (۲۲)، λ را با استفاده از نرخ یادگیری $e(t)$ در جهت گرادیان تابع نسبت به λ به‌روزرسانی می‌کند. نماد $e(t)$ بیان‌گر اندازهٔ گام الگوریتم «زیر-گرادیان تصادفی» است.

قانون آموزش (۲۱) و (۲۲) تخمین‌های $\tilde{V}^*(\tilde{s})$ و λ_t را به‌طور هم‌زمان ولی در مقیاس‌های زمانی متفاوت به‌روزرسانی می‌کند. بنابراین نرخ به‌روزرسانی‌های این دو تخمین یعنی $f(t)$ و $e(t)$ باید طوری انتخاب شوند که شرایط استاندارد تقریب تصادفی[۲۵]، رابطه (۲۰) در مورد آن‌ها صدق نماید تا همگرایی فرایند یادگیری به استناد قضیه همگرایی مارتینگل در تئوری احتمال کاربردی [۲۶]، تضمین شود.

$$\sum_t (f(t)^2 + e(t)^2) < \infty \qquad \lim_{t\to\infty} \frac{e(t)}{f(t)} \to 0 \quad (۲۳)$$

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 روش‌های پیشنهادی را شرح خواهیم داد و در آخر نمودارها و نتایج شبیه‌سازی را ارائه خواهیم کرد.

۵-۱- بستر آزمایش و تنظیم مقادیر شبیه‌سازی

تنظیم و مقداردهی پارامترهای شبیه‌سازی مطابق جدول ۱ می‌باشد. هر سه روش شرح داده شده در بخش ۴ را در شرایط محیطی کاملاً یکسان مورد آزمایش قرار دادیم. پارامترهای شبیه‌سازی را مطابق با جدول ۱ تنظیم و مقداردهی شده‌اند.

در این پژوهش تأثیر افزایش حجم داده حس‌شده در هر برهه زمانی را بر مقدار متوسط تطابق داده‌ها، متوسط انرژی مصرف شده و هم‌چنین بر میزان هدر رفت بسته‌ها در هر سه روش پیشنهادی دیده شده است. هم‌چنین اثر افزایش مقدار انرژی هر خانه از بافر را بر روی این سه پارامتر مشاهده کردیم.

۵-۲-ملاک‌های ارزیابی

برای ارزیابی الگوریتم یادگیری PDS، VE و مقایسه با الگوریتم استاندارد Q-learning، متوسط تابع هدف $\bar{R}_f$ و قید متوسط بافر $\overline{C}_B$ را ملاک قرار می‌دهیم.

درواقع، الگوریتمی کاراتر است که متوسط تطابق داده‌ها در بلندمدت بیشتر شده باشد، هدر رفت بسته کمتر داشته باشد و ضمناً قید متوسط بافر آن زیر مقدار آستانه تعیین شده δ قرار بگیرد.

ملاک دیگر سرعت همگرایی این الگوریتم‌ها به سیاست بهینه است درواقع الگوریتمی کاراتر است که سرعت همگرایی بیشتری برای به دست آوردن سیاست بهینه داشته باشد.

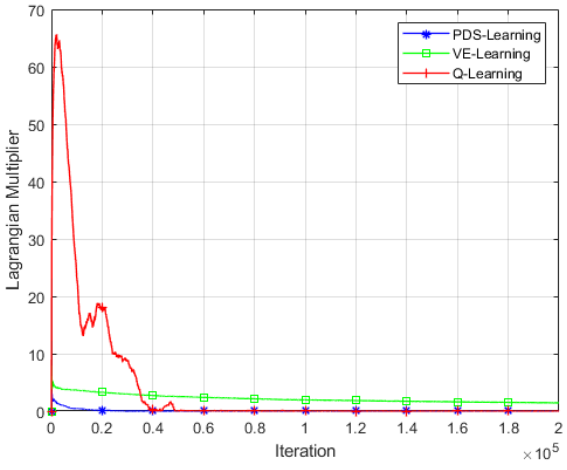
ملاک دیگر سرعت همگرایی مقادیر تابع ارزش موقعیت-عمل در Q-Learning و تابع ارزش حالت در روش‌های PDS و VE می‌باشد.

