مروری بر روش‌های کنترل و مسیریابی در شبکه‌های حسگر بی‌سیم با بهره‌گیری از یادگیری تقویتی

علی فرقانی اله آبادی[۱]، محمدرضا بینش مروستی[۲]، سید امیر اصغری توچائی[۳]

[۱]دانشجوی کارشناسی ارشد مهندسی کامپیوتر- گرایش هوش مصنوعی و رباتیکز، دانشگاه خوارزمی تهران، std_forghani@khu.ac.ir

[۲]استادیار دانشکده برق و کامپیوتر، دانشگاه خوارزمی تهران، marvasti@khu.ac.ir

[۳]دانشیار دانشکده برق و کامپیوتر، دانشگاه خوارزمی تهران، asghari@khu.ac.ir

چکیده

این مقاله به بررسی روش‌های کنترل و مسیریابی در شبکه‌های حسگر بی‌سیم می‌پردازد که با بهره‌گیری از تکنیک یادگیری تقویتی که یکی از روش‌های یادگیری ماشین به شمار می‌رود توانسته‌اند بازدهی انرژی را در شبکه‌های حسگر بی‌سیم بهبود بخشند. این تکنیک بر مبنای روش پاداش و تنبیه که رفتاری مشابه فرآیند یادگیری در کودکان دارد، بنا نهاده شده است. همواره مدیریت صحیح انرژی و به سبب آن افزایش طول‌عمر در شبکه‌های حسگر بی‌سیم یکی از چالش‌های اصلی این نوع از شبکه ها به دلیل محدودیت انرژی در گره‌های آن به شمار می‌آید. هدف از نگارش مقاله حاضر، آشنایی بیشتر با روش‌های ارائه شده مربوطه می‌باشد. در این مقاله روش‌های مختلف که سعی بر استفاده از فرآیند یادگیری تقویتی در جهت بهبود رفتار شبکه‌های حسگر بی‌سیم و هوشمندتر شدن آن‌ها را دارند معرفی شده و مورد بررسی قرار می‌گیرند. همچنین سیر تکامل این روش‌ها و نسبت برتری هر یک به دیگری مورد بررسی قرار گرفته است.

کلیدواژه

یادگیری تقویتی، شبکه حسگر بی‌سیم، محدودیت انرژی، طول عمر، پاداش و تنبیه

۱ مقدمه

امروزه می‌توان شبکه‌های حسگر بی‌سیم را به عنوان یکی از پرکاربردترین روش‌های جمع‌آوری و تحلیل داده‌های محیطی دانست. به علت محدودیت انرژی گره‌های حسگر در این شبکه‌ها، استفاده از روشی بهینه جهت مسیریابی و کنترل شبکه‌های حسگر بی‌سیم می‌تواند در افزایش بازدهی انرژی راه‌گشا باشد[۱].

نوآوری پژوهش مروری حاضر عبارت است از گردآوری روش های کلاسیک و نوین کنترل و مسیریابی شبکه های حسگر بی‌سیم و بررسی جزئیات کارکردی هر یک به همراه اشاره به نقاط قوت و ضعف و همچنین حیطه کاربرد هر یک از روش های مذبور به شکلی است که ذهن خواننده را با استفاده از ساختار دهی صحیح آماده ادامه پژوهش های لازم در این حوزه، و یا استفاده از یکی از تکنیک های ارائه شده در چالش های خود گرداند.

در ادامه به ارائه مقدمه ای از دو حوزه شبکه های حسگر بی‌سیم و یادگیری تقویتی خواهیم پرداخت.

۱.۱ شبکه‌های حسگر بی‌سیم

شبکه حسگر بی‌سیم که از آن به عنوان روشی جهت جمع‌آوری و تحلیل داده‌های محیطی استفاده می‌شود، دارای کاربرد وسیعی در حوزه‌هایی همچون کشاورزی، پزشکی، صنعتی و نظامی و همچنین نظارت بر محیط های دور از دسترس است. از این جهت است که افزایش بازدهی انرژی و در نتیجه آن طول عمر شبکه حائز اهمیت است. این شبکه‌ها دارای تعدادی گره هستند. هر گره خود به تنهایی یک ماژول سخت‌افزاری است که شامل حسگر، باتری، حافظه و پردازنده است. انرژی هر گره با استفاده از باتری آن تامین شده، داده‌های محیطی با استفاده از حسگرها اندازه‌گیری شده و اطلاعات مورد نیاز در حافظه آن ذخیره شده و محاسبات مورد نیاز توسط پردازنده آن انجام شده و از طریق ماژول ارتباطی موجود با سایر گره‌ها ارتباط برقرار نموده و در نهایت اقدام به تبادل اطلاعات با آن‌ها می‌نماید. گره اصلی در این شبکه‌ها که چاهک[۱] نامیده می‌شود در نهایت با دریافت تمامی نتایج و تجمیع[۲] آن‌ها می‌تواند به یک دید کلی از پارامتر-

1 Sink

۲ Aggregation

تاریخ دریافت: ۱۳۹۹/۰۴/۲۴ تاریخ پذیرش: ۱۴۰۲/۰۸/۲۸

های محیطی مورد نظر برسد تا بتوان تحلیل‌های مورد نیاز را بر اساس آن‌ها انجام داد[۲].

حال لازم است به معرفی سه تکنیک پرکاربرد در پروتکل‌های کنترلی شبکه‌های حسگر بی‌سیم شامل خوشه‌بندی، تجمیع و فشرده‌سازی بپردازیم. در تکنیک خوشه‌بندی به طور کلی در هر منطقه یک گره به عنوان گره نماینده انتخاب می‌گردد و اطلاعات گره‌های اطراف هر گره نماینده، در داخل آن تجمیع شده و به گره اصلی ارسال می‌شود تا تبادل اطلاعات از طریق گره‌های نماینده صورت گیرد. با در نظر گرفتن یک روال خوشه‌بندی درست، این تکنیک می‌تواند تاثیر بسزایی در افزایش کارایی و بهره‌وری شبکه و همچنین افزایش بازدهی انرژی مصرفی، داشته باشد. تکنیک‌های تجمیع و فشرده‌سازی باعث کاهش حجم داده‌های ارسالی و در نتیجه ترافیک شبکه شده و باعث می‌شود که نیاز به حجم ذخیره‌سازی بالا در تمام گره‌ها وجود نداشته باشد. در نتیجه هزینه‌های ارتباطی کاهش یافته و بازدهی مصرف انرژی افزایش می‌یابد[۳-۶].

۱.۲ یادگیری تقویتی

یادگیری تقویتی یکی از روش‌های یادگیری دارای کاربرد در این حوزه به شمار می‌رود. این روش که بر مبنای فرآیند تصمیم گیری مارکوف[3] بنا نهاده شده است با در نظر گرفتن یک پاداش برای هر عمل سعی دارد مقدار مناسبی از پاداش (یا تنبیه) را به موجود هوشمند اختصاص دهد. تا موجود هوشمند یاد بگیرد که در هر موقعیت بهتر است چه عملی را انجام دهد تا در نهایت مجموع پاداش اکتسابی توسط موجود هوشمند بیشینه گردد. در این روش پارامتری با نام مؤلفه تخفیف[4] وجود دارد که دارای مقداری بین صفر و یک بوده و بیان کننده این است که همواره پاداش مراحل فعلی باید بیش از پاداش مراحل آینده بر روی محاسبه پاداش کل، اثرگذار باشد. این مورد در رابطه (۱) قابل مشاهده است[۷-۱۸].

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \gamma^3 r_{t+4} + \cdots = \sum_{i=t}^{\infty} \gamma^{i-t} r_{i+1} \quad (۱)$$

که در آن R_t میزان پاداش دریافتی از زمان t بر اساس پاداش های آینده است. r_{t+1} بیانگر پاداش دریافتی در زمان t+1 و γ مؤلفه تخفیف است.

حال به ذکر تعاریف مورد استفاده در این نوع از یادگیری می‌پردازیم. این تعاریف مواردی شامل تابع سیاست، تابع ارزش وضعیت‌ها، تابع ارزش وضعیت-عمل، معادلات بهینگی بلمن، روش تکرار ارزش و روش تکرار سیاست را در بر می‌گیرد:

- **تابع سیاست**: این تابع در مسائل یادگیری تقویتی بیانگر سیاست تصمیم‌گیری است که به ما اعلام می‌کند که در هر وضعیت مسئله، اتخاذ تصمیم برای چه عملی، احتمالا پاداش کلی دریافتی را بیشینه خواهد نمود. درواقع تابع سیاست نگاشتی است از هر وضعیت محیط به عملی که لازم است در آن وضعیت انجام شود تا پاداش دریافتی نهایی را بیشینه نماید[۷].
- **تابع ارزش وضعیت‌ها**: تابع ارزش، بیان کننده آن است که به ازای هر وضعیت در مسئله، امید ریاضی مجموع پاداش دریافتی به چه میزان خواهد بود. این مقدار توسط رابطه (۲) قابل بیان است[۷]:

$$V^{\pi}(s_t) = \mathbb{E}_{E.\pi}[R_t|s_t] \quad (۲)$$

که در آن $V^{\pi}(s_t)$ نشان دهنده ارزش وضعیت s در زمان t بر اساس سیاست π و $\mathbb{E}_{E.\pi}[R_t|s_t]$ امید ریاضی پاداش اکتسابی از زمان t و با حضور در وضعیت s است[۷].
- **تابع ارزش وضعیت-عمل**: این تابع با دریافت یک وضعیت مسئله و عملی قابل انجام در آن وضعیت، بیان می‌دارد که امید ریاضی مجموع پاداش دریافتی در وضعیت فعلی چقدر است. این تابع در قالب رابطه (۳) قابل بیان است[۷].

$$V^{\pi}(s_t, a_t) = \mathbb{E}_{E.\pi}[R_t|s_t = s, a_t = a] \quad (۳)$$

که در آن $V^{\pi}(s_t, a_t)$ میزان ارزش انجام عمل a در وضعیت s و در زمان t است. همچنین $\mathbb{E}_{E.\pi}[R_t|s_t = s, a_t = a]$ امید ریاضی پاداش اکتسابی به ازای یک وضعیت و انجام یک عمل خاص است.

- **معادلات بهینگی بلمن**: این معادلات بیان می‌دارد که بهترین سیاست، سیاستی است که به ازای آن، ارزش هر یک از وضعیت‌ها بیشینه شود. با استفاده از برنامه‌سازی پویا اثبات شده است که ارزش یک وضعیت در مسئله به ازای یک سیاست خاص، با ارزش هر یک از وضعیت‌ها در آن مسئله در زمان آینده،

[3] Markov Decision Process

[4] Discount Factor

ارتباط مستقیم دارد. معادله مذکور در رابطه (۴) آمده است:

$$V^{\pi}(s_t) = \mathbb{E}_{E.\pi}[r_{t+1} + \gamma V^{\pi}(s_{t+1})] \quad (٤)$$

جهت حل معادلات بهینگی بلمن، از تکنیک برنامه‌سازی پویا در قالب دو روش تکرار ارزش و تکرار سیاست بهره گرفته می‌شود و بر اساس آن سیاست بهینه نهایی به دست خواهد آمد.

در طی روش تکرار ارزش، در ابتدا مقدار تابع ارزش بر اساس یک توزیع تصادفی به دست آمده و به ازای هر وضعیت، مقدار تابع ارزش وضعیت-عمل محاسبه شده و هر بار مقدار تابع ارزش آن وضعیت، برابر با مقدار بیشینه تابع ارزش وضعیت-عمل به ازای آن وضعیت و عمل خاص، قرار داده می‌شود. این عملیات آنقدر تکرار می‌گردد تا بر اساس روش ارزیابی سیاست، تابع ارزش به حالت بهینه خود برسد.

روش تکرار سیاست در دو مرحله ارزیابی سیاست و بهبود سیاست انجام می‌شود. در واقع هر بار به وسیله روش ارزیابی سیاست، سیاست فعلی ارزیابی شده و اگر بهینه نباشد، از روش بهبود سیاست جهت اصلاح آن کمک گرفته شده تا در نهایت سیاست بهینه نهایی به دست آید[۷].

روش بهبود سیاست به این صورت است که در ابتدا یک سیاست تصادفی را به عنوان سیاست فعلی در نظر می‌گیرد و هر بار در طی روال تکرار ارزش، با محاسبه تابع ارزش بر مبنای آن سیاست، بررسی می‌شود که آیا این تابع، بهینه است یا خیر. در صورتی که تابع ارزش بهینه نبود، سیاست بهبود یافته را به دست آورده و مجددا تابع ارزش به ازای سیاست جدید به دست آمده و بهینه بودن آن بررسی می‌شود[۷].

۲ روش‌های گذشته

در مورد روند افزایش کارایی و طول عمر شبکه‌های حسگر بی-سیم، زمانی می‌توان میزان بازدهی را در این شبکه‌ها افزایش داد که یک روال صحیح تصمیم‌گیری در مورد فرآیندهای موجود در شبکه‌های حسگر بی‌سیم و از جمله مهم‌ترین آن‌ها، فرآیند مسیریابی، ارائه شود. بینش روش‌های گذشته بر خلاف روش-هایی همچون یادگیری تقویتی بر روال‌هایی استوار بود که سعی نداشتند رفتار خود را نسبت به گذشته بهبود دهند و در طی این بهبود از تجربیات و نتایج بهره گیرند. در یک دید کلی دسته بندی روش‌های کنترل و مسیریابی شبکه‌های حسگر بی‌سیم در شکل ۱ قابل مشاهده است[۲۳].

