

تخصیص طیف در شبکه‌های رادیوشناختی با استفاده از آتوماتای یادگیر

احسان کریم زاده^۱، پریسا رحمانی^۲

^۱ دانشجوی ارشد، مهندسی کامپیوتر دانشگاه آزاد اسلامی واحد الکترونیکی

^۲ هیئت علمی دانشگاه آزاد اسلامی واحد پردیس

نام و نشانی ایمیل نویسنده مسئول:

احسان کریم زاده

K7108478@gmail.com

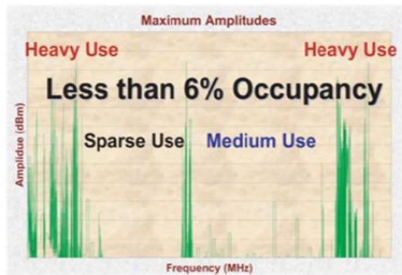
چکیده

سیاست‌های سنتی تخصیص طیف بصورت ایستا، کمبود شدید طیف و بهره‌برداری کم بعنوان یک مشکل شبکه‌های بی‌سیم است، که نیاز فوری به حل دارد برای حل این مشکل و استفاده مفیدتر از طیف، مفهومی بنام شبکه‌های رادیویی شناختی (CR) مطرح شد، تا بتوان از فضاهای خالی طیف مجوزدار استفاده کرد. یکی از مسایل مهم در شبکه‌های رادیو شناختی پروتکل‌های کنترل دسترسی به رسانه (MAC) است. و همچنین به دنبال روشی برای پیش‌بینی وضعیت کانال در شبکه‌های رادیویی شناختی هستیم. تکنیک‌های زیادی به‌عنوان اعمال هوش لازم برای یادگیری برای پیش‌بینی حالت کانال در شبکه‌های بی‌سیم مورد مطالعه و بررسی قرار گرفته‌اند از جمله تئوری بازی، محاسبات تکاملی، منطق فازی، تئوری قیمت گذاری و یادگیری ماشین که از متداول ترین روش‌های یادگیری می‌باشد. ما قصد داریم با استفاده از روش‌های هوشمند یک تخصیص طیف سازگار با تغییرات شبکه را در هر گره از شبکه داشته باشیم، برای این منظور از آتوماتای یادگیر استفاده می‌شود. هدف پایان نامه آرایه الگوریتمی برای تخصیص طیف در شبکه‌های رادیوشناختی است که فاقد یک کنترل‌گر مرکزی است. الگوریتم پیشنهادی قصد دارد احتمال دسترس پذیری هر یک از کانال‌ها را برای هر یک از کاربران ثانویه محاسبه کند و با گذشت زمان انتظار می‌رود که کاربران ثانویه بتوانند بصورت توزیع شده مطابق با ماهیت شبکه‌های رادیوشناختی از فضاهای خالی طیف استفاده بهینه را داشته باشد.

واژگان کلیدی: شبکه‌های رادیوشناختی، تخصیص طیف، پروتکل دسترسی به رسانه، آتوماتای

یادگیری

مقدمه



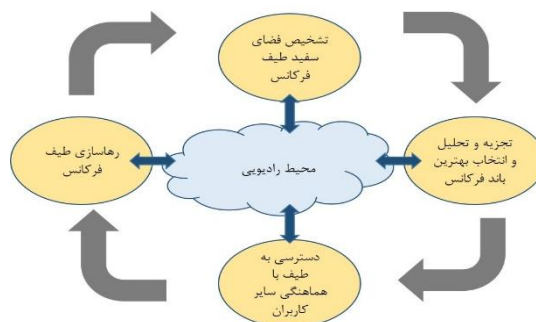
شبکه‌های بی‌سیم^۱ کنونی از یک سیاست تخصیص طیف، به صورت ثابت (ایستا) بهره می‌گیرند. که در آن، سازمان‌ها طیف فرکانس را به صورت بلند مدت به کاربران مجاز اختصاص می‌دهند. این سیاست بکار گرفته شده، به دلیل افزایش تقاضاهای تخصیص طیف، در برخی از باندهای خاص با کمبود شدید روبرو گردیده است. از سوی دیگر بخش عظیمی از طیف اختصاص داده شده، به طور پراکنده و نامناسب استفاده می‌شود، همین امر منجر به استفاده کم و غیر بهینه از مقدار قابل توجهی از طیف فرکانسی می‌گردد. همان طور که در شکل روبرو نشان داده شده، بخش عظیمی از طیف اختصاص داده شده توسط کاربران مجاز مورد استفاده قرار نمی‌گیرد. یعنی با توجه به مقتضیات زمانی و مکانی کمتر از شش درصد طیف فرکانسی توسط کاربران مجاز اشغال می‌شود.

بررسی گزارش‌هایی که توسط کمیسیون مرکزی مخابرات آمریکا^۲ (FCC)، در رابطه با اندازه‌گیری طیف فرکانس صورت گرفته، نشان می‌دهد، در باند صفر تا سه گیگاهرتز، رقابت شدیدی برای استفاده از طیف فرکانس وجود دارد. یک مشکل قابل توجه در این زمینه کمبود فیزیکی طیف و مدیریت آن است. از این رو برای حل مشکل ناکارآمدی استفاده از طیف فرکانس روش‌های دسترسی پویا پیشنهاد شد. به این ترتیب که اجازه داده می‌شود که در کنار کاربران مجوزدار، یکسری کاربر بدون مجوز از فضای بلا استفاده طیف فرکانس استفاده نمایند. به شرط آنکه مزاحمتی برای کاربران مجاز به وجود نیاید. عبارتی دیگر، تکنیک دسترسی به طیف فرکانس به صورت پویا استفاده گردد.