جدول ۱- پارامترهای شبیه‌سازی تنظیم و مقداردهی

پارامتر	مقادیر	توضیحات
τ	۳۰۰ میلی‌ثانیه	طول برهه زمانی
n	یک‌میلیون تکرار	تعداد تکرار الگوریتم
E	۱۰بسته انرژی(ژول)	ظرفیت بافر انرژی
L	۱۲۵۰۰ بیت	حجم بسته‌های داده
B	۱۵ بسته	ظرفیت بافر داده
p_4	۰٫۰۲۵	ضریب جریمه هدر رفت بسته
k	۶	تعداد سطح‌های فشرده‌سازی
H	[۰٫۰۱، ۰٫۱، ۰٫۳، ۰٫۶، ۰٫۹، ۱٫۲، ۱٫۵، ۱٫۸، ۲٫۱، ۲٫۴]	فضای حالت بهره کانال ارتباطی بی‌سیم
W	۵ مگاهرتز	پهنای باند کانال
S	$S=\{10\times15\times10\times2\}$	فضای حالت سیستم
δ	۱۲بسته داده	قید آستانه بافر
λ	صفر	مقدار اولیه ضریب لاگرانژ
N_b	۲۰۰۰۰ بایت	حجم داده حس‌شده در هر برهه زمانی

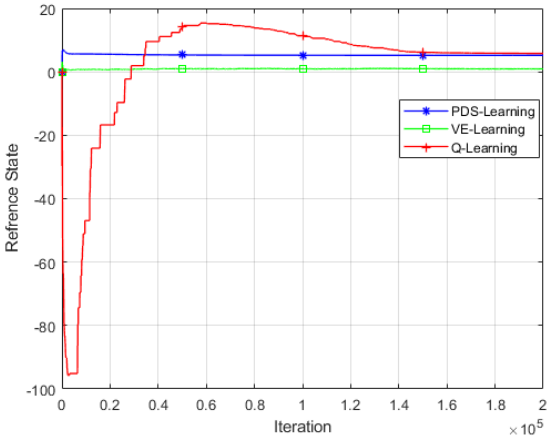
۵-۳- نتایج شبیه‌سازی

ابتدا تکامل ضریب لاگرانژ را به‌عنوان شاخص اصلی رفتار همگرایی الگوریتم در مسئله بهینه‌سازی مقیّد نشان می‌دهیم. برای این منظور، تخمین‌های λ_t حاصل از تکرار فرایند (۱۹) برای هر سه روش در شکل ۲ رسم شده است. همان‌طور که مشاهده می‌شود ضریب لاگرانژ مقیّد به قید بافر بعد از تقریباً ۱۰۰۰ دور همگرا می‌شود. همگرایی به یک مقدار غیر صفر بیان‌گر حالتی است که قید بافر به‌طور مرزی تأمین شده و تأمین هدف و قید با توجه به شرایط مسئله آسان نبوده است. همان‌طور که در شکل ۲ نشان داده شده، ضریب لاگرانژ برای روشVE سریع‌تر از روشPDS و Q-learning همگرا شده است.

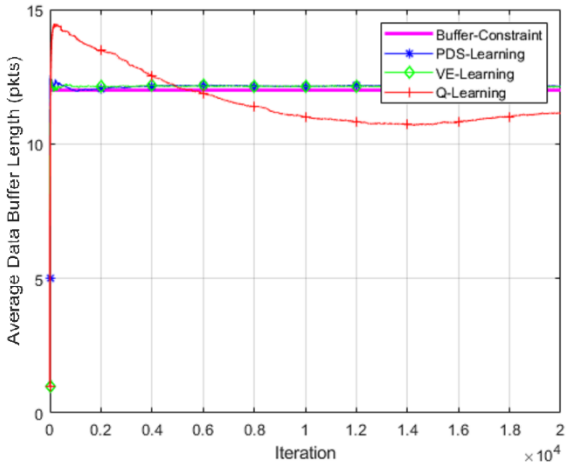


شکل ۲- همگرایی ضریب لاگرانژ در سه روش Q، PDSو VE

همان‌طور که در شکل ۳ نشان داده شده، این شکل همگرایی حالت مرجع را در هر سه الگوریتم نشان می‌دهد. در روش PDS و VE وقتی‌که ارزش حالت مرجع همگرا می‌شود به این معنی است که این روش‌ها توانسته به سیاست بهینه دست یابند. همچنین از نظر سرعت همگرایی روش استاندارد Q- Learning از بین روش-های دیگر با سرعت کمتری همگرا می‌شود.