شکل ۱. دسته‌بندی پروتکل‌های مسیریابی در شبکه‌های حسگر بی‌سیم

۲.۱ پروتکل‌های داده محور[5]

در بسیاری از کاربردهای شبکه حسگر بی‌سیم به علت وجود تعداد بالای گره‌های حسگر، اختصاص یک شناسه سراسری به هر یک از آن‌ها امکان پذیر نیست. این در حالی است که به طور معمول گره‌های حسگر به طور تصادفی در محیط پخش می‌شوند و این روال، انتخاب گره‌های منتخب برای ارسال داده و ارتباط با پروتکل را دشوار می‌کند. این مشکل باعث می‌شود به علت محاسبه دشوارتر در جهت انتخاب این گره‌ها، داده‌ها با افزونگی قابل توجهی منتقل شوند که این خود موجب ناکارآمد بودن از منظر مصرف انرژی می‌شود. پروتکل‌های مسیریابی که قادرند مجموعه‌ای از گره‌های حسگر را انتخاب کرده و از تکنیک تجمیع در طی ارسال داده‌ها استفاده نمایند را به علت آن که از مقادیر داده‌های دریافتی حسگرها جهت رسیدن به روالی بهینه در ارسال داده‌ها توسط الگوریتم تجمیع، بهره گرفتهاند، پروتکل‌های داده محور می‌نامند. در این پروتکل‌ها گره مقصد، یک روال پرس و جو را به ناحیه مورد نظر ارسال نموده و در انتظار دریافت داده آن منطقه می‌ماند. دسته‌بندی روش‌های گذشته داده محور در شکل ۲ قابل مشاهده است[۲۳].

شکل ۲. پروتکل‌های داده‌محور مسیریابی در گذشته

در روش داده محور از طریق مذاکره[6] یک روال دست‌دادن سه ایستگاهی ایجاد شده و با در نظر گرفتن یک حد آستانه برای

[5] Data-centric protocols

[6] SPIN: Sensor Protocols for Information via Negotiation

انرژی هر گره، در صورتی که انرژی از این حد آستانه در یک گره، پایین تر بیاید، آن گره مشارکت خود را در ارسال و دریافت داده ها به منظور حفظ انرژی جهت افزایش طول عمر شبکه، کاهش می‌دهد[۲۴,۲۳].

در روش انتشار جهت‌دار[۷]، هر داده بر مبنای زوج‌های مقدار و ویژگی تعریف شده و گره‌ها به دنبال دریافت داده‌هایی نزدیک‌تر به زوج‌های مقدار ویژگی مورد جستجوی خود هستند. این روش از تغییرات ساختاری شبکه پشتیبانی نموده و می‌تواند بیش از یک مسیر را جهت ارسال داده پیشنهاد دهد تا در صورت خرابی یک مسیر بتوان بدون انجام محاسبات مجدد از مسیر بعدی استفاده نمود. این روش نمی‌تواند برای کلیه برنامه‌های شبکه حسگر اعمال شود. زیرا این مدل مبتنی بر پرس و جو است و برنامه‌های کاربردی که نیاز به تحویل مداوم داده‌ها به گره مقصد دارند، با استفاده از این روش، کارایی مناسبی نخواهند داشت [۲۳] [۲۶-۲۵].

در روش مسیریابی آگاه از انرژی[۸]، سه مرحله تنظیم، انتقال و نگهداری در نظر گرفته شده است. این روش در قالب این سه مرحله، در نهایت تنها یک مسیر را در هر لحظه، بر اساس تابع احتمال به دست آمده، کشف می‌نماید[۲۷,۲۳].

کاربرد روش مسیریابی بر مبنای شایعه[۹]، در مواردی است که در آن معیارهای مسیریابی جغرافیایی کاربردی نیستند. در بسیاری از موارد، تنها وقایع محدود و مشخصی در طی فعالیت شبکه به وجود می‌آیند و تعداد رویدادها اندک و تعداد پیام‌های ارتباطی زیاد است. بنابراین استفاده از یک روال سیل آسا که موجب صرف انرژی بیشتر می‌شود، غیرضروری است. بررسی‌ها نشان داده است در هنگامی که تعداد رویدادها اندک باشد، این روش موجب صرفه‌جویی قابل توجه در انرژی خواهد شد[۲۸,۲۳].در روش مسیریابی بر مبنای گرادیان[۱۰]، هر گره با کشف حداقل فاصله تا گره مقصد که ارتفاع آن گره نامیده می‌شود و سنجش نرخ تغییر این مقدار نسبت به همسایگان گره تا رسیدن به گره مقصد، گرادیان مسیرهای مختلف را تا رسیدن به مقصد سنجیده و آن گره‌ای در لیست گره‌های همسایه که واسطه مسیری است که دارای گرادیان بزرگتری است (با شیب بیشتری داده را به گره مقصد می‌رساند) به عنوان گره حامل انتخاب می‌گردد. گره‌هایی که به عنوان رله برای چندین مسیر عمل می‌کنند می‌توانند اقدام به تجمیع داده‌های دریافتی نمایند. در این روش علاوه بر آن که می‌توان از تجمیع بهره گرفت، همچنین می‌توان از تکنیک گسترش ترافیک به منظور متعادل‌سازی ترافیک به صورت یکنواخت در شبکه، استفاده نمود. این روش دارای سه رویکرد برای پخش اطلاعات شامل انتخاب تصادفی، بر پایه انرژی و بر مبنای جریان است. روش‌های مذکور سعی بر توزیع یکنواخت ترافیک در تمام سطح شبکه به منظور متعادل‌سازی بار گره‌ها و در نتیجه افزایش طول عمر شبکه دارند. این روندهای متوازن-سازی بار ترافیکی و همچنین تکنیک تجمیع در سایر پروتکل‌های مسیریابی نیز کاربردی هستند. بر طبق نتایج به دست آمده، پروتکل مسیریابی بر مبنای گرادیان از لحاظ مصرف انرژی و افزایش طول عمر شبکه نسبت به روش انتشار جهت‌دار، کارایی بالاتری دارد[۲۹,۲۳].

روش انتشار ناهمسانگرد محدود[۱۱]، در کنار بهره‌گیری از مقدار هزینه مسیر در کنترل شبکه، با در نظر گرفتن بهره اطلاعاتی سعی بر افزایش کارایی پروتکل مسیریابی شبکه حسگر دارد. این روش تنها حسگرهایی را که نزدیک به رویدادهای مشخصی هستند فعال نموده و اقدام به تنظیم مسیرهای داده به صورت دینامیکی می‌نماید. همچنین ضمن متعادل کردن هزینه، مفیدترین اطلاعات را ارائه داده و در واقع اقدام به تجمیع داده‌ها می‌نماید. این تکنیک سعی بر ارائه ترتیب پیمایش گره‌ها در مسیر به گونه‌ای دارد که بهره اطلاعاتی بیشینه شود[۳۰,۲۳].

ایده اصلی روش مسیریابی با استفاده از انتخاب گره نماینده[۱۲]، انتخاب گره‌هایی از بین گره‌های شبکه به عنوان نماینده است که تنها این گره‌ها با ایستگاه اصلی ارتباط داشته و اقدام به تجمیع داده سایر گره‌ها و سنجش آن نمایند و اگر میانگین تغییرات داده‌ها در حاصل تجمیع هر گره نماینده، از آستانه مشخصی بیشتر شد آن را به ایستگاه اصلی ارسال کنند. این رویکرد موجب صرفه جویی در میزان انرژی مصرفی شده است[۳۱,۲۳].

روش ارسال پرس و جو فعال در شبکه حسگر[۱۳]، سعی می‌کند با تعریف پارامتری، ناحیه ارسالی موارد به روز رسانی را در محدوده با قابلیت ارسال به حداکثر گام مشخصی، محدود نماید. در صورتی که این پارامتر برابر با اندازه شبکه در نظر گرفته شود، رویکرد این روش معادل رویکردهای ارسال سیل‌آسا خواهد بود. در این روش هر گره دریافت کننده پرس‌وجو، سعی می‌کند با استفاده از اطلاعات از قبل دریافت شده خود، بخشی از آن را پاسخ دهد و آن را به حسگر دیگری منتقل کند. اگر اطلاعات ذخیره شده به روز نباشند گره‌ها اطلاعاتی را از همسایگان خود دریافت و آن‌ها را تجمیع می‌نمایند و این گونه در طی روال پرس‌وجو، یک به روز رسانی موضعی را اعمال می‌نمایند. هنگامی که پرس‌وجو کاملا برطرف شد، از طریق مسیر معکوس یا کوتاهترین مسیر، انتقال به سمت گره مبدا انجام می‌شود. یکی از انگیزه‌های اصلی ارائه این روش مقابله با پرس‌وجوهای پیچیده برای داده‌هایی است که می‌توان از طریق بسیاری از گره‌ها، پاسخی برای آن‌ها ارائه نمود[۳۲,۲۳].

[7] DD: Directed Diffusion
[8] EAR: Energy-aware Routing
[9] Rumor Routing
[10] GBR: Gradient-based Routing
[11] CADR: Constrained Anisotropic Diffusion
[12] CAUGAR
[13] ACQUIRE: Active Query Forwarding

۲.۲ پروتکل‌های سلسله مراتبی[۱۴]

این پروتکل‌ها از تکنیک خوشه‌بندی بهره می‌برند. هدف اصلی مسیریابی سلسله مراتبی، حفظ انرژی گره‌های حسگر به وسیله درگیر کردن آن‌ها در ارتباطات درون خوشه ای و انجام تکنیک تجمیع به منظور کاهش تعداد پیام‌های منتقل شده به گره مبدأ است. دسته‌بندی روش‌های گذشته سلسله مراتبی در شکل ۳ قابل مشاهده بوده و توضیحات هر یک از آن‌ها در ادامه آمده است[۲۳].

شکل ۳ . پروتکل‌های سلسله‌مراتبی مسیریابی در گذشته

در روش مسیریابی وفقی کم‌مصرف بر مبنای خوشه‌بندی[۱۵]، یک عملیات خوشه‌بندی بر روی گره‌ها انجام می‌شود. همچنین برای هر خوشه، یک سرخوشه انتخاب می‌گردد و تنها این گره‌ها قابلیت ارتباط با گره مبدا درخواست را دارا می‌باشند. انجام تکنیک تجمیع در این روش به صورت درون خوشه‌ای انجام می‌شود. همچنین در این روش، سرخوشه‌ها با گذشت زمان، به طور تصادفی تغییر می‌کنند تا در اتلاف انرژی گره‌ها، توازن برقرار شود. خوشه‌بندی پویا در این روش موجب افزایش طول عمر سیستم می‌گردد. همچنین این روش کاملا توزیع شده است و نیازی به دانش کلی درباره شبکه ندارد. با این حال از مسیریابی تک‌گامی استفاده می‌کندکه در آن هر گره می‌تواند مستقیما به سرخوشه و گره مبدا، اطلاعات را ارسال نماید. این روش برای شبکه‌های مستقر در مناطق بزرگ کاربردی ندارد. علاوه بر این، ایده خوشه بندی پویا سربار اضافی را به ارمغان می‌آورد[۳۳,۲۳].

روش تجمیع داده با بازدهی بالای انرژی در سیستم‌های اطلاعاتی مبتنی بر حسگر[۱۶] به جای تشکیل خوشه‌های متعدد، زنجیره‌هایی از گره‌های حسگر تشکیل می‌دهد به طوری که ارسال و دریافت داده در هر گره تنها با همسایه آن صورت گرفته و فقط یک گره از آن زنجیره برای انتقال اطلاعات به ایستگاه پایه انتخاب می‌شود. داده‌ها به صورت گره به گره جمع‌آوری و تجمیع می‌گردند و در نهایت به ایستگاه پایه ارسال می‌شود. این روش به جای استفاده از مسیریابی تک گامی از مسیریابی چند گامی با تشکیل زنجیره‌ها و انتخاب تنها یک گره برای انتقال داده به ایستگاه پایه به جای استفاده از گره‌های متعدد، استفاده می‌کند[۳۴,۲۳].

روش مسیریابی حساس به آستانه[۱۷]، دارای مکانیزم سلسله‌مراتبی و داده محور است که برای پاسخگو بودن به تغییرات ناگهانی در ویژگی‌های حساس مانند دما ساخته شده است. در این روند با در نظر گرفتن دو آستانه نرم و سخت، تنها در صورتی داده را توسط یک گره به ایستگاه پایه منتقل می‌نماید که تغییرات ناگهانی به وقوع بپیوندد. همچنین در این روش نیز از تکنیک خوشه‌بندی استفاده می‌شود. این روند برای سیستم‌هایی که به گزارش دوره‌ای نیازمند هستند، مناسب نیست؛ زیرا در صورتی که شبکه در تمامی لحظات مقادیر جزئی تغییر را حس کند، کاربر به هیچ وجه داده‌ای را دریافت نخواهد نمود[۳۵,۲۳].

در روش مسیریابی آگاه از انرژی در شبکه‌های مبتنی بر خوشه-بندی، حسگرها قبل از بهره برداری از شبکه، در خوشه‌ها گروه بندی می‌شوند. همچنین در این روش، ایستگاه پایه تنها با دروازه‌ها ارتباط برقرار می‌نماید. در این روش، گره‌های حسگر در یک خوشه، می‌توانند در یکی از چهار حالت سنجش به تنهایی، وظیفه بازپخش به تنهایی، فعال بودن هر دو حالت و غیرفعال بودن قرار بگیرند. در حالت سنجش، گره توسط حسگر خود در حال سنجش محیط و به دست آوردن داده است. در حالت بازپخش، مدار ارتباطی آن برای انتقال اطلاعات از گره‌های فعال روشن است اما به خودی خود محیط را نمی‌سنجد. در حالت غیر فعال مدار سنجش و ارتباط گره خاموش است. یکی از رویکردهای موجود در این روش، امکان محدود نمودن محدوده انتقال داده با هدف کاهش تاخیر می‌باشد[۳۶,۲۳].