۱-۱ معرفی شبکه‌های رادیوشناختی^۳:

با توسعه سریع دستگاه‌های بی‌سیم و برنامه‌های کاربردی، شاهد تقاضاهای رو به رشد برای اختصاص طیف فرکانس هستیم. سیاست‌های سنتی تخصیص ایستای طیف، تبدیل به تنگنایی برای استفاده کارآمد از طیف فرکانسی گشته است و عملاً بخش بزرگی از طیف‌های فرکانس مجوزدار کمتر مورد استفاده قرار می‌گیرد. به همین دلیل مفهوم رادیوی هوشمند و یا همان رادیوشناختی، اولین بار توسط جوزف میتولا در سال ۱۹۹۸ اینگونه تعریف شد که عبارت است از: "دسترسی فرصت طلبانه به باندهایی که مالک آنها نیستیم. بدون اینکه تداخلی با کاربران مجوزدار در آن باندها داشته باشیم".

۱-۱-۱ وظایف شبکه رادیوشناختی دارای یک چرخه چهار مرحله‌ای مطابق با شکل ۲ است؛ که این چهار مرحله عبارتند از: ۱- تشخیص فضای سفید طیف فرکانسی، ۲- تجزیه و تحلیل و انتخاب بهترین باند فرکانس، ۳- دسترسی به طیف با هماهنگی سایر کاربران، ۴- رها سازی طیف فرکانس. نهایت آزادسازی طیف فرکانس^۷، هنگامی که یک کاربر اولیه ظاهر می‌گردد.



شکل ۱. چرخه شبکه رادیوشناختی

¹ Wireless network

² Federal Communications Commission (FCC)

³ Cognitive Radio Networks

⁴ Sensing : Real-time wide-band monitoring

⁵ Analasis : Rapid characterization of environment

⁶ Reasoning : Determine the best response strategy

⁷ Adaptation : transition to new operating parameters

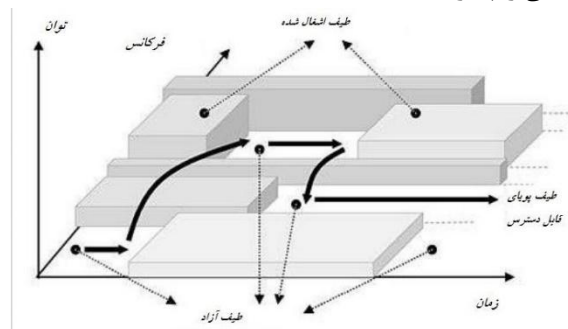
۱-۱-۲ دو ویژگی اصلی شبکه رادیوشناختی، عبارت اند از:

قابلیت شناخت

در این شبکه‌ها می‌توان از طریق تعامل با محیط‌های رادیویی، بخشی از طیف فرکانسی که در یک زمان یا مکان خاص مورد استفاده قرار نگرفته است را شناسایی نمود. همان‌طور که در شکل (۲) نشان داده شده است. در شبکه‌های رادیوشناختی، این توانایی وجود دارد که به طور موقت از طیف بلااستفاده که اصطلاحاً سوراخ طیف یا فضای سفید^۸ نامیده می‌شود، استفاده نماید. در نتیجه بهترین طیف بدون تداخل با کاربر مجاز انتخاب می‌شود و با کاربران دیگر به اشتراک گذاشته می‌شود. به عبارت دیگر، قابلیت شناخت به توانایی حس کردن و جمع‌آوری اطلاعات از محیط اطراف اشاره دارد.

قابلیت تنظیم مجدد

یکی از ویژگی‌های دیگر شبکه رادیوشناختی، قابلیت تنظیم مجدد است؛ که می‌تواند فناوری دسترسی مختلفی را برای برنامه‌ریزی ارسال و دریافت در انواع فرکانس‌ها را داشته باشد. بعبارتی دیگر، از طریق این قابلیت است که می‌تواند بهترین طیف فرکانس را انتخاب نموده و مناسب‌ترین پارامترهای عملیاتی را پیکربندی کند.



شکل ۲. مفهوم سوراخ طیف فرکانس

بنابراین کاربر یک شبکه رادیوشناختی باید قادر باشد، اعمال زیر را انجام دهد:

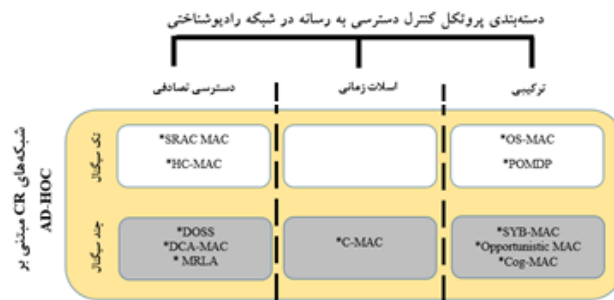
- (۱) قسمت‌هایی از طیف را که مورد استفاده دیگران نیست را تشخیص دهد؛
- (۲) بهترین کانال موجود را انتخاب کند؛
- (۳) دسترسی به کانال را با کاربران دیگر هماهنگ کند؛
- (۴) به محض تشخیص حضور یک کاربر مجاز در کانال از آن خارج شود؛