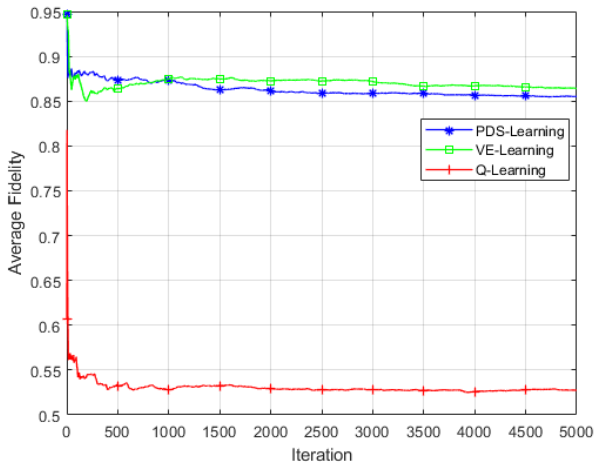


شکل ۳- همگرایی حالت مرجع در روش‌های Q، PDSو VE



شکل۴- همگرایی قید بافر داده در روش‌های Q، PDS وVE

در شکل ۴ همگرایی قید بافر در هر سه روش نشان داده شده است که در روش Q-learning مقدار قید بافر با گذشت زمان پایین‌تر از مقدار مشخص شده می‌شود ولی در دو روش دیگر قید بافر همواره در روی مقدار مشخص شده قرار می‌گیرد.

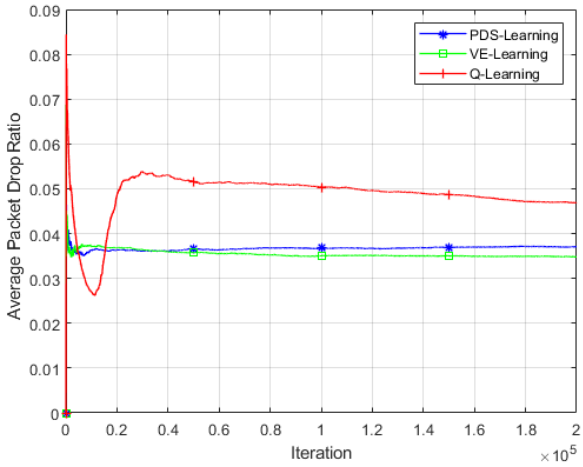


شکل۵- متوسط تطابق داده‌ها در سه روش Q، PDS و VE

در شکل ۵ همگرایی به مقدار متوسط تطابق داده‌ها در سه روش نشان داده شده است. همان‌طور که در شکل نشان داده شده مقدار متوسط تطابق داده‌ها در روش VE از روش PDS بیشتر است و هم‌چنین این مقدار برای روش PDS از روش استاندارد Q بیشتر است.

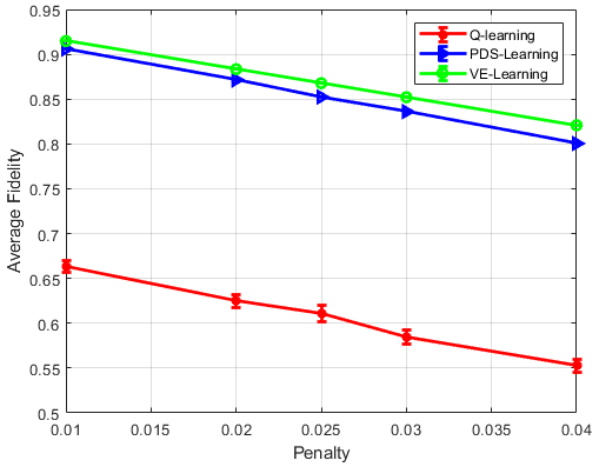
مقدار متوسط تعداد بسته‌هایی هدر رفت در شکل ۶ برای روش استاندارد Q از دو روش دیگر بیشتر است.

در این کار پارامترهای شبیه‌سازی تغییر داده شده و تأثیر آن‌ها را بر روی مقدار متوسط تطابق داده‌ها، مقدار متوسط انرژی مصرفی و هم‌چنین متوسط نسبت بسته-های هدر رفت در هر روش ، مقایسه شده‌اند. آزمایش‌ها را برای هر نقطه ۵۰ بار با تعداد تکرار یک‌میلیون انجام شده است و هم‌چنین فاصله اطمینان با دقت ۹۵ درصد نیز در نظر گرفته شده است.