[14] Hierarchical Protocols
[15] LEACH: Low-energy Adaptive Clustering Hierarchy
[16] PEGASIS: Power-efficient Gathering in Sensor Information Systems
[17] TEEN: Threshold Sensitive Energy-efficient Sensor Network Protocol

روش مسیریابی خودسازمانده[18]، از حسگرهای ناهمگن پشتیبانی می‌کند. ویژگی این حسگرها آن است که هم می‌توانند ثابت باشند و هم متحرک. برخی از این حسگرها محیط را مورد بررسی قرار داده و داده‌ها را به سمت گره‌های تعیین شده که به عنوان دروازه عمل می‌کنند، هدایت می‌کنند. گره‌های دروازه ثابت بوده و ستون فقرات شبکه را تشکیل می‌دهند. در نهایت داده‌های تجمیع شده از طریق دروازه‌ها به ایستگاه‌های پایه منتقل می‌گردند. در این روش هر یک از گره‌هایی که وظیفه سنجش را بر عهده دارند از طریق گره دروازه به ایستگاه پایه مشخصی متصل هستند و گره‌های سنجش متصل به یک ایستگاه پایه از طریق آدرس ایستگاه پایه در آن‌ها قابل شناسایی است. به منظور افزایش تحمل‌پذیری اشکال، از الگوریتم حلقه‌های محلی مارکوف که یک قدم زدن تصادفی را بر روی درختان پوشا انجام می‌دهد، استفاده می‌شود. این روش شامل چهار مرحله کشف، سازماندهی، نگهداری و در نهایت، خودسازماندهی مجدد است[۳۷,۲۳].

۲.۳ پروتکل‌های مبتنی بر مکان[19]

در اغلب اوقات، اطلاعات مکان برای محاسبه فاصله بین دو گره خاص به منظور برآورد انرژی مصرفی مورد نیاز است. از آنجایی که روال‌های آدرس دهی همچون IP برای شبکه‌های حسگر وجود ندارد، می‌توان از اطلاعات مکانی، برای مسیریابی در آن‌ها استفاده نمود. با این روش می‌توان اطلاعات را به منطقه مکانی مشخصی ارسال نموده و در نتیجه، تعداد انتقال اطلاعات را به طور قابل توجهی در جهت کاهش انرژی مصرفی، کاهش داد. این روش از گره‌های متحرک نیز پشتیبانی می‌نماید. دسته‌بندی روش‌های گذشته مبتنی بر مکان در شکل ۴ قابل مشاهده بوده و توضیحات هر یک از آن‌ها در ادامه آمده است[۲۳].

شکل ۴ . پروتکل‌های مبتنی بر مکان در گذشته

روش شبکه ارتباطی با حداقل انرژی[20]، از GPSهای کم مصرف جهت مکان‌یابی گره‌ها در شبکه استفاده می‌نماید. همچنین این شبکه برای حسگرهای متحرک و ثابت قابل استفاده است. ناحیه رله در این شبکه‌ها در واقع ناحیه‌ای است که انتقال از طریق گره‌های موجود در آن از انتقال مستقیم، مصرف انرژی کمتری دارد. این پروتکل دارای قابلیت تنظیم مجدد است و بنابر این می‌تواند به صورت پویا با خرابی گره‌ها یا استقرار حسگرهای جدید، سازگار شود[۳۸,۲۳].

پروتکل وفاداری تطبیقی جغرافیایی[21]، یک الگوریتم مبتنی بر مکان است که با خاموش کردن گره‌های غیرضروری در شبکه، بدون این که روی سطح وفاداری مسیریابی تاثیر بگذارد، موجب صرفه‌جویی در مصرف انرژی می‌گردد. این روش یک شبکه مجازی برای منطقه تحت پوشش تشکیل می‌دهد. هر گره بر اساس حاصل سنجش مکانی خود با استفاده از GPS سعی می‌کند خود را با نقطه‌ای از شبکه مجازی مرتبط کند. در گره‌های متحرک، تحرک توسط گره به همسایگان آن اطلاع‌رسانی می‌گردد. قبل از پایان زمان و یا عمر یک گره فعال، گره‌های خواب، بیدار می‌شوند و یکی از آن ها فعال می‌گردد. این الگوریتم در تلاش است تا همواره در هر منطقه یک گره را در حالت فعال نگه دارد تا اتصال کلی شبکه از بین نرود. این روش علاوه بر امکان کار با دیدگاه مبتنی بر مکان می‌تواند به صورت سلسله مراتبی عمل نماید که در این حالت برعکس روش‌های قبل، گره‌های دروازه هیچ تجمیع یا تلفیقی را انجام نمی‌دهند[۳۹,۲۳].

هدف از ارائه روش مسیریابی جغرافیایی آگاه از انرژی[22]، در نظر گرفتن یک منطقه خاص، به جای ارسال علایق به کل شبکه است. در روش حاضر، هر گره یک تخمین هزینه و یک تخمین هزینه یادگیری برای رسیدن به مقصد از طریق همسایگان خود را نگه می‌دارد. هزینه تخمینی، ترکیبی از انرژی باقیمانده و فاصله تا مقصد است. هزینه آموخته شده، پالایش هزینه تخمین زده شده برای مسیریابی در اطراف حفره‌های شبکه است. حفره در هنگامی به وجود می‌آید که یک گره هیچ همسایه‌ای به طوری که از خود گره به مقصد نزدیک تر باشد، ندارد. این روش، شامل دو مرحله ارسال بسته‌ها به سمت منطقه هدف و انتقال بستهها

[18] Self-organizing Protocol
[19] Location-based Protocols
[20] MECN: Minimum Energy Communication Network
[21] GAF: Geographic Adaptive Fidelity
[22] GEAR: Geographic and Energy-aware Routing Protocol

در داخل منطقه است. این روش باعث کاهش مصرف انرژی برای راه اندازی مسیر می‌شود[۴۰,۲۳].

۲.۴ پروتکل‌های آگاه بر جریان شبکه و کیفیت سرویس[۲۳]

برخی از پروتکل‌های مسیریابی ارائه شده آگاه از جریان شبکه و کیفیت سرویس هستند. این پروتکل‌ها ضمن تنظیم مسیرها در شبکه حسگر، الزامات تاخیر در روال ارسال انتها به انتها را در نظر می‌گیرند. دسته‌بندی روش‌های گذشته آگاه بر جریان شبکه و کیفیت سرویس در شکل ۵ قابل مشاهده است[۲۳].

شکل ۵ . پروتکل های آگاه بر جریان شبکه و کیفیت سرویس در گذشته

روش مسیریابی آگاه از انرژی با طول عمر بیشینه[۲۴]، یک رویکرد مبتنی بر جریان شبکه است که هدف اصلی آن به حداکثر رساندن طول عمر شبکه به وسیله تعیین هزینه پیوند به عنوان تابعی از انرژی باقی‌مانده گره و انرژی مورد نیاز جهت انتقال با استفاده از آن لینک است. این روش سعی می‌کند با استفاده از یافتن توزیع ترافیک در شبکه، راه حل خود را ارائه نماید. این روش با استفاده از روش کوتاهترین مسیر بلمن-فورد[۲۵]، بهترین مسیر به طوری که انرژی باقی‌مانده، بیشینه شود را می‌یابد[۴۱,۲۳].

پروتکل مسیریابی تخصیص متوالی[۲۶]، اولین رویکرد مبتنی بر مفهوم کیفیت سرویس[۲۷] است. در این پروتکل با در نظر گرفتن معیار سنجش کیفیت سرویس، منبع انرژی در هر مسیر و سطح اولویت هر بسته، ریشه درختانی را در همسایگی یک گامی مقصد ایجاد می‌کند. یکی از این مسیرها با توجه به منابع انرژی و کیفیت سرویس در مسیر انتخاب می‌شود. هر گونه نارسایی محلی توسط یک روش خودکار، ترمیم می‌شود. این روش مسیرهای مختلفی را نگه می‌دارد و بنابراین از باز نگه داشتن جداول و حالت‌ها در هر گره حسگر، رنج می‌برد. به خصوص هنگامی که تعداد گره‌ها بسیار زیاد است[۲۳] [۴۳-۴۲].

پروتکل مسیریابی مبتنی بر کیفیت سرویس و آگاه از انرژی[۲۸]، مسیریابی و یافتن مسیر کم‌هزینه و کارآمد از نظر انرژی مصرفی را با در نظر گرفتن هزینه پیوند بر طبق انرژی گره‌ها، انرژی مورد نیاز جهت انتقال، میزان خطا و سایر پارامترهای ارتباطی، انجام می‌دهد. این روش برای یافتن کوتاه ترین مسیرها از روش دایجکسترا[۲۹] استفاده می‌کند[۴۴,۲۳].

پروتکل مسیریابی بر مبنای کیفیت سرویس با تضمین ارتباط انتها به انتهای بی‌درنگ، از مسیریابی جغرافیایی بهره می‌گیرد. این روش سعی دارد تا برای هر بسته سرعت رسیدن آن به مقصد را تخمین بزند. همچنین این روش میتواند از ازدحام شبکه جلوگیری نماید. این روش در تخمین تاخیرها از تکنیک انتقال غیرقطعی جغرافیایی بدون تابعیت[۳۰] بهره می‌گیرد. همچنین، مسیر جدیدی را در هنگامی که یک گره نتواند گام بعدی را پیدا کند، با استفاده از ارسال پیام به گره‌های منبع می‌یابد که موجب جلوگیری از ازدحام در شبکه خواهد شد[۴۵,۲۳].

همانطور که مشاهده نمودید این روش‌ها در طی اجرا سعی بر بهینه نمودن تصمیمات خود داشتند، اما به طور واقعی بررسی نمی‌نمودند که آیا واقعا تصمیماتی که اتخاذ می‌نمایند، بهینه است یا خیر. این در حالی است که روش‌های مبتنی بر یادگیری-ماشین سعی بر یادگیری تابع توزیع رویدادهای محیط دارند، تا بتوانند بر اساس وقایع گذشته، در اتفاقات پیش رو، تصمیمات بهینه‌ای را اتخاذ نمایند.

در شکل ۶ بهره مندی و یا عدم بهره مندی روش های مذکور از ویژگی های کلیدی پروتکل های مسیریابی مورد مقایسه قرار گرفته است[۲۳].

[23] Network flow and QoS Aware Protocols
[24] Maximum Lifetime Energy-aware Routing Protocol
[25] Bellman-Ford
[26] SAR: Sequential Assignment Routing Protocol
[27] QoS: Quality of Service
[28] Energy-aware QoS-based Routing Protocol
[29] Dijkstra's Algorithm
[30] SNFG: Stateless Geographic Non-deterministic forwarding

Routing protocol	Data-centric	Hierarchical	Location-based	QoS	Network-flow	Data aggregation
SPIN	✓					✓
Directed Diffusion	✓					✓
Rumor routing	✓					✓
Shah and Rabaey	✓		✓			
GBR	✓					✓
CADR	✓					
COUGAR	✓					✓
ACQUIRE	✓					
Fe et al.					✓	
LEACH		✓				✓
TEEN and APTEEN	✓	✓				✓
PEGASIS		✓				✓
Younis et al.		✓	✓			
Subramanian and Katz		✓				✓
MECN and SMECN			✓			
GAF		✓	✓			
GEAR			✓			
Chang and Tassiulas		✓			✓	
Kalpakis et al.			✓		✓	
Akkaya et al.		✓		✓		
SAR				✓		
SPEED			✓	✓		

شکل ۶- مقایسه روش های کلاسیک از منظر ویژگی های کلیدی[۲۳]

$$t = min_{y \in neighbors\ of\ x} Q_x(d.y) \qquad (۵)$$

اگر در هر q واحد از زمان، بسته‌ای به صف گره x وارد شود و در هر s واحد از زمان از گره x به گره y ارسال شود، اصلاح مقدار تخمینی توسط رابطه (۶) محاسبه خواهد شد[۴۶].

$$\Delta Q_x(d.y) = \eta \left(\overbrace{q + s + t}^{new\ estimate} - \overbrace{Q_x(d.y)}^{old\ estimate} \right). \qquad (۶)$$

η ضریب یادگیری است و بر طبق تجربه نویسنده مقاله، توصیه شده است مقدار آن برابر ۰.۵ در نظر گرفته شود. تفاوت این روش با الگوریتم یافتن کوتاه‌ترین مسیر بلمن‌فورد در دو مورد زیر خلاصه می‌گردد[۴۶].

- مراحل تخفیف مسیر[۳۱] را به طور یکنواخت و آنلاین انجام می‌دهد. شبه‌کد این مراحل در شکل ۶ آمده است.

```
For the edge from the vertex u to the vertex v:
    if d[u] + w(u,v) < d[v] is satisfied:
        update d[v] to d[u] + w(u,v)
```

شکل ۷ . شبه‌کد روال Path Relaxation [۴۲]

- طول مسیر را صرفا بر اساس تعداد گام‌ها محاسبه ننموده بلکه بر اساس مدت زمان کل انتقال بسته، اندازه‌گیری می‌نماید.

از سمت دیگر این روش بیان می‌کند که محاسبه تمامی مقادیر $Q_x(d.y)$ دارای هزینه محاسباتی بالایی است و به حافظه زیادی جهت ذخیره سازی این مقادیر نیازمند است. این روش با اعلام این نکته که برای رفع چنین مشکلی می‌توان یک تابع تخمین بر اساس تکنیک شبکه‌های عصبی طراحی نمود، الگوریتم Q-Routing را روشی شامل بازنمایی تابع $Q_x(d.y)$

۳ بهره‌گیری از یادگیری تقویتی در شبکه حسگر بی‌سیم

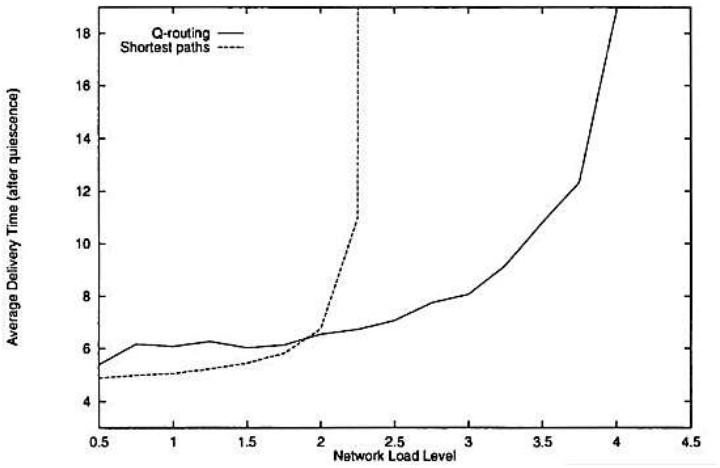
در این بخش به بررسی پروتکل‌های کنترل و مسیریابی‌ای می‌پردازیم که با بهره‌گیری از یادگیری تقویتی سعی بر افزایش بازدهی انرژی در شبکه‌های حسگر بی‌سیم دارند.