۲- پروتکل MAC^۹ در شبکه‌های رادیویی شناختی

پروتکل‌ها MAC شناختی، به منظور استفاده و بهره‌برداری مناسب از منابع طیف و بدون تاثیر بر خصوصیات کارایی و عملکرد کاربران اولیه طراحی و ساخته شده است. استفاده از طیف اغلب به علت مشکلاتی که در پیش بینی ظهور آنها وجود دارد، نیاز به روش‌های تصادفی و اتفاقی دارد. بستر کاری مبتنی بر تکنیک‌های دستیابی هماهنگ نیز برای بسیاری از کاربردها و باندهای طیفی خصوصاً در بحث شبکه‌های حسگر و شبکه‌های بی‌سیم محلی، دوام‌پذیر و ماندگار نیستند. ما در این گزارش؛ توصیفی از پروتکل‌های دسترسی به رسانه، برای شبکه‌های CR مبتنی بر زیرساخت و ادهاک، در جدول (۱) ارائه دادیم و روش‌های موجود را به دسته‌ها زیر تقسیم نمودیم: (۱) پروتکل‌های دسترسی تصادفی، (۲) پروتکل‌های اسلات زمانی (۳) پروتکل‌های ترکیبی.

^۸ Spectrum white space

^۹ MAC (Media Access Control)



شکل ۳. دسته‌بندی پروتکل‌های دسترسی به رسانه در شبکه CR

- پروتکل‌های دسترسی تصادفی: پروتکل‌های دسترسی به رسانه این دسته، به همگام‌سازی زمانی نیاز ندارند و عموماً مبتنی بر CSMA/CA^{۱۰} (دسترسی چندگانه با اجتناب از برخورد) هستند. کاربر CR بر باند طیفی نظارت می‌کند تا بفهمد که چه زمانی ارسالی توسط بقیه کاربران CR انجام نمی‌شود و پس از یک مدت زمان backoff (به منظور جلوگیری از ارسال‌های همزمان) عمل ارسال داده را انجام می‌دهد.
- پروتکل‌های اسلات زمانی: این پروتکل‌های دسترسی به رسانه، به همگام‌سازی گسترده نیاز دارند. در این نوع پروتکل‌ها، که برای کانال کنترلی و ارسال داده، زمان را به چند اسلات تقسیم می‌کند.
- پروتکل‌های ترکیبی: این پروتکل‌ها از یک ارسال حدوداً اسلاتی استفاده می‌کنند و سیگنال کنترلی در هر اسلات زمانی همگام شده تولید می‌شود. اما ارسال بعدی، ممکن است طبق الگوی دسترسی تصادفی به کانال و بدون عملیات همگام‌سازی انجام شود. در یک روش متفاوت، ممکن است دوره‌های کنترل و ارسال داده دارای مدت زمان‌های از قبل تعریف شده که شامل، یک سوپرفریم مشترک بین همه کاربران شبکه باشد. در دوره کنترلی یا ارسال داده، ممکن است دسترسی به کانال کاملاً تصادفی باشد.

۳- هوش و یادگیری ماشینی

یادگیری به عنوان یک جنبه مهم از رفتار هوشمند شناخته شده است. در طول چند دهه گذشته، فرآیند یادگیری که قبلاً توسط روانشناسان یا زیست‌شناسانی که آزمایش‌های یادگیری را بر روی حیوانات و انسان‌ها انجام می‌دهند. یک تعریف مفید مهندسی در زمینه یادگیری به این صورت است که: به یک ماشین یا سیستم در صورتی گفته می‌شود یادگیر، که با تجربه کسب کرده در طول زمان و بدون داشتن اطلاعات کامل از محیط عملیاتی خودش، بتواند عملکرد خود را بهبود ببخشد.

به‌طور کلی، یادگیری به سه شکل قابل انجام است: [25]

- ✚ یادگیری با نظارت^{۱۱}: در این حالت، ناظر (معلم) عمل یا خروجی مطلوب برای هر ورودی آموزشی را در اختیار عامل قرار می‌دهد و عامل از این طریق یاد می‌گیرد که برای ورودی‌های جدید، خروجی مورد انتظار را ارائه دهد. استفاده از این روش، اجرای فرآیند یادگیری را بسیار ساده می‌کند.
- ✚ یادگیری بی‌نظارت^{۱۲}: برخلاف روش یادگیری بانظارت، در این روش، بازخوردی به ازاء هر عمل به عامل داده نشده و عامل با سعی و خطا برای تولید خروجی مطلوب و یافتن رابطه میان ورودی و خروجی تلاش می‌کند. این روش، مشکل‌ترین راه برای یادگیری توسط عامل می‌باشد.
- ✚ یادگیری تقویتی^{۱۳}: در این روش، فرد خبره در ازاء هر عمل انجام شده توسط عامل، مقداری عددی به‌عنوان پاداش یا جریمه (سیگنال تقویتی) به عامل ارائه می‌دهد و عامل از طریق این سیگنال، میزان سودمندی عمل خود را درک کرده و اعمال بعدی را در جهت حداکثر کردن میزان پاداش دریافتی تنظیم می‌نماید. این روش بسیار انعطاف‌پذیر است و به‌خصوص هنگامی استفاده می‌شود که هیچ نمونه‌ای از آن‌چه باید یاد گرفته شود، در اختیار عامل نباشد و عامل مجبور باشد یادگیری را بدون داشتن مدلی از محیط آغاز نماید. مشکل اساسی یادگیری تقویتی به مشخصات معلم مربوط است؛ مثلاً توزیع احتمال سیگنال تقویت مرتبط با اقدامات متفاوت نامشخص هستند. در حوزه بهینه‌سازی

¹⁰ Carrier Sense Multiple Access/Collision Avoidance

¹¹ Supervised Learning

¹² Unsupervised Learning

¹³ Reinforcement Learning (RL)

کارایی، این مسئله را می‌توان به عنوان یک مسئله تپه‌نوردی تصادفی یا بهینه‌سازی تصادفی آنلاین در نظر گرفت. ولی مزیت اصلی یادگیری تقویتی نسبت به سایر روش‌های یادگیری عدم نیاز به هیچگونه اطلاعاتی از محیط (بجز سیگنال تقویتی) است.