شکل۶- متوسط تعداد بسته‌های هدر رفت در روش‌های Q، PDS و VE

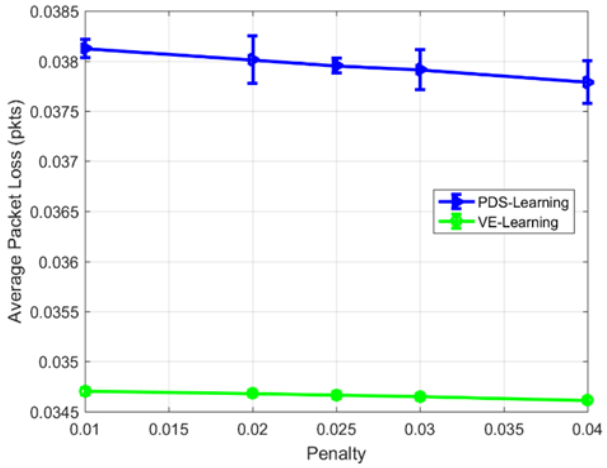
همان‌طور که در شکل ۷ نشان داده شده است در این نمودار تأثیر مقدار جریمه p_4 که برای تابع پاداش در نظر گرفته بودیم را نشان می‌دهد هر چه مقدار این جریمه افزایش یابد مقدار متوسط تطابق داده‌ها در هر سه روش نیز کاهش می‌یابد.



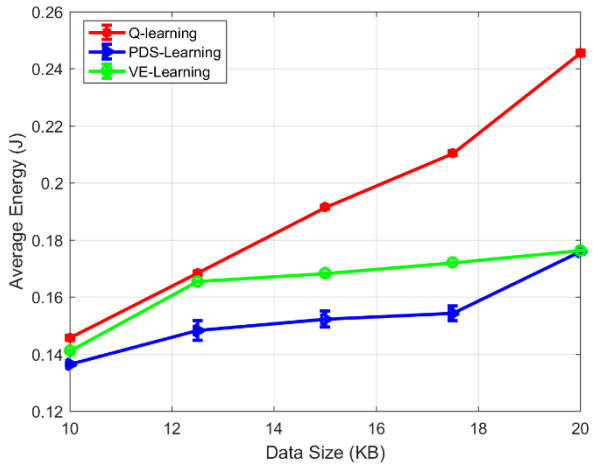
شکل۷- اثر افزایش جریمه p_4 در رابطه تابع پاداش بر متوسط تطابق داده‌ها

در شکل ۸ تأثیر افزایش مقدار جریمه p_4 بر نسبت تعداد بسته‌های هدر رفت در دو روش PDS و VE را ملاحظه می‌کنید. با افزایش این مقدار جریمه نسبت بسته‌های هدر رفت کاهش می‌یابد.

برای بررسی تأثیر حجم داده اولیه بر متوسط مصرف انرژی ، شبیه‌سازی را مطابق با مقادیر ذکرشده در جدول ۱، برای حجم‌های مختلف داده‌های اولیه با مقادیر ۱۰،۱۲،۵،۱۵،۱۷،۵،۲۰ هزار بایت هر سه الگوریتم اجرا شده است. همان‌طور که شکل ۹ نشان می‌دهد با افزایش حجم داده‌ها، متوسط انرژی مصرفی در هر سه روش افزایش می‌یابد.