۳.۱ روش Q-Routing

این روش[۴۶] با به دست آوردن تخمینی از مدت زمان ارسال بسته، سعی بر کمینه کردن این مقدار در حالت کلی دارد. در واقع این روش سعی بر پاسخ دادن به این سوال را دارد که یک بسته در مرحله پیش رو به سمت کدام یک از همسایه‌های خود ارسال شود تا زودتر به مقصد برسد. بررسی کارایی این روش در نهایت با توجه به مجموع زمان ارسال بسته‌ها، سنجیده خواهد شد. تخمین مدت زمان ارسال بسته در روش حاضر با نماد $Q_x(d, y)$ نشان داده می‌شود که تخمینی از مدت زمان ارسال بسته P از گره x به سمت گره d از مسیر یکی از همسایگان گره x همچون گره y است که شامل کل مدت زمانی که بسته در صف گره x حضور دارد و زمانی که به سمت گره y ارسال می‌شود، خواهد شد. پس از محاسبه این تخمین‌ها کافی است هربار همسایه‌ای جهت ارسال بسته به آن انتخاب شود که این مقدار تخمینی به ازای آن در حالت کمینه قرار داشته باشد. این روال در رابطه (۵) قابل مشاهده است[۴۶]:

[31] Path Relaxation

توسط انجام فرآیند تخمین به وسیله شبکههای عصبی دانسته و بیان می‌دارد که روش حاضر از این شبکه برای تصمیم‌گیریهای خود بهره می‌گیرد و نه از جدول $Q_x(d.y)$ که دارای حجم بسیار بالایی است. در قسمت نتیجه‌گیری و ارزیابی الگوریتم، این روش تنها با الگوریتم کوتاه‌ترین مسیر بلمن‌فورد مقایسه می‌گردد که نتیجه‌گیری آن بیان می‌دارد که با افزایش بار شبکه، کارایی روش حاضر نسبت به الگوریتم بلمن‌فورد افزایش می‌یابد. همچنین این مقاله بیان می‌کند که الگوریتم ارائه شده می‌تواند در آن دسته از شبکه‌های حسگر بی‌سیم که دارای تغییرات دینامیکی در ساختار، الگوی ترافیکی و سطح بار شبکه هستند نیز به خوبی کار کند. مقایسه میانگین زمان ارسال در روش‌های Shortest path و Q-Routing با افزایش میزان بار اعمالی به شبکه در شکل ۷ آمده است.



شکل ۸ . مقایسه میانگین زمان ارسال در روش های Shortest path و Q-Routing با افزایش میزان بار اعمالی به شبکه[۴۲]

۳.۲ روش مسیریابی وفقی[۳۲]

این روش[۴۷] با هدف یافتن یک سیاست مسیریابی بهینه با در نظر گرفتن امکان بهینه‌سازی اهداف چندگانه و بر پایه تکنیک یادگیری تقویتی، بنا نهاده شده است. مواردی را که مقاله حاضر از موارد پر اهمیت، جهت ایجاد یک شبکه حسگر بیسیم با کارایی و طول عمر بالا میداند شامل موارد زیر است:

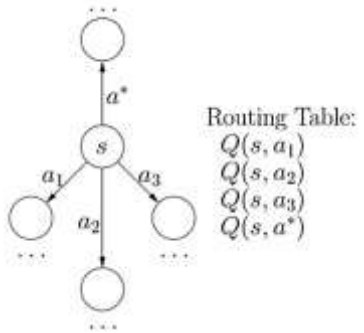
- مسیریابی صحیح انتقال بسته
- برقراری تعادل در میزان بار اعمالی به شبکه
- پایداری لینک ها
- تجمیع داده‌های دارای وابستگی و سپس ارسال آنها

این روش بر پایه ترکیبی از روش تکرار سیاست با حداقل مربعات[۳۳] و روش Q-Learning شکل گرفته است. این روش، یکی از روش‌های بر پایه یادگیری تقویتی است که از مدل مسئله مستقل است. همچنین یکی از ویژگی‌های آن، امکان یافتن سیاست بهینه با تعداد کمی تلاش[۳۴] (جستجو) است.

روش تکرار سیاست با حداقل مربعات، روشی بر مبنای یادگیری تقویتی است که مقدار Q را بر اساس سیاستی خاص همچون π با استفاده از یک تخمین زننده پارامتری توابع، تخمین می‌زند. در حقیقت این روش با توجه به رابطه (۷)، تابع ارزش را در روش یادگیری تقویتی بر اساس ترکیب خطی وزن‌دار تعدادی تابع پایه، تخمین می‌زند.

$$\hat{Q}^{\pi}(s,a,\omega) = \sum_{i=1}^{k} \emptyset_i(s,a)\omega_i = \emptyset(s,a)^T\omega \qquad (۷)$$

این روش، هر گره در یک شبکه حسگر را معادل با یک وضعیت و همسایه‌های هر گره را نیز معادل با وضعیت‌هایی دیگر در نظر گرفته و ارسال هر بسته از هر گره به هر یک از همسایگان آن را یک عملگر در نظر می‌گیرد. در نهایت مقادیر Q-Value محاسبه شده توسط این روش، یک سیاست مسیریابی بهینه در آن شبکه حسگر بی‌سیم خواهد بود. طریقه مسیریابی در این روش به این شکل است که وقتی یک گره، قصد ارسال بسته به یکی از همسایگان خود را دارد، مقدار Q-Value همسایگانش را با یکدیگر مقایسه نموده و هر یک که مقدار بیشینه داشته باشد به عنوان گره مقصد ارسال بسته در نظر گرفته می‌شود. مجددا همین روال در گره مقصد ارسال، انجام شده و آن قدر ادامه می‌یابد تا مقصد انتخابی، برابر مقصد نهایی باشد. این رویکرد در شکل ۸ آمده است.



شکل ۹ . مسیریابی با استفاده از مؤلفه Q-Value [۴۳]

الگوریتم حاضر با استفاده از روال LSPI و در نظر گرفتن پارامتر هایی شامل اختلاف فاصلههای s و s' از ایستگاه پایه که با $d(s.a)$، انرژی باقی مانده برای s' که با $e(s.a)$، تعداد مسیرهای ممکن رسیدن از s به s' که با $c(s.a)$ و قابلیت اطمینان لینک اتصال بین s و s' که با $l(s.a)$ نمایش داده میشود و دخالت این پارامتر ها در محاسبه $\phi(s.a)$ در رابطه (۸) قابل مشاهده است.

$$\phi(s.a) = \{d(s.a).e(s.a).c(s.a).l(s.a)\}. \qquad (۸)$$

[32] AdaR: Adaptive Routing Protocol
[33] LSPI: Least Square Policy Iteration
[34] Attempt

به طوری که تمامی این مقادیر در بازه منهای یک و یک نرمالیزه[35] شده باشند. تابع پاداش در این الگوریتم به صورت زیر است:

- اگر به ایستگاه پایه دست پیدا کنیم در این مورد پاداشی مثبت برابر با حداکثر پاداش ممکن در نظر میگیریم.
- اگر طول مسیر از آستانه مشخصی بگذرد در این مورد پاداش صفر در نظر خواهیم گرفت.
- در حالتی که در مسیر ایجاد شده حلقهای مشاهده شود(یک گره دو بار یا بیشتر ملاقات شده باشد) پاداش به صورت $R = -R_{max}/2$ در نظر گرفته می‌شود.

الگوریتم AdaR از لحاظ میزان پاداش دریافتی و همچنین نرخ موفقیت، عملکرد بهتری نسبت به الگوریتم Q-Learning دارد.

۳.۳ روش ATP[36]

در این روش[۴۸]، پارامتر Q-Value عکس مسائل دیگر، میزان هزینه به حساب آمده و در هر مرحله، گره‌ای انتخاب می‌شود که مقدار این پارامتر به ازای آن گره از بقیه گره‌ها کمتر باشد. هر گره مقدار پارامتر Q-Value همسایگان خود را نیز در جهت استفاده، در مرحله ارسال بسته به همسایه مناسب، در خود ذخیره نموده و نام NQ-Value را به آن اختصاص می‌دهد. پارامترهای مذکور، هر بار بر اساس ایده اصلی نظریه یادگیری، به روز رسانی می‌شوند. این روال به روزرسانی در روابط (۹) و (۱۰) آمده است:

$$NQ_m(n) \leftarrow rNQ_m(n) + (1-r)Q(n). \quad (۹)$$

$$Q_m \leftarrow (1-\alpha)Q_m + \alpha\left(o_m + \min_n NQ_m(n)\right). \quad (۱۰)$$

در روابط بالا m نوع بسته و n گره مبدا بسته است که دارای مؤلفه هزینه $Q(n)$ می‌باشد. همچنین پارامتر r نرخ به‌روزرسانی و α نرخ یادگیری است. همچنین o_m مقدار فعلی تابع هدف و n گره همسایه مورد بحث است. یکی از محدودیت های این روش لزوم ثابت بودن گره چاهک می‌باشد. روش حاضر از سه مرحله به شرح زیر تشکیل شده است که فرآیند یادگیری به طور همزمان در این سه مرحله اتفاق می‌افتد.

- **فاز پیکربندی اولیه**[37]: در این مرحله، یک درخت پوشا که بیانگر روال ارتباط و مسیر بین گره‌ها می‌باشد، تشکیل شده و در طی الگوریتم این درخت پوشا به حالت بهینه خود می‌رسد. حالت بهینه حالتی است که با شروع از گره مبدا انتخابی در درخت پوشا و با حرکت به سمت گره هدف از طریق مسیر موجود، با کمترین هزینه ممکن، داده را به مقصد برساند. ریشه درخت پوشا همان گره هدف است و روال ارتباطی گره‌ها در درخت پوشا به این صورت است که هر گره دارای اشاره‌گری به والد خود می‌باشد و والد نهایی در این درخت پوشا، گره هدف خواهد بود و در ابتدا آن گره‌ای که به ازای گره هدف در همسایگی آن قرار داشته باشد و دارای NQ-Value[38] کمینه باشد به عنوان فرزند آن گره در درخت در نظر گرفته می‌شود.
- **فاز ارسال**[39]: در این مرحله بسته‌های دریافتی هر گره با توجه به مسیر تشکیل شده در درخت پوشا، به سمت گره هدف حرکت می‌کنند. بدیهی است در لحظه رسیدن یک داده به هر گره، در صورت ایجاد تغییر در مقدار لحظه‌ای NQ-Value، ممکن است گره انتخاب شده به عنوان گره والد با NQ-Value کمینه تغییر کند و بر همین اساس انتخاب گره بعدی در لحظه فعلی نسبت به گره‌ای که در مرحله قبل به ازای این مرحله پیشنهاد می‌شد تغییر نماید. در نتیجه مسیر بهینه هر بار بر اساس مقادیر NQ-Value به دست آمده در آن لحظه به روز رسانی می‌گردد و این خود موجب بهبود کارایی الگوریتم و وفقی بودن آن نسبت به تغییرات موجود در شبکه از جمله تغییرات مکانی گره‌ها می‌گردد. همچنین گره هدف، همواره مقدار صفر را به ازای Q-Value خود، پخش همگانی می‌نماید تا این گونه به عنوان آخرین گره والد انتخاب گردد.
- **فاز تایید**[40]: در طی این فاز در صورتی که برای دوره زمانی مشخصی داده از گرهای همچون m به گرهای همچون v ارسال نشود، مقدار پارامتر NQ-Value به وسیله افزایش آن به اندازه یک ثابت مشخص، به‌روز رسانی شده و انتخاب گره با NQ-Value کمینه مجددا انجام می‌گردد.

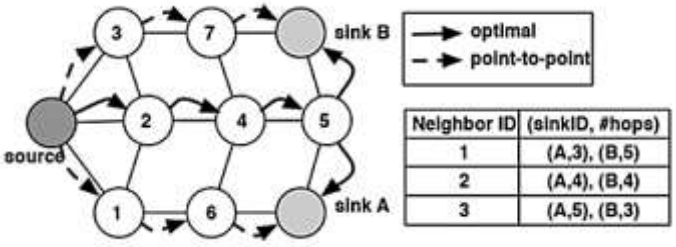
در نهایت این الگوریتم برای انتخاب مقدار مناسب برای پارامتر نرخ به روز رسانی، بیان می‌کند که مقدار این پارامتر به صورت تعداد بسته‌هایی که طی ارسال، موفق نبوده اند نسبت به مجموع کل بسته‌ها در نظر گرفته شود. همچنین روش حاضر بیان می‌کند که اگر نرخ از دست رفتن اتصالات، بالا باشد، پارامترهای مربوط به همسایگان یک گره به خوبی تغییر نمی‌کنند و اگر این نرخ به صفر نزدیک باشد، روند به روزرسانی این پارامترها به

[35] Normalize
[36] Adaptive Tree Protocol
[37] Initialization Phase
[38] Neighbor Q-Value
[39] Forwarding Phase
[40] Confirmation phase

سرعت و صحت انجام می‌شود. در فاز تایید، مقدار پارامتر همسایگان یک گره به طور نمایی با توان نسبت ارسالهای ناموفق به موفق، جهت حل این مشکل افزایش می‌یابد.