۳-۱- اتوماتای یادگیری

یکی از مدل‌هایی که براساس یادگیری تقویتی عمل می‌نماید و توسط Tsetlin و همکارانش ارایه شده، اتوماتای یادگیر می‌باشد اتوماتای یادگیر، ماشینی است که می‌تواند تعدادی متناهی عمل را انجام دهد. هر عمل انتخاب شده توسط یک محیط احتمالی ارزیابی می‌شود و نتیجه ارزیابی در قالب سیگنالی مثبت یا منفی به اتوماتا داده می‌شود و اتوماتا از این پاسخ در انتخاب عمل بعدی تأثیر می‌گیرد. هدف نهایی این است که اتوماتا یاد بگیرد تا از بین اعمال خود، بهترین عمل را انتخاب کند. بهترین عمل، عملی است که احتمال دریافت پاداش از محیط را به حداکثر برساند. کارکرد اتوماتای یادگیر در تعامل با محیط، در شکل ۱۲ مشاهده می‌شود. [26]



شکل ۱۴. ارتباط بین اتوماتا یادگیر و محیط

یک اتوماتای یادگیر از دو قسمت اصلی تشکیل شده است.

الف: یک اتوماتای تصادفی با تعداد محدودی عمل و یک محیط تصادفی که اتوماتا با آن در ارتباط است.

ب: الگوریتم یادگیری که اتوماتا با استفاده از آن عمل بهینه را یاد می‌گیرد.

یک اتوماتای تصادفی بصورت پنج‌تایی $SA \equiv \{\alpha, \beta, F, G, \phi\}$ تعریف می‌شود، $\alpha \equiv \{\alpha_1, \alpha_2, \dots, \alpha_r\}$ مجموعه اعمال اتوماتا، $\beta \equiv \{\beta_1, \beta_2, \dots, \beta_r\}$ مجموعه ورودی‌های اتوماتا، تابع تولید وضعیت جدید، $G \equiv \phi \rightarrow \alpha$ تابع خروجی که وضعیت فعلی را به خروجی بعدی نگاشت می‌کند و $\phi(n) \equiv \{\phi_1, \phi_2, \dots, \phi_k\}$ مجموعه وضعیت‌های داخلی اتوماتا در لحظه n ، می‌باشند. مجموعه α شامل خروجی‌های (اعمال) اتوماتا است که اتوماتا در هر گام یک عمل از r عمل این مجموعه را برای اعمال بر محیط انتخاب می‌نماید. مجموعه ورودی‌ها (β) ورودی‌های اتوماتا را مشخص می‌کند. توابع F و G وضعیت فعلی ورودی را به خروجی بعدی (عمل بعدی) اتوماتا نگاشت می‌کنند. اگر نگاشت‌های F و G قطعی باشند، اتوماتا یک اتوماتای قطعی^{۱۴} نامیده می‌شود. در حالتیکه نگاشت‌های F و G تصادفی باشند، اتوماتا یک اتوماتای تصادفی نامیده می‌شود.

۳-۲ ویژگی‌های اتوماتای یادگیر

اتوماتاهای یادگیر دارای ویژگی‌های برتر زیر می‌باشند:

- اتوماتاهای یادگیر در شرایطی که هیچگونه اطلاعاتی در دسترس نیست، به خوبی عمل می‌کند.
- اتوماتاهای یادگیر در شرایطی که عدم اطمینان وجود دارد، دارای عملکرد خوبی می‌باشد.
- اتوماتاهای یادگیر عمل جستجو را در فضای احتمال انجام می‌دهد.
- اتوماتاهای یادگیر برای بهبود وضعیت خود، در هر مرحله به یک بازخورد ساده از محیط نیاز دارد.
- اتوماتاهای یادگیر به عنوان مدلی برای یادگیر در محیط‌های توزیع‌شده بسیار کارساز است.
- اتوماتاهای یادگیر در کاربردهای بهینه‌سازی، نیاز به استفاده از معیار کارایی و مشتق‌پذیری ندارد.
- اتوماتاهای یادگیر دارای بار محاسباتی بسیار کم می‌باشد.
- روش‌های تحلیل ریاضی قدرتمندی برای تجزیه و تحلیل اتوماتاهای یادگیر وجود دارد.

۴ - روش پیشنهادی برای تخصیص طیف در شبکه‌های رادیوشناختی

روش پیشنهادی ارایه یک روش تخصیص طیف در شبکه رادیو شناختی است که، در این شبکه کاربران رادیوشناختی می‌توانند در هر دو باند با مجوز و بدون مجوز بدون نیاز به اتصال با ایستگاه مرکزی ثانویه، یک ارتباط را با سایر کاربران ثانویه برقرار نمایند. به عنوان مثال، گره‌های موجود در یک شبکه ادهاک ممکن است دارای درجه بالای سیاری به همراه همگام‌سازی زمانی باشند، که منجر به سخت شدن

تطبیق حس کردن می‌شود. به همین دلیل ما برای تخصیص طیف در شبکه‌های رادیوشناختی بدون زیرساخت از روش آتوماتای یادگیر استفاده می‌کنیم که با گذشت زمان هر نود یاد می‌گیرد که با استفاده از آتوماتای یادگیری، دسترس پذیری طیف را در یک ناحیه داشته باشد و منجر به بهبود عملکرد الگوریتم تخصیص طیف فرکانسی با گذشت زمان شود. از طرفی به کارگیری آتوماتای یادگیر در تخصیص طیف و شناسایی طیف‌های مجوزدار در طول زمان منجر به کاهش سوچینگ میان کانال‌ها می‌شود. چرا که سوچینگ مکرر بین کانال‌های موجود بسیاری از مشکلاتی را مانند افزایش تاخیر، از دست دادن بسته و افزایش هزینه ارتباطات را ایجاد می‌کند. مخصوصاً در محیط‌هایی که مانند شبکه‌های رادیو شناختی بسیار پویا هستند این مساله منجر به کاهش چشمگیر کارایی شبکه و استفاده نامطلوب از طیف مجوزدار می‌شود.