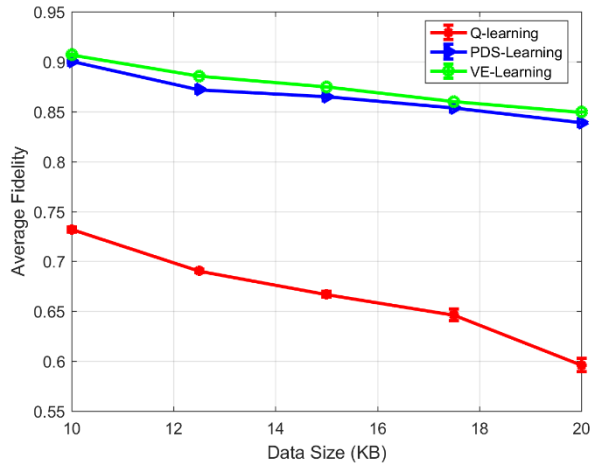


شکل۸- اثر افزایش جریمه p_4 در رابطه تابع پاداش بر نسبت تعداد بسته‌های هدر رفت



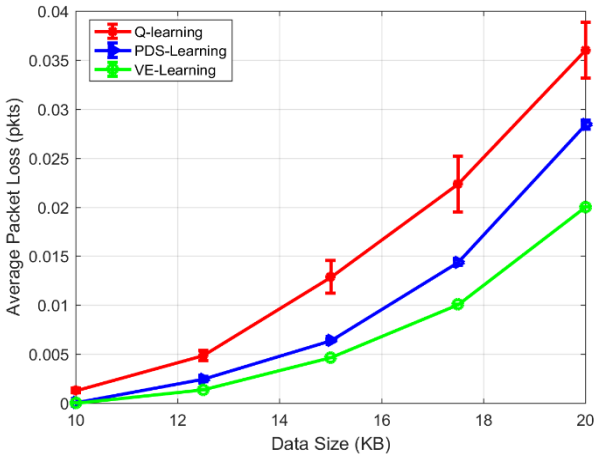
شکل۹- تأثیر افزایش حجم داده ورودی بر متوسط مصرف انرژی

شکل ۱۰ نشان می‌دهد که متوسط تطابق داده‌ها در گیرنده با افزایش مقدار حجم داده کاهش می‌یابد و هم‌چنین روش مبتنی بر یادگیری VE از روش استاندارد Q-learning بهتر این هدف را فراهم می‌آورد.



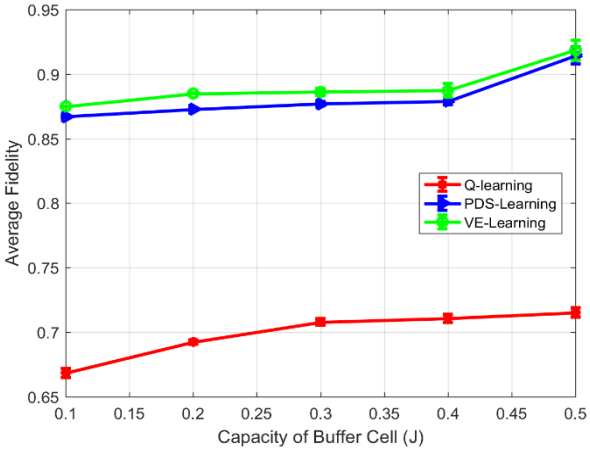
شکل۱۰- تأثیر افزایش حجم داده ورودی بر متوسط تطابق داده‌ها

در شکل ۱۱ تأثیر افزایش حجم داده وردی بر میزان هدر رفتن بسته‌ها نشان داده شده است. این شکل نشان می‌دهد که میزان هدر رفت بسته‌ها در روش VE کمتر از روش PDS و روش استاندارد Q است و هم‌چنین با افزایش سایز داده مقدار هدر رفتن بسته‌ها در هر سه روش افزایش می‌یابد.

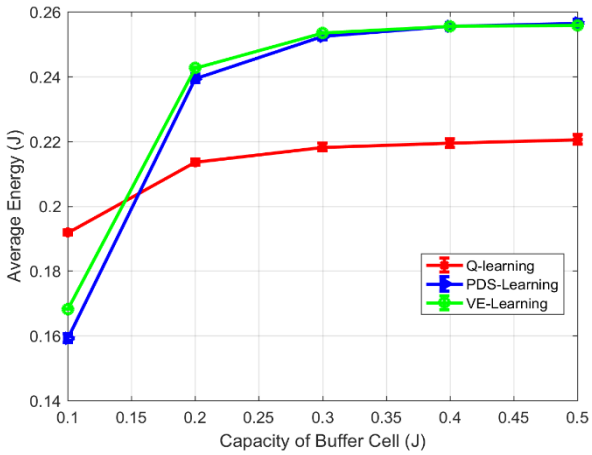


شکل ۱۱- تأثیر افزایش حجم داده ورودی بر متوسط میزان هدر رفت بسته‌ها

برای بررسی تأثیر گنجایش هر خانه از بافر انرژی بر متوسط مصرف انرژی، شبیه‌سازی را مطابق با مقادیر ذکرشده در جدول۱، با یک‌میلیون تکرار برای بافر انرژی با گنجایش‌های متفاوت ۰,۱، ۰,۲، ۰,۳، ۰,۴و ۰,۵ ژول کوانتوم هر خانه بافر اجرا شده است.