۳.۴ روش FROMS

یکی از مهم ترین ویژگی های این الگوریتم[۴۹]، توانایی پردازش و یافتن مسیر بهینه به ازای چندین گره هدف است. این الگوریتم در حقیقت با تشکیل یک درخت به اشتراک گذاری مسیر سعی دارد تا به ازای هر گره اعلام کند که چه تعداد قدم تا رسیدن به هدف از طریق گره حاضر باید طی شود. با داشتن چنین داده‌ای در هر لحظه بر اساس گره هدف انتخاب شده و جدول همسایه ها که در شکل ۹ آمده است، می‌توان گره با مقدار کمینه تخمین تعداد قدم‌ها تا هدف را به عنوان گره بعدی مسیر انتخاب نموده و داده را به سمت آن گره ارسال نمود.



Neighbor ID	(sinkID, #hops)
1	(A,3), (B,5)
2	(A,4), (B,4)
3	(A,5), (B,3)

شکل ۱۰ . جدول همسایگی و گراف مسیر در الگوریتم FROMS [۴۵]

فرآیند یادگیری در این روش به دو مرحله تقسیم شده است. مرحله اول، شناسایی و انتخاب عملگرها و مرحله دوم، ارسال بازخورد گونه پاداش است. در مرحله اول ساختاری با نام درخت به اشتراک گذاری مسیر استفاده شده است. مهم ترین اشکال موجود در درخت به اشتراک گذاری مسیر، اندازه آن است که افزایش آن به علت تعداد بالای اقدامات ممکن جهت اجرا، محتمل است. در این روش، تعدادی تابع اکتشاف[۴۱] با هدف کاهش اندازه درخت به اشتراک‌گذاری مسیر، پیشنهاد شده است. این توابع از شیوه هرس کردن جهت کاهش اندازه درخت به اشتراک گذاری مسیر، استفاده می‌کنند. این توابع شامل مواردی همچون محدودسازی تعداد مسیرها از گره هدف به هر یک از گره‌ها در درخت و محدودسازی بیشینه هزینه هر یک از مسیرها تا گره هدف است. همچنین این روش از سیاست اکتشاف تصادفی وزن‌دار[۴۲] به جهت افزایش سرعت همگرایی، استفاده می‌نماید. در مرحله دوم روال پخش همگانی به هر یک از همسایگان اجازه ذخیره سازی اطلاعات افزوده موجود را داده است تا جهت بهبود الگوریتم از این اطلاعات استفاده شود. مقدار پاداش نیز یکی از این مقادیر ذخیره شده است. در طی الگوریتم حاضر هر بار تعداد قدم‌های بهینه تا هدف، با برگشت به سمت گره مبدا و به روز رسانی مقادیر، اصلاح می‌گردد تا در نهایت بتوان به مقادیر بهینه دست یافت. برقراری مصالحه‌ای منطقی بین اکتشاف و بهره‌گیری از اهمیت ویژه‌ای برخوردار است. الگوریتم حاضر پس از طی تعدادی تکرار در اجرای مراحل یادگیری، با رسیدن به نقطه‌ای که در آن تغییرات Q-Value از حد مشخصی کمتر شود، به همگرایی رسیده و متوقف می‌شود. با بررسی‌های انجام شده، این روش، از نظر پایداری در مقادیر نرخ ارسال و انرژی مصرفی، نسبت به روش انتشار مستقیم، دارای برتری است.

این روش با روشهای انتشار مستقیم و حریصانه از منظر سربار شبکه بر روی هر بسته اطلاعاتی، در حالاتی شامل شبکه با تعداد گره هدف متغیر، شبکه با تعداد کل گره‌های متغیر و شبکه با تعداد گره مبدا متغیر، مقایسه شده است و نتیجه این مقایسه گویای برتری روش FROMS نسبت به روش های انتشار مستقیم و حریصانه است.

۳.۵ روش QELAR

در این الگوریتم[۵۰]،که به طور خاص منظوره برای شبکه‌های حسگر بی‌سیم زیر آب، طراحی شده است، فارغ از آن که یک گره به عنوان فرستنده بعدی انتخاب شده باشد و یا نشده باشد اطلاعاتی از جمله انرژی باقیمانده و انرژی متوسط گره را به دست آورده و در لیست همسایگان محلی، این مقادیر را به روز رسانی می‌کند. در این الگوریتم در هنگامی که بسته دریافتی، یک بسته اطلاعاتی باشد و با بررسی شناسه گره بعدی (Forwarder ID)، در صورتی که مشخص شود گره تعیین شده واجد شرایط ارسال بسته نیست، بسته از بین می‌رود. اما در صورتی که فرستنده تعیین شده، واجد شرایط ارسال داده باشد آن گاه بر اساس مقادیر Q-Value فعلی و رابطه (۱۱) مقادیر جدید Q-Value محاسبه شده و گره با بیشینه مقدار Q-Value به عنوان فرستنده بعدی انتخاب می‌گردد.

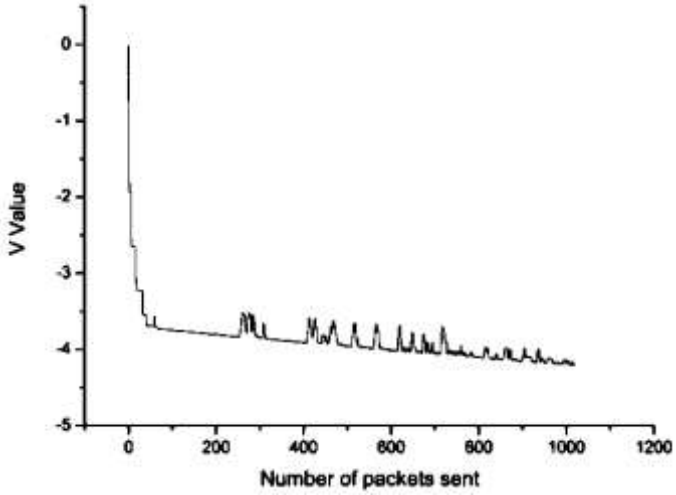
$$Q^*(s_t,a_t) = r_t + \gamma \sum_{s_{t+1}\in S} P^{a_t}_{s_t s_{t+1}} \max_a Q^*(s_{t+1},a). \qquad (۱۱)$$

در این رابطه، $Q^*(s_t,a_t)$ نشان دهنده پاداش حاصل از انجام عمل a_t در وضعیت s_t است. γ پارامتر کاهنده و P ماتریس احتمالات جابجایی بین وضعیت ها است. در نهایت، قبل از ارسال داده به گره با Q-Value بیشینه، مقادیر موجود در Packet Header بر اساس مقادیر به دست آمده به روز رسانی شده و سپس ارسال داده به سمت فرستنده انتخابی، انجام خواهد شد. جهت تشخیص ارسال های ناموفق در الگوریتم حاضر، از روش ارسال بسته تصدیق[۴۳] استفاده می‌شود. در صورتی که با ارسال

[41] Heuristic Function
[42] Stochastic Weighted Explore

[43] ACK: Acknowledge Packet

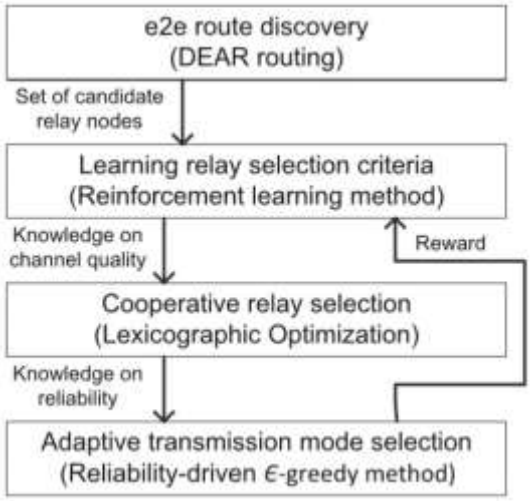
یک بسته و در طی مدت زمان مشخصی بسته در مقصد دریافت نشود به این معنی که پیام تصدیق آن به مبدا نرسد، آنگاه بسته مجددا از مبدا، ارسال می‌گردد. در نهایت در لحظه‌ای که مقصد پیام ارسال شده را دریافت می‌کند آن را به فرستنده بعدی انتخابی ارسال نموده و همچنین پیام تصدیق را برای مبدا، ارسال می‌نماید. همچنین در صورتی که به تعداد بار مشخصی که توسط مقدار ثابت max_trans مشخص می‌شود، بسته دریافت نشود، تلاش برای ارسال، ناموفق تلقی شده و دیگر ارسال مجددی اتفاق نمی‌افتد. این روش با مشکل عدم ارسال موفق بعضی از بسته‌ها به وسیله به روز رسانی تابع ارزش بر اساس سنجش محیط، وفق پیدا می‌کند. همچنین به دست آوردن مقادیر صحیح Q-Value با بهره‌گیری از به روز رسانی مقادیر ماتریس احتمالات حرکت بین وضعیت‌ها، انجام می‌گردد. همگرایی تابع ارزش در شکل ۱۰ بر اساس تعداد بسته‌های ارسالی تا هر لحظه، قابل مشاهده است.



شکل ۱۱ . روال همگرایی به تابع ارزش واقعی بر اساس تعداد بسته‌های ارسالی در روش QELAR [۴۶]

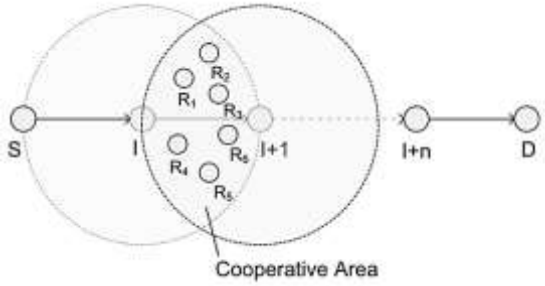
۳.۶ روش DACR[۴۴]

این روش[۵۱]، در حقیقت با نگاه به عنوان یک مسئله بهینه سازی به پروتکل اگاه از کیفیت سرویس و ارائه یک بهینه‌ساز خطی برای آن، شکل گرفته است. این روش جهت انتخاب گره-های بازپخش[۴۵] از الگوریتم DEAR [۴۶] که با بهره‌گیری از تکنیک اجتناب از وارد شدن به ناحیه بحرانی انرژی، در گره‌ها، یک مسیر بهینه به صورت انتها به انتها از مبدا به مقصد را با حداقل انرژی مصرفی می‌یابد، استفاده می‌کند که در آن از بهره‌گیری گره‌های با انرژی پایین جهت افزایش طول عمر شبکه، اجتناب می‌گردد. معماری الگوریتم DACR در شکل ۱۱ قابل مشاهده است.



شکل ۱۲ . معماری روش DACR [۴۷]

الگوریتم DACR به شناسایی مجموعه گره‌های بازپخش در گره مبدا و گره‌های واسط در مسیر کمک می‌کند. این مجموعه از گره‌ها با کمک یکدیگر مسیرهای صحیح انتقال داده را تشکیل می‌دهند. به ناحیه‌ای که این مجموعه از گره‌های بازپخش را تشکیل می‌دهند، منطقه تعاونی[۴۷] گفته می‌شود. این ساختار در شکل ۱۲ قابل مشاهده است.



شکل ۱۳ . مفهوم منطقه تعاونی[۴۷]

در حقیقت الگوریتم DEAR بر پایه الگوریتم مسیریابی بردار فاصله بر اساس تقاضا در شبکه های Ad Hoc[۴۸] و اعمال تغییراتی در آن، ایجاد شده است. مقادیر hopcount و reserved در AODV با مقادیری شامل یک برچسب زمانی و آستانه کمینه انرژی باقی‌مانده جایگزین شده است. همچنین در این الگوریتم، گره مبدا جهت آغاز ارسال داده به مقصد موردنظر یک پیام درخواست مسیر یابی[۴۹] را به صورت پخش‌همگانی ارسال می-نماید. این درخواست هر بار توسط همسایگانی که به عنوان واسط در مسیر ارسال انتخاب شده اند، پخش‌همگانی می‌گردد، تا در نهایت، داده به مقصد برسد. هر گره با دریافت داده در صورتی درخواست RREQ را پخش همگانی می‌کند که انرژی باقی‌مانده آن بیشتر از مقدار تعیین شده به عنوان آستانه انرژی باشد. در

[44] Distributed Adaptive Cooperative Routing Protocol
[45] Relay Nodes
[46] Delay- and Energy-aware Routing Protocol
[47] Cooperative Area
[48] AODV: Ad hoc On-Demand Distance Vector Routing
[49] RREQ: Route Request

طی الگوریتم حاضر هر گره قبل از بازپخش مجدد یک پیام RREQ مقدار دوره زمانی حضور در یک گره[50] را از برچسب زمانی موجود در درخواست RREQ کسر می‌کند. در نهایت همچون الگوریتم AODV بسته‌های RREQ از مسیرهای مختلفی به مقصد می‌رسند. مقصد بر اساس آن که کدام یک از بسته‌های دریافتی دارای مقدار برچسب زمانی بیشینه هستند، تنها به آن بسته در قالب پیام پاسخ به درخواست مسیریابی پاسخ داده و آن را برای گره مبدا ارسال می‌کند. همانطور که مشخص است به هر مقدار که یک مسیر، فاصله کمتری را بین مبدا و مقصد برقرار نماید، مقدار کمتری از برچسب زمانی آن، کسر خواهد شد و در نتیجه مسیر بهتری برای ارسال داده خواهد بود. در ادامه به بررسی روند محاسبه دوره زمانی حضور در یک گره خواهیم پرداخت. رابطه (۱۲)، این روند محاسبه را بیان می‌کند.

$$T_{sojourn} = T_{departure} + T_{transDelay} - T_{arrival} \quad (۱۲)$$

بر طبق آن که $T_{departure}$ زمان ارسال اولین بیت از بسته به پیوند فیزیکی، $T_{transDelay}$ تاخیر ارسال بسته و $T_{arrival}$ مدت زمان مورد نیاز جهت دریافت بسته از کانال و پردازش آن باشد، $T_{sojourn}$ از رابطه بالا به دست خواهد آمد. بر خلاف AODV جهت خطایابی و تشخیص وقوع خطا در طی مسیریابی در DEAR دو شرط، بررسی می‌شود. با برقراری هر یک از این شروط، نتیجه گرفته می‌شود که در طی مسیریابی و ارسال داده ها، با خطا مواجه شده ایم. این شروط شامل موارد زیر است:

- از دست رفتن پیوند ارتباطی بین گره فعلی با گره بعدی در مسیر انتخاب شده
- پایین آمدن انرژی گره جاری تا حد عبور از آستانه انرژی تعیین شده

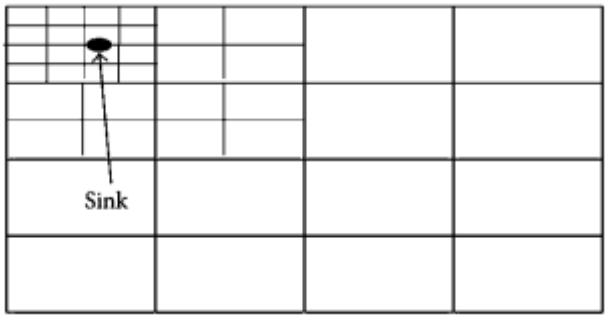
در روال انتخاب گره‌های بازپخش، به عنوان واسط ارسال داده‌ها، گره‌هایی در هر لحظه می‌توانند به عنوان گره بازپخش انتخاب گردند که پیام RREQ را از گره جاری و پیام RREP از گره بعدی در مسیر انتخابی، دریافت نمایند. در این الگوریتم، یک ملاک انتخاب بهینه گره‌های بازپخش در نظر گرفته شده است که در آن از تکنیک یادگیری تقویتی جهت به دست آوردن سیاست بهینه انتخاب گره‌های بازپخش، استفاده شده است. در طی این روال، دو مقدار تاخیر و قابلیت اطمینان در مرحله آموزش، کاملا تاثیر گذار بوده و تابع پاداش بر اساس این مقادیر، پاداش مناسب هر عمل خاص در هر وضعیت خاص را به دست می‌آورد. وضعیت‌ها در اینجا شامل گره جاری، مجموعه گره‌های بازپخش انتخاب شده و گره بعدی می‌باشد. عملگرهای موجود شامل ارسال مستقیم بدون بهره‌گیری از گره‌های بازپخش به عنوان واسط ارسال و ارسال با استفاده از گره‌های بازپخش می‌باشد. سیاست بهینه‌ای که در نهایت به دست خواهد آمد، بیانگر لزوم و یا عدم لزوم استفاده از گره‌های بازپخش در هر وضعیت خاص است. تابع پاداش دارای دو خروجی میزان تاخیر و قابلیت اطمینان است. در حقیقت عملی، بهترین عمل ممکن در وضعیت حاضر خواهد بود که دارای کمینه تاخیر و بیشینه قابلیت اطمینان باشد. این تابع بر اساس مقادیر تاخیر و قابلیت اطمینان در لحظه قبل و لحظه فعلی بر طبق رابطه (۱۳) به دست آمده است. مقدار α سرعت همگرایی به مقادیر تاخیر و قابلیت اطمینان در لحظه فعلی را تعیین می‌کند که مقداردهی صحیح به آن از اهمیت ویژه ای جهت همگرایی به سیاست بهینه، برخوردار است.

$$Rwd = \begin{cases} K_{rel}^{m}(t) = (1-\alpha)K_{rel}^{m}(t-1) + \alpha r_{l,l+1}^{m}(t). \\ K_{del}^{m}(t) = (1-\alpha)K_{del}^{m}(t-1) + \alpha t_{l,l+1}^{m}(t). \end{cases} \quad (۱۳)$$

در نهایت با به دست آمدن سیاست بهینه انتخاب گره‌های بازپخش، با حل یک مسئله بهینه‌سازی به منظور یافتن مسیرهای بهینه، هدف نهایی الگوریتم محقق خواهد شد. همچنین در الگوریتم حاضر یک روال انتخاب وفقی حالت ارسال در نظر گرفته شده است. در مواقعی استفاده از گره‌های بازپخش در مسیر ارسال به علت وجود یک لینک مناسب با قابلیت اطمینان بالا و تاخیر پایین، لزومی ندارد. این شیوه در طی اجرا هر بار حالت ارسال را به صورت وفقی، انتخاب می‌نماید.

۳.۷ روش FTIEE

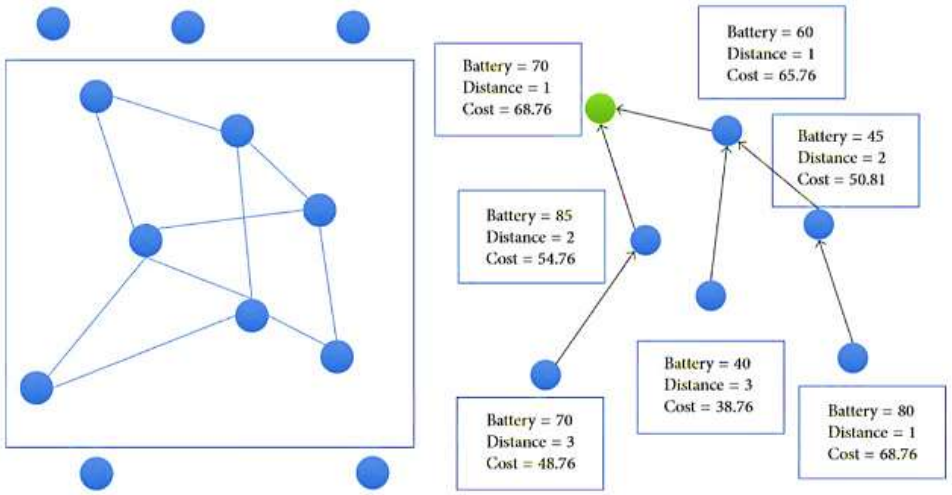
این روش[۵۲]، بر اساس ایده خوشه‌بندی که در روش‌های ساختاری مسیریابی با آن آشنا شده اید، شکل گرفته است. فضای پوشش‌دهی خوشه‌ها در این الگوریتم مربعی است. تعداد خوشه-ها در این روش مقداری ثابت در نظر گرفته شده و اندازه‌ای مربع شکل دارد. اندازه (ظرفیت) هر خوشه بر اساس فاصله خوشه جاری از خوشه‌ای که گره چاهک در آن قرار دارد، تعیین می-شود(بر این اساس که خوشه جاری در همسایگی چندم خوشه شامل گره چاهک قرار گرفته است). در شکل ۱۳ روال تعیین اندازه خوشه‌ها قابل مشاهده است.



شکل ۱۴ . روال تعیین اندازه خوشه‌ها در روش FTIEE [۴۸]

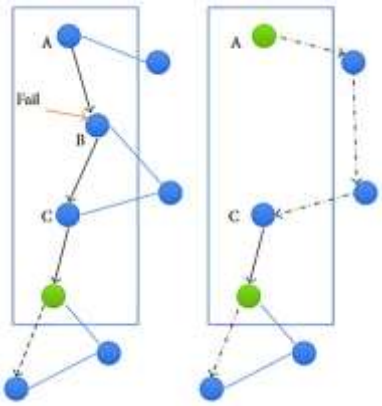
[50] Sojourn Period

این روش جهت انتخاب گره‌های هر خوشه و سرخوشه‌ها، از تکنیک Q-learning در یادگیری تقویتی، استفاده نموده است. به روز رسانی مقدار Q-value در این روش بر اساس دو مؤلفه فاصله گره تا گره چاهک و انرژی باقی مانده گره، انجام شده است. در شکل ۱۴، یک نمونه از روال خوشه‌بندی مربعی و انتخاب سرخوشه و در نهایت شیوه و جهت جریان داده‌ها در یک خوشه قابل مشاهده است.



شکل ۱۵ . روال (الف) خوشه‌بندی مربعی و (ب) انتخاب سرخوشه و تعیین جهت جریان داده‌ها [۴۸]

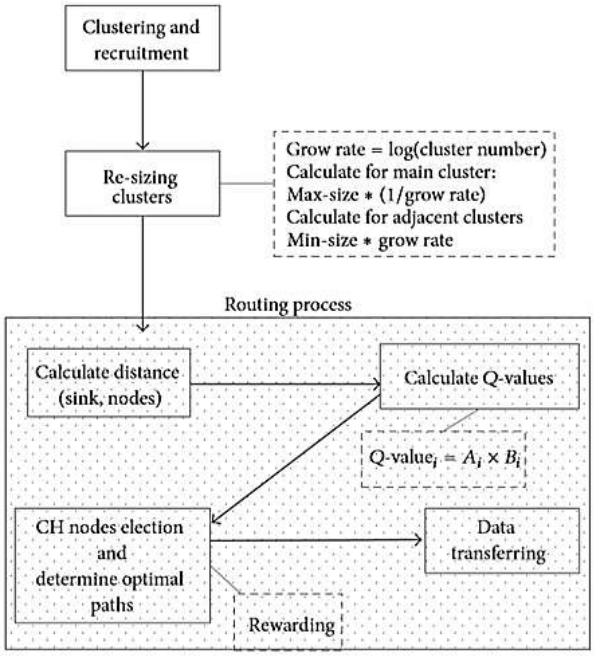
ویژگی دیگر این روش طراحی با در نظر گرفتن یک میزان تحمل‌پذیری اشکال مطلوب است که در شرایطی همچون از دست رفتن یک پیوند در مسیر، کارایی الگوریتم را حفظ خواهد کرد. نمونه‌ای از این تصمیم گیری در لحظات بحرانی، در شکل ۱۵ قابل مشاهده است.



شکل ۱۶ . تحمل‌پذیری اشکال در الگوریتم FTIEE [۴۸]

یک نگاه کلی به الگوریتم، همانطور که در شکل ۱۶ نشان داده شده است به ترتیب در بر گیرنده مواردی شامل خوشه بندی و اصلاح اندازه خوشه‌ها، محاسبه Q-Value ها، انتخاب گره‌های هر خوشه، انتخاب سرخوشه‌ها، محاسبه مسیرهای بهینه، انتقال داده بر اساس مسیر بهینه به دست آمده و اعمال پاداش به دست آمده جهت ادامه یافتن فرآیند همگرایی به سمت مسیرهای بهینه در طی عملیات یادگیری است.

در نهایت روش حاضر با پروتکل LEACH از منظر زمان اتمام انرژی گره‌ها در سه حالت شامل زمان سپری شده تا اولین اتمام انرژی یک گره، زمان اتمام انرژی نیمی از گره‌های شبکه که تا پایان کار، انرژی خود را از دست داده اند و حالت سوم، زمان سپری شده تا اتمام انرژی آخرین گره‌ای که تا انتهای کار شبکه، انرژی آن به اتمام نرسیده است، دارای برتری است.



شکل ۱۷ . نمودار روند اجرایی الگوریتم روش FTIEE [۴۸]

۳.۸ روش MRL-SCSO

این الگوریتم[۵۳]، برای هر گره چندین حالت را در نظر گرفته و بیان می‌کند که هر گره توانایی حضور در یک حالت خاص را در هر لحظه دارد. این حالات شامل حالت کشف، فعال، غیرفعال

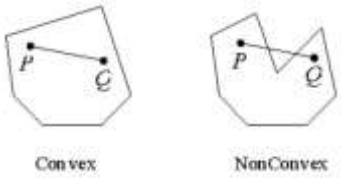
و به خواب رفته است. همچنین روش حاضر با بیان اهمیت برقراری مصالحه‌ای مناسب بین کاوش و بهره‌گیری، روش ε_{greedy} را برای این منظور مناسب دانسته و از این روش جهت برقراری این مصالحه استفاده کرده است. در این روش به ازای گره‌ای با شناسه i مقدار احتمال ε با استفاده از رابطه (۱۴) به دست آمده است.

$$\varepsilon_i = \frac{N_{active}}{N} \qquad (۱۴)$$

در رابطه بالا N_{active} بیانگر تعداد گره‌های دارای حالت active در همسایگی گره i و N تعداد کل گره‌های موجود در همسایگی گره i است. در نتیجه با احتمال $1-\varepsilon$ گره جاری، بسته را به سمت همسایه‌ای که ارسال داده به آن موجب اکتساب بیشینه پاداش شود، ارسال می‌کند و با احتمال ε بسته موجود در گره فعلی، به طور تصادفی برای یکی از همسایگان آن ارسال می‌شود. مقدار انرژی باقی‌مانده هر گره در هر لحظه از رابطه (۱۵) محاسبه می‌گردد.