در روش پیشنهادی مساله تخصیص طیف در سه قسمت بررسی می‌شود؛ در قسمت اول برای تخصیص طیف مجوز دار از مدل آتوماتای یادگیر L_{R-P} , L_{R-I} , L_{R-EP} استفاده می‌شود ولی از آنجاییکه در الگوریتم‌هایی مثل الگوریتم L_{R-I} ، بروزسانی احتمال عملیات در یک نقطه ثابت تنها به عمل انتخاب شده بستگی دارد؛ و به تاریخچه عملیات و تقویت‌های گذشته بستگی ندارد. در بخش دوم جهت بهبود عملکرد تخصیص طیف از الگوریتم‌های تخمین (Pursuit Estimator) استفاده می‌شود. در این دسته از روش‌های یادگیر بروزسانی احتمال اعمال (کانال‌ها) براساس تاریخچه اعمال انجام می‌شود. دلیل اصلی استفاده از این روش در جهت بهبود کارایی از منظر سرعت و دقت یادگیری است.

۱-۴ تخصیص طیف با استفاده از L_{R-I}

در هر نود {کاربر ثانویه} یک آتوماتای یادگیر گنجانده می‌شود. فرض می‌کنیم کل طیف قابل دسترس مجوز دار به یک سری کانال‌ها با عرض یکسان تقسیم شده‌اند. یعنی تعداد کانال‌ها بصورت پیش فرض $[ch_1, ch_2, \dots, ch_m]$ مشخص شده‌اند. الگوریتم به اینصورت عمل می‌کند که در ابتدای امر احتمال انتخاب هر کانال برای کاربران ثانویه یکسان است. در هر تکرار یک کاربر ثانویه از لیست کانال‌های موجود کانال را براساس مقدار توزیع احتمال آن عمل انتخاب می‌کند. حال با توجه به فیدبک دریافتی از محیط بردار احتمال کانال‌ها را به روز رسانی می‌کند. وقتی یک نود کانالی را از لیست اعمال انتخاب می‌کند چند حالت اتفاق می‌افتد. ۱- کانال نه در اختیار کاربر اولیه نه کابر ثانویه است. یعنی کانال آزاد است. در اینصورت مقدار فیدبک دریافتی برابر با ۱ در نظر گرفته می‌شود و بردار احتمال مطابق با فرمول (۳) به روزرسانی می‌شود. یعنی احتمال کانال انتخابی افزایش و احتمال بقیه تغییر نمی‌کند. ۲- کانال انتخابی در اختیار کابر ثانویه است در اینصورت مقدار فیدبک دریافتی ۰٫۵ در نظر گرفته می‌شود و احتمال اعمال بصورت فرمول (۱) به روزرسانی می‌شود. ۳- کانال انتخابی در اختیار کاربر اولیه باشد مقدار فیدبک دریافتی ۰ می‌شود و احتمال کانال انتخابی مطابق با فرمول (۲) به روز رسانی می‌شود. با گذشت زمان الگوریتم یادگیر رفتار کابر اولیه و کاربر ثانویه را در زمینه استفاده از طیف از طریق محاوره با محیط یاد می‌گیرد و در نهایت بهترین عمل برای نود ثانویه بدست می‌آید.

۱-۴-۴ الگوریتم تخصیص طیف با استفاده از L_{R-I}

کد الگوریتم بصورت زیر است:

۱- مقدار دهی اولیه پارامترها

- فرض کنید مجموعه اعمال آتوماتای موجود در هر کابر ثانویه بصورت زیر است، جاییکه r تعداد کانال هاست:

$$\alpha_i = \{ch_1, ch_2, \dots, ch_r\}$$

$$p(t)_i = \frac{1}{r} \text{ where } i = 1..r$$

- در شروع کار احتمال همه اعمال یکسان است :

$$2 \text{ - مراحل زیر را تکرار ، تا به شرط همگراشدن برسیم: } \{300 * (1 - p_{Ti}(t))\} \text{ CW}$$

- یک کانال ch_j از مجموعه اعمال α_i براساس توزیع احتمال آن انتخاب می‌شود؛
- فیدبک دریافتی از محیط را بصورت زیر محاسبه می‌شود:

$$\beta(t) = 0 \text{ اگر } ch_j \text{ اشغال و در اختیار کاربر اولیه است مقدار فیدبک دریافتی از محیط می‌شود.}$$

$$\beta(t) = 0.5 \text{ اگر } ch_j \text{ اشغال و در اختیار کاربر ثانویه است مقدار فیدبک دریافتی از محیط می‌شود.}$$

$$\beta(t) = 1 \text{ اگر } ch_j \text{ آزاد است. مقدار فیدبک دریافتی از محیط می‌شود.}$$

- سپس بردار احتمال را براساس فرمول (۱) به روزرسانی می‌شود.

$$\begin{aligned} p_i(t+1) &= p_i(t) + \lambda\beta(t)(1 - p_i(t)) \\ p_j(t+1) &= p_j(t) - \lambda\beta(t)p_j(t), \quad \text{for all } j \neq i \end{aligned} \quad \text{رابطه (۱)}$$

جاییکه λ پارامتر یادگیری (یا اندازه گام) در بازه $0 < \lambda < 1$ است.