شکل۱۲- تأثیر افزایش ظرفیت هر خانه بافر بر متوسط تطابق داده‌ها

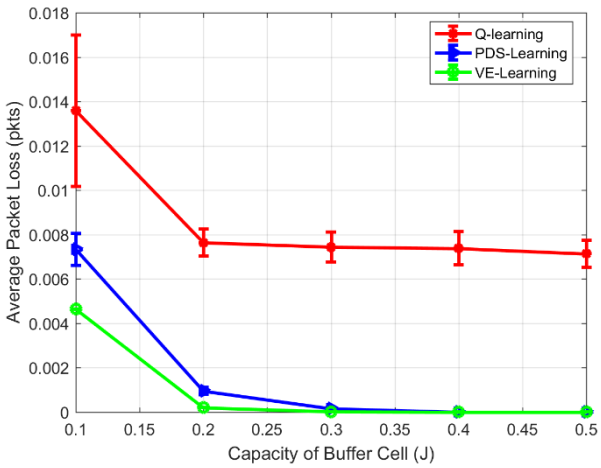


شکل۱۳- تأثیر افزایش ظرفیت هر خانه بافر انرژی بر متوسط مصرف انرژی

در شکل ۱۲ نشان داده شده است که با افزایش ظرفیت هر خانه بافر مقدار متوسط تطابق داده‌ها در بلندمدت در هر سه روش افزایش می‌یابد و در روش VE نیز متوسط تطابق داده‌ها از سایر روش‌ها بیشتر است.

در شکل ۱۳ میزان متوسط مصرف انرژی با افزایش مقدار ظرفیت هر خانه از بافر انرژی را نشان می‌دهد. که در این نمودار نیز مصرف انرژی در هر سه روش افزایش می‌یابد.

همان‌طور که در شکل ۱۴ نشان داده‌شده است میزان متوسط هدر رفت بسته‌ها در هر سه روش با افزایش ظرفیت هر خانه از بافر انرژی کاهش می‌یابد و برای روش VE این کاهش محسوس‌تر خواهد بود.



شکل۱۴- تأثیر افزایش ظرفیت هر خانه از بافر انرژی بر متوسط هدر رفت بسته‌ها

۵- نتیجه‌گیری

در پژوهش جاری، در راستای افزایش تطابق داده‌های گزارش شده ما مسئله کنترل توأم نرخ فشرده‌سازی (با اتلاف) و تعداد بسته‌های ارسالی در واحد زمان را برای یک گره اینترنت اشیاء مجهز به منبع انرژی تجدید پذیر مطرح کردیم؛ و آن را به‌صورت یک مسئله بهینه‌سازی تصادفی با هدف بیشینه کردن متوسط تطابق داده‌های گزارش شده در بلندمدت، ضمن ایجاد محدودیت در متوسط تأخیرِ گزارش رویدادهای حسگری مورد بررسی قرار دادیم. راهکارهای ارائه شده در این پژوهش در مقایسه با پژوهش‌های پیشین به بهبود کارایی الگوریتم‌های پیشنهادی نسبت به الگوریتم استاندارد Q-learning و تأمین آستانه تأخیر به‌عنوان قید کیفیت سرویس دست یافته است و همچنین مقایسه و اندازه‌گیری میزان هدر رفت بسته‌ها در الگوریتم‌های پیشنهادی PDS و VE نسبت به روش‌های استاندارد یادگیری تقویتی است.

۶- مراجع

[1] R. Khan, S. Ullah Khan, R. Zaheer, and S. Khan, “Future Internet: The Internet of Things Architecture,Possible Applications and Key Challenges”, *in the proceedings of 10th International Conference on Frontiers of Information Technology, Islamabad, Pakistan*, 17-19 December, 2012.

[2] R. Sharma, “A data compression application for wireless sensor networks using LTC algorithm,” IEEE International Conference on Electro/Information Technology (EIT), pp. 598–604, 2015.

[3] C. J. Deepu, C.-H. Heng, and Y. Lian, "A hybrid data compression scheme for power reduction in wireless sensors for IoT," *IEEE transactions on biomedical circuits and systems*, vol. 11, pp. 245-254, 2017.