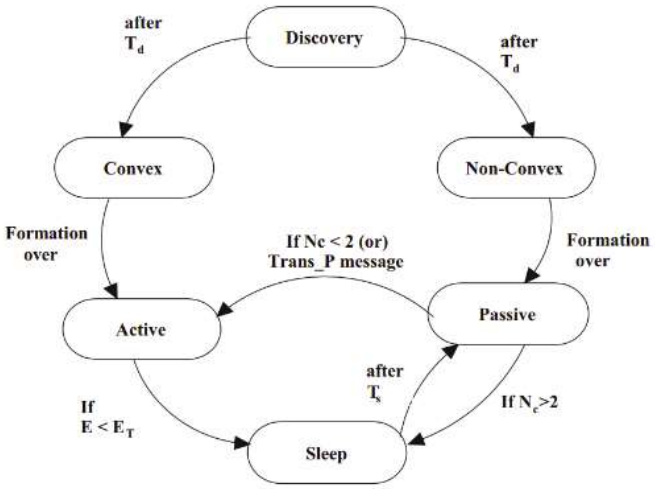
$$E_r = E_i - [I_a t_a + I_l t_l + I_t t_t + I_r t_r] \times V. \qquad (۱۵)$$

در رابطه بالا E_r انرژی باقی‌مانده گره، I_a و t_a به ترتیب میزان جریان مصرفی حالت فعال در واحد زمان و مجموع زمان فعال بودن گره، I_l و t_l به ترتیب میزان جریان مصرفی حالت توان پایین در واحد زمان و مجموع زمان خواب بودن گره، I_t و t_t به ترتیب میزان جریان مصرفی و مجموع زمان در فرآیند ارسال داده، I_r و t_r به ترتیب میزان جریان مصرفی و مجموع زمان در فرآیند دریافت داده و در نهایت V برابر با ولتاژ سیستم است. همچنین مقدار آستانه انرژی در این روش برابر ضریبی از انرژی اولیه گره تعیین شده است. این ضریب برابر با 0.5 در نظر گرفته شده است. در صورتی که انرژی گره‌ای از این حد آستانه پایین تر آید، آن گره وارد وضعیت توان پایین خواهد شد. در حقیقت گره‌هایی که در وضعیت توان پایین قرار می‌گیرند جهت حفظ انرژی توسط الگوریتم مدیریت می‌شوند. این الگوریتم گره‌ها را به دو دسته گره‌های محدب و گره‌های غیر محدب تقسیم می‌کند. گره‌های محدب با استفاده از الگوریتم پوش محدب[۵۱]، تعیین می‌شوند. این گره‌ها در اکثر مواقع در حالت فعال قرار دارند. الگوریتم پوش محدب و رویکرد آن در شکل ۱۷ قابل مشاهده است. در حقیقت اتصال گره‌هایی که الگوریتم پوش محدب آن‌ها را انتخاب می‌کند به یکدیگر موجب پدید آمدن یک شکل هندسی محدب شده است که به عنوان یک مرز، تمام گره‌ها را احاطه نموده است. یک شکل هندسی محدب شکلی است که با انتخاب هر دو نقطه در آن، در اتصال آن دو نقطه با یک خط راست، خط ترسیم شده از شکل خارج نشود. در شکل ۱۷ تفاوت مفهوم شکل محدب و غیرمحدب نمایش داده شده است.



شکل ۱۸ . تفاوت مفهوم محدب و غیرمحدب در اشکال هندسی

روال جابجایی بین وضعیت‌ها در الگوریتم MRL-SCSO در شکل ۱۸ آمده است. در وضعیت کشف، گره‌های محدب شناسایی شده و اگر گره‌جاری یک گره محدب باشد در وضعیت فعال و اگر غیر محدب باشد در وضعیت غیر فعال قرار می‌گیرد. در صورتی که یک گره در وضعیت فعال قرار داشته باشد و انرژی باقی‌مانده آن از آستانه انرژی در نظر گرفته شده پایین تر بیاید، به خواب می‌رود. همچنین گره‌هایی که به خواب رفته اند در یک دوره زمانی مشخص همچون T_s وارد وضعیت غیر فعال می‌شوند. در صورتی که تعداد همسایگان فعال گره جاری بیش از دو باشد، گره به حالت به خواب رفته باز می‌گردد. همچنین اگر تعداد همسایگان گره جاری که در وضعیت غیر فعال قرار دارد کمتر از دو بوده و یا در صورتی که به علت سنگینی بار شبکه، شبکه درخواست گره همسایه در دسترس و فعال نماید(این درخواست با نام Trans_P شناخته شده است.) وضعیت گره، به حالت فعال تغییر می‌کند.



شکل ۱۹ . نمودار تغییر حالت روش MRL-SCSO [۴۹]

۳.۹ روش RLBR[۵۲]

در این روش[۵۴]، پس از انجام پیکربندیهای اولیه شبکه، هر گره در انتظار دریافت یک بسته می‌ماند. با دریافت یک بسته، اطلاعات آن را استخراج نموده و جدول همسایگان آن گره را بر اساس اطلاعات به دست آمده به روز رسانی می‌کند.

در صورتی که هیچ گره بازپخشی در نزدیکی گره حاضر وجود نداشته باشد، بسته دور انداخته می‌شود. در غیر این صورت در

[51] Convex-hull

[52] Reinforcement-learning-based Routing

ابتدا وجود گره چاهک در محدوده ارتباطی گره حاضر، بررسی می‌شود. در صورتی که گره در این محدوده قرار داشته باشد، بسته مستقیما به سمت گره چاهک ارسال می‌گردد. در غیر این صورت، اگر گره همسایه‌ای به عنوان نماینده ارسال بسته به آن وجود نداشته باشد، در صورت وجود انرژی کافی با تنظیم قدرت انتقال سعی می‌شود که داده به صورت مستقیم به گره چاهک ارسال شود اما در صورت عدم وجود انرژی کافی، بسته دور انداخته می‌شود. در صورتی که گره نماینده‌ای برای ارسال داده به سمت آن وجود داشته باشد پارامتر Q-Value به ازای تمامی این نمایندگان به دست آمده و گره‌ای که دارای بیشترین مقدار Q-Value باشد به عنوان گره بازپخش بعدی انتخاب می‌گردد. در نهایت با به روز رسانی مقادیر Q-Value و hop count و همچنین packet header، بسته به سمت گره بازپخش بعدی انتخابی، ارسال می‌شود. مقدار Q-Value توسط روابط (۱۵) و (۱۶) به دست آمده است.

$$Q_{new}(cur,nbr) = (1-\alpha)Q_{old}(cur,nbr) + \alpha\big(R(cur,nbr) + Q(nbr)\big). \quad (۱۵)$$

$$Q(cur) = max_{nbr\in N} Q(cur,nbr). \quad (۱۶)$$

α نرخ یادگیری و $R(cur,nbr)$ نشان دهنده میزان پاداش دریافتی در ارسال حاضر به همسایه انتخابی است. همچنین این میزان پاداش توسط رابطه (۱۷) قابل محاسبه است.

$$R(cur,nbr) = E(nbr)/(d^n(cur,nbr) \times h(nbr)) \quad (۱۷)$$

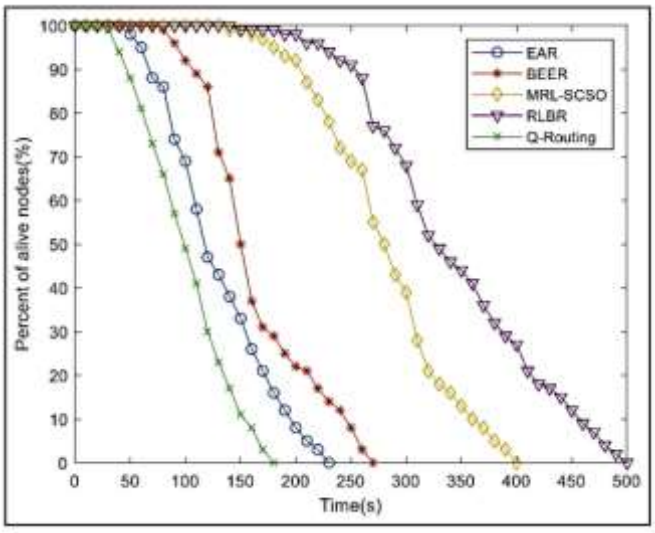
$d(\mathrm{cur,nbr})$ فاصله اقلیدسی بین گره جاری و همسایه انتخاب شده است و $\mathrm{E(nbr)}$ انرژی باقی‌مانده گره همسایه انتخابی می‌باشد. همچنین $\mathrm{h(nbr)}$ تعداد گام باقی‌مانده از گره همسایه انتخابی تا گره چاهک است. پارامتر n بر اساس رابطه (۱۸) به دست می‌آید.

$$n = \begin{cases} 2 & d \le d_0. \\ 4 & otherwise. \end{cases} \quad (۱۸)$$

d_0 یک آستانه برای فاصله بین دو گره است که اگر فاصله آن‌ها از این آستانه کوچکتر باشد پارامتر n برابر ۲ و در غیر این صورت برابر با ۴ در نظر گرفته می‌شود. همانطور که مشخص است این روال موجب می‌شود که پاداش به ازای دو گره با فاصله‌ای کمتر از حد آستانه بیش از پاداش به ازای گره‌هایی با فاصله بیش از آستانه در نظر گرفته شود. این روش در نهایت با روش‌های EAR، BEER[۵۳]، MRL-SCSO و Q-Routing مقایسه شده و برتری خود را از منظر نرخ کاهش تعداد گره‌های فعال، طی زمان و اولین زمان پدید آمدن گره جدا شده، تعداد بسته‌های ارسالی در واحد زمان، طول عمر شبکه و در نهایت نسبت نرخ ارسال بسته به انرژی مصرفی، اعلام می‌دارد.

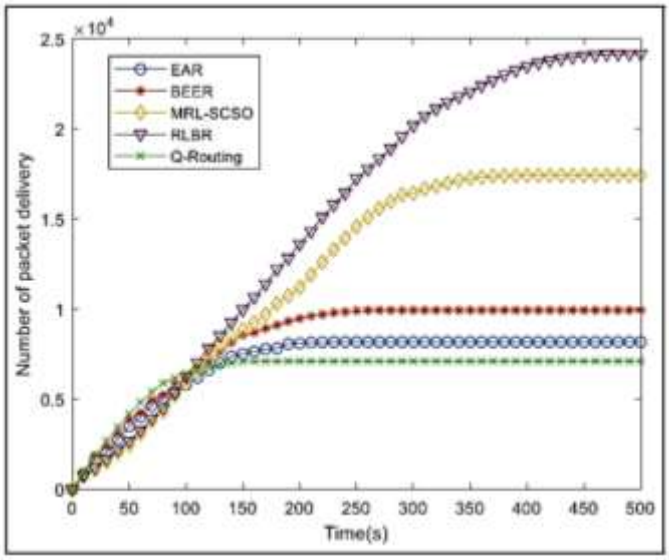
[53] Balanced Energy Efficient Rectangular Routing Protocol

مقایسه نرخ زنده ماندن گره ها در روش های مذبور در شکل ۲۰ [۲۳] آمده است.



شکل ۲۰. مقایسه نرخ زنده ماندن گره ها بر حسب زمان

همچنین مقایسه نرخ انتقال داده در شکل ۲۱ آمده است.



شکل ۲۱. مقایسه نرخ انتقال داده بر حسب زمان

۳.۱۰ روش های Actor-Critic

در این دسته از روش ها [۵۵] هنگامی که یک عامل اقداماتی را انجام میدهد و در یک محیط حرکت میکند، یاد میگیرد که وضعیت مشاهده شده محیط را به دو خروجی ممکن ترسیم میکند.

الف) اقدام توصیه شده: برای هر یک از اعمال ممکن در فضای عمل، احتمالی در نظر گرفته میشود. یکی از بخش هایی در عامل که به خروجی مذبور، عکس العمل نشان میدهد، بازیگر (Actor) نامیده میشود.

ب) پاداش های تخمینی مربوط به آینده: مجموع تمامی پاداش هایی که پیش بینی میشود که در آینده دریافت شود. یکی از بخش هایی در عامل که به خروجی مذبور، عکس العمل نشان میدهد منتقد (Critic) نامده میشود.

بازیگر و منتقد یاد میگیرند که وظایف خود را به گونه ای انجام دهند که اقدامات توصیه شده برای بازیگر بتواند پاداش را به حداکثر برساند.

الارغم کارایی مناسب روش های Actor-critic اما تا این لحظه هیچ پژوهشی در زمینه پروتکل های مسیریابی با استفاده از این روش ارائه نشده است.

۳.۱۱ روش های مبتنی بر Policy Gradient

روش‌های مبتنی بر Policy Gradient [۵۶] نوعی از تکنیک‌های یادگیری تقویتی هستند که بر بهینه‌سازی سیاست‌های پارامتری با توجه به بازده مورد انتظار (پاداش تجمعی بلندمدت) با نزول گرادیان تکیه دارند. آنها از بسیاری از مشکلاتی که رویکردهای یادگیری تقویتی سنتی را مخدوش کرده اند، مانند فقدان تضمین یک تابع ارزش، مشکل حل نشدنی ناشی از اطلاعات وضعیت نامشخص و پیچیدگی ناشی از حالت ها و اقدامات مستمر، رنج نمی برند.

یکی از چالش های یادگیری در Q-Learning پیوسته بودن اقدامات است. یکی از روش هایی که توانسته است برای این چالش راه حل مطلوبی را ارائه نماید DDPG [۵۴]است. درست مانند روش Actor-Critic ، در این روش نیز از دو شبکه بازیگر و منتقد بهره گرفته میشود. این روش با استفاده از به روز رسانی تدریجی شبکه های هدف و بهره گیری از تجربیات انباشته شده در گذشته سعی بر بهبود سیاست انتخابی دارد.

همچون روش Actor-Critic، روش های مبتنی بر Policy Gradient با این حال که در بسیاری از کاربرد ها مفید واقع شده اند اما هنوز در حوزه مسیریابی در شبکه های حسگر بیسیم از آن ها استفاده قابل توجهی صورت نپذیرفته است.

٤ نتیجه‌گیری

در مقاله حاضر سعی شد تا ضمن بیان و معرفی انواع دسته بندی پروتکل‌های مسیریابی در شبکههای حسگر بی‌سیم به معرفی تکنیک یادگیری تقویتی به عنوان روشی جهت همگرایی بهتر به یک روال مسیریابی بهینه در جهت افزایش طول عمر شبکه‌های حسگر بی‌سیم در کنار حفظ بهره‌وری آن بپردازیم. نشان داده شد که توجه به مواردی از جمله امکان برنامه‌ریزی حالت هر یک از گره‌ها، خوشه‌بندی و انتخاب نماینده برای هر خوشه، حفظ اطلاعات مسیریابی قبلی در جهت بهره‌گیری از آن در مسیریابی‌های آینده، لزوم اهمیت دهی به پارامترهایی همچون کیفیت سرویس و مواردی از این جمله، در کنار بهره گیری از تکنیک یادگیری تقویتی می‌تواند تاثیر بسزایی در عملکرد و افزایش طول عمر شبکه‌های حسگر بی‌سیم داشته باشد. بدیهی است که با در نظر گرفتن توابع پاداش و ارزیابی دقیق تر می‌توان در توسعه روش‌های هوشمندتر قدم برداشت.