۳ اگر در پایان الگوریتم براساس اینکه کاربر ثانویه توانسته است کانال مناسب را در اختیار بگیرد و داده خود را ارسال کند یا خیر،

یعنی کانال انتخابی آزاد بوده است یا خیر، احتمال p_{Tri} بصورت زیر به روزرسانی می‌شود.

$$p_{Tri}(t+1) = p_{Tri}(t) + \lambda.(1 - p_{Tri}) \quad \text{if the channel is busy} \quad \text{رابطه (۲)}$$

$$p_{Tri}(t+1) = p_{Tri}(t) - \lambda.(p_{Tri} - \alpha) \quad \text{if the channel is free} \quad \text{رابطه (۳)}$$

پایان الگوریتم.

۴-۵ تخصیص طیف با استفاده از Pursuit Estimator

الگوریتم تعقیب از تاریخچه عملیات انتخاب شده و فیدبک‌های بدست آمده برای محاسبه احتمال پاداش‌های عملیات استفاده می‌کند. در این الگوریتم؛ ما از لیست کانال‌ها، کانالی را براساس مقدار توزیع احتمال انتخاب و سپس پاداش آن براساس روش زیر تخمین زده می‌شود. در هر زمان t ، الگوریتم برآوردهای جاری احتمال پاداش‌ها، $\hat{d}_i(t)$ را محاسبه می‌کند. برای محاسبه $\hat{d}_i(t)$ یا پاداش کانال i ام دو بردار اضافه $(Z_1(t) \dots Z_r(t))^T$ و $(\eta_1(t) \dots \eta_r(t))^T$ را نگه می‌دارد. بردار $\eta_i(t)$ مطابق با فرمول (۴) تعداد دفعاتی است که عمل α_i یا کانال ch_i انتخاب می‌شود.

$$\begin{aligned} \eta_i(t) &= \eta_i(t-1) + 1 \\ \eta_j(t) &= \eta_j(t-1), \forall j \neq i \end{aligned}$$

رابطه (۴)

$Z_i(t)$ مطابق با فرمول (۵) فیدبک‌های بدست آمده در پاسخ به عمل α_i یا کانال ch_i در زمان t است.

$$\begin{aligned} Z_i(t) &= Z_i(t-1) + \beta(t) \\ Z_j(t) &= Z_j(t-1), \forall j \neq i \end{aligned} \quad \text{رابطه (۵)}$$

این الگوریتم از بردار $\eta_i(t)$ و $Z_i(t)$ برای برآورد پاداش کانال i یعنی $\hat{d}_i(t)$ استفاده می‌کند.

$$\hat{d}_i(t) = \frac{Z_i(t)}{\eta_i(t)}, \quad i = 1, \dots, r. \quad \text{رابطه (۶)}$$

در نهایت برای عملی که بیشترین پاداش را بدست آورده احتمال را بصورت فرمول (۷) به روزرسانی می‌کنیم.

$$P(t+1) = P(t) + \lambda(e_{M(t)} - P(t)) \quad \text{رابطه (۷)}$$

در این رابطه $0 < \lambda \leq 1$ پارامتر یادگیری و شاخص $M(t)$ براساس رابطه (۸) بدست می‌آید.

$$\hat{d}_{M(k)}(t) = \max_i \hat{d}_i(t) \quad \text{رابطه (۸)}$$

توجه داشته باشید که برای بدست آوردن $\hat{d}_i(t)$ ، لازم است تا $\eta_i(t) > 0$ باشد. برای برقراری این شرط، در شروع مقداردهی اولیه به برآوردها، هر کانال را چند مرتبه انتخاب می‌کنیم. یا می‌توانیم تا زمانی که به نقطه‌ای برسیم که شرط $\eta_i(t) > 0$ درست باشد، $\hat{d}_i(t)$ را برابر صفر نگه داریم. یعنی تا زمانی که یک عمل انتخاب شود. باید توجه کنیم که بروزرسانی احتمال عملیات در الگوریتم تعقیب برخلاف الگوریتم L_R-I ، به شناسایی کانال انتخاب شده بستگی ندارد. در این الگوریتم از کانال انتخاب شده و تقویت بدست آمده برای بروزرسانی برآورد احتمال پاداش‌ها استفاده می‌شود و سپس احتمال کانال‌ها بروزرسانی می‌گردد تا احتمال بهترین کانال تخمین زده شده جاری افزایش یابد. و ویژگی دیگر الگوریتم تعقیب این است که بروزرسانی $P(t)$ مستقیماً شامل $\beta(t)$ (پاسخ محیط) نیست. یکی از مزایای این ویژگی این است که، دیگر لازم نیست $\beta(t)$ (به عنوان یک فاکتور از ابزارهای مهم در کاربردهای مختلف) یک مقدار ثابت در بازه $[0,1]$ باشد. مزیت دیگر الگوریتم تعقیب این است از الگوریتم‌های دیگر مثل L_R-I سریعتر است. در این الگوریتم اگر λ به اندازه کافی کوچک باشد، هر کانال به تعداد کافی انتخاب می‌شود و برآوردها به مقادیر واقعی احتمال پاداش‌ها نزدیک خواهند بود.