[4] M. J. Kang, S. Jeong, I. Yoon, and D. K. Noh, “Energy-aware determination of compression for low latency in solar-powered wireless sensor networks,” International Journal of Distributed Sensor Networks, vol. 13, no. 2, p. 155014771769416, Feb. 2017.

[5] S. Kim, C. Cho, K.-J. Park, and H. Lim, "Increasing network lifetime using data compression in wireless sensor networks with energy harvesting," *International Journal of Distributed Sensor Networks*, vol. 13, no. 1, p. 1-10, Jan. 2017.
[6] D. Zordan, T. Melodia, and M. Rossi, "On the Design of Temporal Compression Strategies for Energy Harvesting Sensor Networks," *IEEE Transactions on Wireless Communications,* vol. 15, no. 2, pp. 1336–1352, Feb. 2016.
[7] M. A. Razzaque, C. Bleakley, and S. Dobson, "Compression in wireless sensor networks: A survey and comparative evaluation," *ACM Transactions on Sensor Networks (TOSN)*, vol. 10, no. 1, p. 5, 2013.
[8] C. Pielli, A. Biason, A. Zanella, and M. Zorzi, "Joint optimization of energy efficiency and data compression in TDMA-based medium access control for the IoT," *in Globecom Workshops (GC Wkshps),* 2016 IEEE, pp. 1–6, 2016.
[9] R. V. Bhat, M. Motani, and T. J. Lim, "Distortion minimization in energy harvesting sensor nodes with compression power constraints," in *2016 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2016.
[10] H. Ghasemi, I. Stupia, and L.Vandendorpe, "Optimal Compression and Transmission Policies for Energy Harvesting Nodes," *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, 2018.
[11] C. Tapparello, O. Simeone, and M. Rossi, "Dynamic Compression-Transmission for Energy-Harvesting Multihop Networks With Correlated Sources," *IEEE/ACM Transactions on Networking*, vol. 22, no. 6, pp. 1729–1741, Dec. 2014.
[12] P. Castiglione, O. Simeone, E. Erkip, and T. Zemen, "Energy Management Policies for Energy-Neutral Source-Channel Coding," *IEEE Transactions on Wireless Communications.*, vol. 60, no. 9, pp. 2668–2678, Sep. 2012.
[13] A. Biason, C. Pielli, A. Zanella, and M. Zorzi, "Energy/distortion tradeoffs in joint source coding and MAC scheduling for the IoT," *arXiv:1702.03695, submitted to IEEE Trans. on Wireless Communications,* Nov. 2016.
[14] A. Biason, C. Pieli, A. Zanella, and M. Zorzi, "Access Control for IOT Nodes With Energy And Fidelity Constraints," *IEEE Transactions on Wireless Communications,* vol. 17, no.5, pp. 3242-3257, May. 2018.
[15] C. Pielli, C. Stefanovic, P. Popovski, and M. Zorzi, "Joint Retransmission, Compression and Channel Coding for Data Fidelity under Energy Constraints," *arXiv Prepr. arXiv1706.09183*, pp. 1–14, Jun. 2017.
[16] N. Toorchi, J. Chakareski, and N. Mastronarde, "Fast and low-complexity reinforcement learning for delay-sensitive energy harvesting wireless visual sensing systems," in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 1804–1808, 2016.
[17] C. Pielli, C. Stefanovic, P. Popovski, and M. Zorzi, "Minimizing Data Distortion of Periodically Reporting IoT Devices with Energy Harvesting*," in 2017 14th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pp. 1– 9, 2017.
[18] P. Bertsekas, Dynamic Programming and Optimal Control, 4th ed. Athena Scientific, vol, 2, 2012.
[19] C. J. C. H. Watkins and P. Dayan, "Q-learning," Machine Learning, vol. 8, no. 3-4, pp. 279–292, 1992.
[20] T. Schoellhammer, B. Greenstein, E. Osterweil, M. Wimbrow, and D. Estrin, "Lightweight temporal compression of microclimate datasets [wireless sensor networks]," in Annual IEEE International Conference on Local Computer Networks, Tampa, FL, US, Nov. 2004, pp. 516– 524.
[21] D. Zordan, B. Martinez, I. Vilajosana, and M. Rossi, "On the Performance of Lossy Compression Schemes for Energy Constrained Sensor Networking," *ACM Transactions on Sensor Networks*, vol. 11, no. 1, pp. 15:1–15:34, Aug. 2014.
[22] R. Wang, J. Zhang, S. H. Song and K. B. Letaief, "Optimal QoS-Aware Channel Assignment in D2D Communications With Partial CSI," IEEE Transactions on Wireless Communications, vol. 15, no. 11, pp. 7594-7609, Nov. 2016.
[23] N. Salodkar, A. Bhorkar, A. Karandikar, and V. S. Borkar, "An on-line learning algorithm for energy efficient delay constrained scheduling over a fading channel," *IEEE Journal on Selected Areas in Communications,* vol. 26, no. 4, 2008
[24] R. Aslani, V. Hakami, and M. Dehghan, "A token-based incentive mechanism for video streaming applications in peer-to-peer networks," *Multimedia Tools Applications,* pp. 1-29, 2017.
[25] M. Crowder. Stochastic approximation: A dynamical systems viewpoint by Vivek Borkar. International Statistical Review, 77(2), 2009.
[26] J. Davidson, Stochastic limit theory An introduction for econometricians. Oxford: Oxford University Press, 1994.