[54] Deep Deterministic Policy Gradient

جدول ۱. مقایسه پروتکل های مسیریابی مبتنی بر یادگیری تقویتی

پروتکل مسیریابی	داده محور	سلسله مراتبی	مبتنی بر مکان	مبتنی بر کیفیت سرویس	مبتنی بر جریان شبکه	بهره‌گیری از تجمیع داده‌ها	ویژگی‌ها
Q-Routing			✓		✓		• تلاش در جهت کمینه کردن مدت زمان ارسال بسته • در نظر گرفتن یک تابع تخمین بر مبنای شبکه‌های عصبی • کارایی مناسب در شبکه‌های دارای تغییرات دینامیکی
مسیریابی وفقی (AdaR)			✓			✓	• امکان بهینه‌سازی اهداف چندگانه • استقلال از مدل مسئله • امکان یافتن سیاست بهینه با تعداد پایین تلاش(جستجو) • کارایی مناسب در شبکه‌های دارای تغییرات دینامیکی
ATP		✓			✓		• بهره گیری از درخت پوشا در یافتن مسیرهای بهینه • در نظر گرفتن Q-Value همسایگان در روال مسیریابی • کارایی مناسب در شبکه‌های دارای تغییرات دینامیکی
FROMS		✓					• توانایی پردازش و یافتن مسیر بهینه به ازای چندین گره هدف • بهره‌گیری از درخت به اشتراک گذاری مسیر و هرس نمودن آن.
QELAR	✓				✓		• مورد کاربرد در شبکه‌های حسگر بی‌سیم زیر آب • بهره‌گیری از روش ارسال پیام تصدیق
DACR	✓			✓	✓		• بهره‌گیری از روش مسیریابی بردار فاصله بر اساس تقاضا
FTIEE		✓	✓				• تعیین ظرفیت خوشه‌ها بر اساس فاصله آن‌ها از خوشه شامل گره چاهک • کارایی مناسب در شبکه‌های دارای تغییرات دینامیکی
MRL-SCSO					✓		• تعیین هوشمندانه وضعیت کاری هر گره
RLBR	✓				✓		• دادن بازخورد بسته اطلاعاتی و تنظیم قدرت انتقال

۵ مراجع

[1] J.N. Al-Karaki, A.E. Kamal, "Routing techniques in wireless sensor networks: a survey", IEEE, 2004, vol. 11, no. 6, pp. 6-28.

[2] G. Anastasi, M. Conti, M. Di. Francesco, A. Passarella, "Energy conservation in wireless sensor networks: A survey", Elsevier, 2009, vol.7, pp. 537-568.

[3] J. Yick, B. Mukherjee, D. Ghosal, "Wireless sensor network survey", Elsevier, 2008, vol. 52, no. 12, pp. 2292-2330.

[4] X. Liu, "A Survey on Clustering Routing Protocols in Wireless Sensor Networks", Sensors, 2012, vol. 12, no. 8, pp. 11113-11153.

[5] G. Dhand, S. S. Tyagi, “Data aggregation techniques in WSN: Survey”, 2nd International Conference on Intelligent Computing, Communication & Convergence (ICCC-2016), Odisha, India, January 2016.
[6] C. Cappiello, F. A. Schreiber, “Quality- and energy-aware data compression by aggregation in WSN data streams”, IEEE International Conference on Pervasive Computing and Communication, Galveston, TX, USA, March 2009.
[۷] و. درهمی، ف. اعلمیان هرندی، م. ب. دولت‌شاهی، "یادگیری تقویتی"، انتشارات دانشگاه یزد، ۱۳۹۶، ISBN: 978-600-8571-26-1
[8] A. Gosavi, “Reinforcement Learning: A Tutorial Survey and Recent Advances”, INFORMS journal on computing, 2008, vol. 21, no. 2, pp. 177-345.
[9] K. Arulkumaran, M. P. Deisenroth, M. Brundage, A. A. Bharath, “Deep Reinforcement Learning: A Brief Survey”, IEEE, 2017, vol. 34, no. 6, pp. 26-38.
[10] J. Kober, J. A. Bagnell, J. Peters, “Reinforcement Learning in Robotics: A Survey”, International Journal of Robotics Research, 2013, vol. 32, no. 11.
[11] SS. Keerthi, B. Ravindran, “A tutorial survey of reinforcement learning”, Springer, 1994, vol. 19, no. 6, pp. 851-889.
[12] L. Peshkin, V. Savova, “Reinforcement Learning for Adaptive Routing”, Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02, Honolulu, HI, USA, May 2002.
[13] S. Natarajan, P. Tadepalli, “Dynamic preferences in multi-criteria reinforcement learning”, Proceedings of the 22nd international conference on Machine Learning, USA, August 2005.
[14] D. Ernst, P. Geurts, L. Wehenkel, “Tree-Based Batch Mode Reinforcement Learning”, Journal of Machine Learning Research, 2005, vol. 6, pp. 503-556.
[15] Y. Niv, “Reinforcement learning in the brain”, Journal of Mathematical Psychology, 2008, vol. 53, pp. 139-154.
[16] W. D. Smart, L. Pack Kaelbling, “Effective reinforcement learning for mobile robots”, Proceedings IEEE International Conference on Robotics and Automation, Washington, DC, USA, May 2002.
[17] P. Dayan, Y. Niv, “Reinforcement learning: The Good, The Bad and The Ugly”, Elsevier, 2008, vol. 18, no. 2, pp. 185-196.
[19] N. Mazyavkina, S. Sviridov, S. Ivanov, E. Burnaev, “Reinforcement learning for combinatorial optimization: A Survey”, Elsevier, , vol. 134, August. 2021, DOI. https://doi.org/10.1016/j.cor.2021.105400
[20] J.Czech, “Distributed methods for Reinforcement Learning Survey”, springer, pp. 151-161, 2021, DOI. 10.1007/978-3-030-41188-6_13.
[21] S. Pateria, et al, “Hierarchical Reinforcement Learning: A Comprehensive Survey”, ACM, Computing Surveys, vol.54, no.5, pp. 1-35, 2021, https://doi.org/10.1145/3453160.
[22] A.T.D. Perera, P. Kamalaruban, “Applications of Reinforcement Learning in energy systems”, Elsevier, vol.137, 2021, https://doi.org/10.1016/j.rser.2020.110618.
[23] K. Akkaya, M. Younis, “A survey on routing protocols for wireless sensor networks”, Elsevier, 2003, vol. 3, pp. 325–349.
[24] W.R. Heinzelman, J. kulik, H. Balakrishnan, “Adaptive protocols for information dissemination in wireless sensor networks”, ACM/IEEE international conference on Mobile computing and networking, Seattle Washington USA, September 1999.
[25] C. Intanagonwiwat, R. Govindan, D. Estrin, “Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks”. Proceedings of the 6th Annual ACM/IEEE International Conference on Mobile Computing and Networking, USA, June 2000.
[26] D. Estrin, R. Govindan, J. Heidemann, “Next Century Challenges: Scalable Coordination in Sensor Networks”, 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking, USA, 1999.
[27] R. C. Shah, J. M. Rabaey, “Energy Aware Routing for Low Energy Ad Hoc Sensor Networks”, IEEE Wireless Communications and Networking Conference, Orlando, FA, USA, March 2002.
[28] D. Braginsky, E. Forward, “Rumor routing algorithm for sensor networks”. 1st ACM international workshop on Wireless sensor networks and applications, atlanta, Georgia, USA, September 2002.
[29] C. Schurgers, B. Srivastava, Energy Efficient Routing In Wireless Sensor Networks, IEEE, 2001.
[30] M. Chu, H. Haussecker, F. Zhao, “Scalable Information-Driven Sensor Querying and Routing for ad hoc Heterogeneous Sensor Networks”, The International Journal of High

[43] I.F. Akyildiz, Su. WY, Y. sankarasubramaniam, E. Cayirci. Wireless sensor networks: a survey, Elsevier Computer Networks, 2002.
[44] K. Akkaya, M. Younis, “Energy and QoS aware routing in wireless sensor networks”, Proceedings of the IEEE Workshop on Mobile and Wireless Networks, Lake Como, Italy, May 2003.
[45] T. He, “SPEED: a stateless protocol for real-time communication in sensor networks”, Proceedings of International Conference on Distributed Computing Systems, Rhode island, USA, May 2003.
[46] J. A. Boyan, M. L. Littman, “Packet routing in dynamically changing networks: a reinforcement learning approach”, Proceedings of the international conference on neural information processing systems, USA, December 1993.
[47] P. Wang, T. Wang, “Adaptive routing for sensor networks using reinforcement learning”, Proceedings of the IEEE international conference on computer & information technology, East Lansing, MI, September 2006.
[48] Y. Zhang, Q. Huang, “A learning-based adaptive routing tree for wireless sensor networks”, Proceedings of the IEEE 3rd Consumer Communications and Networking Conference, USA, May 2006.
[49] A. Forster, A. Murphy, “FROMS: feedback routing for optimizing multiple sinks in WSN with reinforcement learning”, Proceedings of the international conference on intelligent sensors, Melbourne, Qld, Australia, December 2007.
[50] T. Hu, Y. Fei, “QELAR: a machine-learning-based adaptive routing protocol for energy-efficient and lifetime-extended underwater sensor networks”, IEEE, 2010, vol. 9, no. 6, pp. 796–809.
[51] M. Razzaque, M. Ahmed, C. Hong, “QoS-aware distributed adaptive cooperative routing in wireless sensor networks”, Elsevier Ad Hoc Networks, 2014, vol. 19, no. 8, pp. 28-42.
[52] F. Kiani, E. Amiri, M. Zamani, “Efficient intelligent energy routing protocol in wireless sensor networks”, International Journal of Distributed Sensor Networks, 2014, pp. 5-27.
[53] A. Renold, S. Chandrakala, “MRL-SCSO: multi-agent reinforcement learning-based self-configuration and self optimization protocol for unattended wireless sensor networks”, Springer, 2016, vol. 96, pp. 5061–5079.
[54] W. Guo, C. Yan, T. Lu, “Optimizing the lifetime of wireless sensor networks via reinforcement-learning-based routing”, International Journal of Distributed Sensor Networks, 2019, vol. 15, no. 2.
[55] D. Oh, D.Adams, et.al, “Actor-critic Reinforcement Learning to estimate the optimal Networks”, ACM SIGMOD Record (SIGMOD REC), USA, 2002.
[32] N. Sadagopan, B. Krishnamachari, A. Helmy, “The ACQUIRE Mechanism for Efficient Querying in Sensor Networks”, IEEE, 2003, pp. 149-155.
[33] W. Heinzelman, A. Chandrakasan, H. Balakrishnan,“Energy-efficient communication protocol for wireless sensor networks”, Proceeding of the Hawaii International Conference System Sciences, USA, January 2000.
[34] S. Lindsey, C. S. Raghavendra, “PEGASIS: Power-Efficient Gathering in Sensor Information Systems”, IEEE Aerospace Conference, Big Sky, MT, USA, February 2002.
[35] A. Manjeshwar, D. P. Agrawal, “TEEN: A Routing Protocol for Enhanced Efficiency in Wireless Sensor Networks”, IEEE International Parallel & Distributed Processing Systems, 2001.
[36] M. Younis, M. Youssef, K. Arisha, “Energy-aware routing in cluster-based sensor networks”, Proceedings of the 10th IEEE/ACM International Symposium on Modeling, Rennes, FR, October 2002.
[37] L. Subramanian, R. H. Katz, “An architecture for building self configurable systems”, Proceedings of IEEE/ACM Workshop on Mobile Ad Hoc Networking and Computing, Boston, August 2000.
[38] V. Rodoplu, H. M. Teresa, “Minimum Energy Mobile Wireless Networks”, IEEE Journal on Selected Areas in Communications, 1998.
[39] Y. Xu, J. Heidemann, D. Estrin, “Geography-informed Energy Conservation for Ad Hoc Routing”, Proceedings of the 7th Annual ACM/IEEE International Conference on Mobile Computing and Networking, Seattle, Washington, USA, June 2001.
[40] Y. Yu, D. Estrin, “Geographical and energy aware routing: a recursive data dissemination protocol for wireless sensor networks”, Proceedings of the Seventh Annual ACM/IEEE International Conference on Mobile Computing and Networking, Seattle, Washington, USA, October 2001.
[41] J.H. Chang, L. Tassiulas, “Maximum lifetime routing in wireless sensor networks”, Proceedings of the Advanced Telecommunications and Information Distribution Research Program (ATIRP_2000), College Park, MD, USA, March 2000.
[42] F. Ye, et al, “A scalable solution to minimum cost forwarding in large scale sensor networks”, Proceedings of International Conference on Computer Communications and Networks (ICCCN), Scottsdale, Arizona, USA, October 2001.

operating conditions of the hydrocracking process", Elsevier, 2021, vol.149
[56] W. Jingda, W. Zhongbao, et.al, "Battery-involved Energy Management for Hybrid Electric Bus Based on Expert-assistance Deep Deterministic Policy Gradient Algorithm", IEEE, 2020, , vol. 69, p. 12786-12796, DOI: 10.1109/TVT.2020.3025627.

An Overview of Control and Routing Methods in Wireless Sensor Networks Using Reinforcement Learning

A. Forghani Elah Abadi, S. A. Asghari, M. B. Marvasti

Abstract

This article examines the control and routing protocols of wireless sensor networks that have been able to use reinforcement learning techniques, which is one of the machine learning methods, to achieve energy efficiency in the network. This technique is based on the method of reward and punishment, which is similar to the process of learning in children. Proper energy management and therefore increasing the lifetime of wireless sensor networks is always one of the main challenges of this type of network due to energy limitations in its nodes. The purpose of writing this article is to learn more about the methods presented. In this paper, various methods that try to use the reinforcement learning process to improve the behavior of wireless sensor networks and make them smarter are introduced and examined. Also, the evolution of these methods and the ratio of the superiority of each to the other have been examined.

Keywords:

Reinforcement Learning, Wireless Sensor Network, Energy Constraints, Lifetime, Reward and Punishment