بعد از بررسی دو الگوریتم L_{R-I} و Pursuit در مساله تخصیص طیف فرکانسی، روش‌های دیگر یادگیر از جمله L_{R-P} و L_{R-EP} ارائه می‌شوند. تنها تفاوت این دو الگوریتم به روزرسانی تابع احتمال کانال هنگام تخصیص طیف است. یعنی در الگوریتم L_{R-I} کفایت به جای فرمول (۱) از فرمول به روزرسانی احتمال (۹) استفاده کنیم.

$$p_i(t+1) = p_i(t) + \lambda_1 \beta(t)(1 - p_i(t)) - \lambda_2(1 - \beta(t))p_i(t)$$

$$p_j(t+1) = p_j(t) - \lambda_1 \beta(t)p_j(t) + \lambda_2(1 - \beta(t))\left(\frac{1}{r-1} - p_j(t)\right), j \neq i$$

رابطه (۹)

در مقایسه با الگوریتم L_{R-I} ، تفاوت اصلی این است که حتی زمانی که تقویت نامطلوب است (یعنی $\beta(k) = 0$)، احتمال عملیات بروزرسانی می‌شود. (اما در الگوریتم L_{R-I} در این شرایط، احتمالات بدون تغییر باقی می‌مانند). اگر در رابطه بالا $\lambda_2 = 0$ در نظر بگیریم، الگوریتم L_{R-I} بدست می‌آید. اما در الگوریتم L_{R-P} ، $\lambda_1 = \lambda_2$ در نظر گرفتیم.

در الگوریتم L_{R-P} بردار واحد نمی‌تواند یک نقطه پایدار از فرآیند $P(k)$ باشد. در صورتی که در الگوریتم L_{R-I} مشاهده نمودیم که بردار احتمال عمل به یک بردار واحد همگرا است، چون همواره این شانس وجود دارد که احتمال عمل انتخاب شده کاهش یابد. (توجه داشته باشید که به جز در حالات بدیهی، احتمال پاداش هر عمل کمتر از مقدار واحد است). می‌توان نشان داد که در الگوریتم L_{R-P} احتمال عملیات به داخل توزیع همگرا است و اگر اندازه گام یادگیری به اندازه کافی کوچک باشد، مقدار P_i محدود می‌شود، احتمال عمل نام نیز متناسب با $\frac{1}{1-d_i}$ خواهد بود که d_i احتمال پاداش عمل نام است [۲۷].

می‌توان گفت الگوریتم‌ها مبتنی بر ایده ساده "پاداش-جریمه"؛ یعنی هر زمان عمل انتخاب شده منجر به تقویت مطلوب شود، احتمال عمل را افزایش و در غیراینصورت آن را کاهش می‌دهیم. افزایش و کاهش را می‌توان با اندازه گام‌های مختلف انجام داد (با یک حالت افراطی اندازه گام صفر). نامگذاری این الگوریتم‌ها به نام پاداش-جریمه یا پاداش-انفعال مبتنی بر این معیار است. به همین ترتیب می‌توانیم یک الگوریتم انفعال-پاداش خطی را نیز فرموله کنیم.

الگوریتم مشابه دیگر، الگوریتم پاداش-جریمه L_{R-EP} است که با انتخاب در رابطه (۱۰) و با شرط $\epsilon < 1$ به عنوان یک عدد کوچک بدست می‌آید. ویژگی‌های الگوریتم L_{R-EP} مابین الگوریتم L_{R-I} و L_{R-P} است [۲۷].

اگر $\alpha(k) = \alpha_i$ ، در این صورت $P(k)$ به صورت زیر بروزرسانی می‌شود

$$p_i(t+1) = p_i(t) + \lambda_1 \beta(t)(1 - p_i(t)) - \lambda_2(1 - \beta(t))p_i(t)$$

$$p_j(t+1) = p_j(t) - \lambda_1 \beta(t)p_j(t) + \lambda_2(1 - \beta(t))\left(\frac{1}{r-1} - p_j(t)\right), j \neq i$$

رابطه (۱۰)

می‌توان گفت همه این الگوریتم‌ها مبتنی بر ایده ساده "پاداش-جریمه" هستند. یعنی هر زمان عمل انتخاب شده منجر به تقویت مطلوب شود، احتمال عمل را افزایش و در غیراینصورت آن را کاهش می‌دهیم. افزایش و کاهش را می‌توان با اندازه گام‌های مختلف انجام داد (با یک حالت افراطی اندازه گام صفر). نامگذاری این الگوریتم‌ها به نام پاداش-جریمه یا پاداش-انفعال مبتنی بر این معیار است. به همین ترتیب می‌توانیم یک الگوریتم انفعال-پاداش خطی را نیز فرموله کنیم. هر چند الگوریتم ارائه شده برای هر کابر ثانویه به یک کانال مناسب در طول زمان همگرا می‌شود، ولی با توجه به سرعت کند آتوماتای یادگیر و پویایی محیط به نظر می‌رسد که استفاده از روش‌های یادگیر سریع در بهبود کارایی الگوریتم تاثیر بسزایی داشته باشد.