فرنوش نامجونیا مدرک کارشناسی خود را در سال ۱۳۹۲ از دانشگاه ایلام اخذ کرده‌اند. ایشان اکنون دانشجوی کارشناسی ارشد شبکه‌های کامپیوتری دانشگاه علم و صنعت ایران می‌باشد. زمینه‌های تحقیقاتی ایشان فشرده سازی داده‌ها و یادگیری تقویتی می‌باشد.

آدرس پست الکترونیکی ایشان عبارت است از:

f_namjonia@comp.iust.ac.ir

وصال حکمی مدرک کارشناسی و کارشناسی ارشد و دکترا خود را به ترتیب در سال‌های ۱۳۸۳، ۱۳۸۷، ۱۳۹۴ از دانشگاه صنعتی امیرکبیر اخذ کرده‌اند. ایشان اکنون عضو هیئت‌علمی دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت ایران هستند. زمینه‌های تحقیقاتی ایشان بهینه سازی سیستماتیک کارایی شبکه‌های کامپیوتری و سیستم‌های توزیعی ،رویکردهای بهینه‌سازی غیرمتمرکز و تصادفی، نظریهٔ بازی‌ها، یادگیری تقویتی، یادگیری چند-عاملی است.

آدرس پست الکترونیکی ایشان عبارت است از:

vhakami@iust.ac.ir

[1] Multi-hop Wireless Sensor Networks
[2] Lightweight Temporal Compression
[3] Post Decision State
[4] Virtual Experience
[5] Finite State Markov Chain

The CSI Journal on Computing Science and Information Technology

The CSI Journal on
Computing Science and
Information Technology
Vol. 17, No.2, 2020

Control the Joint Compression and Data Transmission in IoT Equipment with Renewable Energy.

Farnoosh Namjoonia, Vesal Hakami

School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran

Abstract

One of the important challenges in Internet of Things (IoT) is device energy limitation. To reduce the energy consumption, in this paper, we propose an approach to control the joint compression rate (with loss) and the number of transmission packets per second, for an IoT node that is equipped with renewable energy resources. The proposed method focuses on two optimization goals simultaneously, that are considering fidelity of the received data with the original data as well as satisfying the data transmission delay's constraints. To reach these goals, we use Constrained Markov Decision Process (CMDP) to design a stochastic optimization problem to maximize the expected value of fidelity in long term subject to the constraint of average delay of reporting sensor events. The standard Lagrangian technique is applied to make the problem unconstrained. Our proposed approach for calculating adaptive optimal policy is based-on two accelerated reinforcement-learning algorithms that are called Post Decision State (PDS) and Virtual Experience (VE). These algorithms can guarantee the convergence to the optimal policy by separating the system dynamics to known and unknown sections, only by taking a greedy decision without any statistical knowledge of wireless channel stochastic processes, energy harvesting, and sensor event occurrence. To evaluate the novel approach performance, we compare it with the standard Q-learning algorithm in terms of energy consumption, data packet loss, and data fidelity. Consequently, the results demonstrate that the VE, 63.741% and PDS 61.845% improve data fidelity in comparison to standard Q-learning algorithm.

Keywords: Compression, constrained Markov decision process, data fidelity, delay constraint, energy harvesting, energy optimization, Internet of Things, PDS reinforcement learning, VE reinforcement learning