۴ بررسی عملکرد الگوریتم‌های یادگیر

ویژگی‌های همگرایی الگوریتم‌های آتوماتای یادگیر مستقیماً به نوع طرح تقویتی بکار رفته بستگی دارد. تعداد تکرارهای لازم برای همگرایی آتوماتای یادگیر بطور میانگین برای نودها محاسبه شده است. سرعت همگرایی الگوریتم پیگیری Pursuit از همه بیشتر و سرعت همگرایی الگوریتم پاداش-جریمه خطی (L_{R-P}) از همه کمتر است. به عبارتی کاربر ثانویه که از طرح‌های Pursuit یا طرح L_{R-I} استفاده می‌کند پس از تعداد تکرارهای کمتری بهترین کانال را بدست می‌آورد. جدول (۱) نحوه همگرایی و سرعت اجرا الگوریتم‌های ارائه شده با مقدار λ های مختلف را بررسی می‌کند. نتایج نشان می‌دهد که در هر الگوریتم تعداد تکرارها به چه صورت است و بهترین بهره‌وری از کانال در چه مقدار λ برای هر الگوریتم حاصل می‌شود، همانطور که از نتایج مشخص است در λ های بزرگتر سرعت همگرایی بیشتر و تعداد تکرارها کمتر است. ولی در λ کوچکتر سرعت همگرایی خیلی کم و تعداد تکرار خیلی زیاد است، مقدار استفاده بهینه از کانال در هر الگوریتم مشخص شده است. این فاکتور در واقع نشان می‌دهد که با توجه به درخواست کاربران ثانویه و تعداد کانال‌های مجوز دار موجود، چند درصد از پهنای باند توسط کاربران ثانویه استفاده شده و چه درصدی به هدر رفته است.

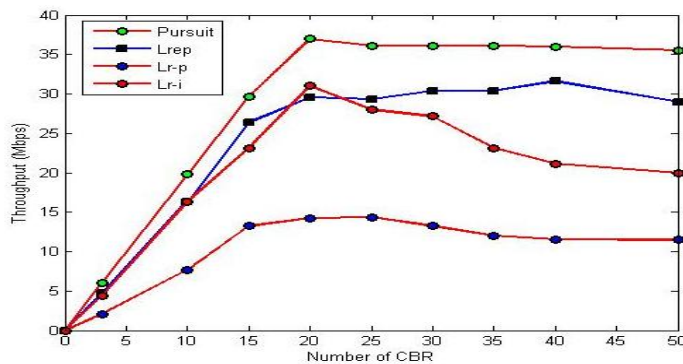
(جدول ۱) نحوه همگرایی و سرعت اجرا الگوریتم‌های ارائه شده با مقدار λ های مختلف

λ		.02	.03	.05	.1	.2	.3	.4
L_{R-P}	Iteration	500	65	00	56	30	20	3
	Channel Utilization	.41	.56	.78	.80	0%	.6	.3
L_{ReP}	Iteration	000	25	20	30	50	6	2
	Channel Utilization	.55	.65	.70	.75	1%	.65	.5
L_{R-I}	Iteration	56	00	30	00	5	0	5
	Channel Utilization	.35	.4	.55	3%	.6	.2	.2
Pursuit Algorithm	Iteration	00	5	6	5	0	0	
	Channel Utilization	.65	.69	9%	.69	.5	.5	.45

۵ بررسی پارامترهای بهبود استفاده از کانال

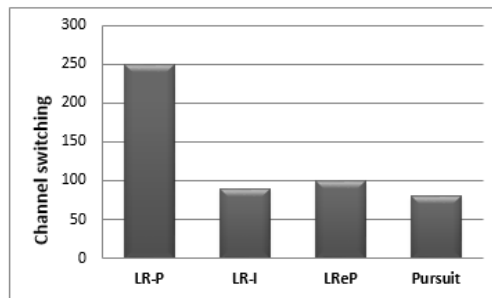
در آزمایش‌های انجام شده، بهره‌وری از طرح تخصیص کانال بالا ذکر شده است و از نظر معیارهای زیر سنجیده می‌شود.

- **نمودار Throughput**: این فاکتور نشان دهنده توان عملیاتی شبکه است که در واقع نسبت تعداد متوسط بسته دریافت شده به نسبت تعداد کل بسته‌هایی است که ارسال شده‌اند. این متریک با استفاده از الگوریتم پیشنهادی که در آن بخش پهنای باند اختصاص داده شده به هر یک از کاربران ثانویه متناسب با تعداد درخواست آنهاست، بهینه شده است.



شکل ۵-۱) نمودار Throughput

- **سوئیچینگ بین کانال‌ها (Channel switching)** هزینه سوئیچینگ در روش پیشنهادی، که از الگوریتم Pursuit استفاده می‌کند، کمتر از سایر الگوریتم‌های یادگیر است. در روش پیشنهادی با توجه به اولویت دادن به کاربران ثانویه که درخواستهای آنها در تکرارهای اول بوده و همچنین با توجه به استفاده از الگوریتم پیگیر که سرعت همگرایی بالایی دارد و براساس تخمین پاداش احتمال اعمال را بروزسانی می‌کند در نتیجه سابقه کانال انتخابی را نیز در نظر می‌گیرد در نتیجه باعث می‌شود که روش پیشنهادی کانال‌های بهتری را انتخاب کرده و در تکرارهای بعدی تعداد سوئیچینگ کانال کمتر شود. هزینه سوئیچینگ در L_{ReP} بیشتر از L_{R-I} و Pursuit است چون برخلاف این دو، دارای شناسی برای انتخاب شدن در کانال‌های دیگر دارد. L_{R-P} در مقایسه با L_{R-I} و L_{ReP} دارای رفتار متفاوتی است. با توجه به اینکه در L_{R-P} ، $\alpha = \beta$ ، زمانی که پاسخ پاداش یا جریمه وجود دارد، احتمال عملیات را با نرخ‌های یکسانی بروزسانی می‌کند. در نتیجه در الگوریتم L_{R-P} ، احتمالاً انتخاب کانالی که بهترین نباشد بیشتر از دیگر روش‌ها است. این امر موجب سوئیچینگ کانال بیشتر و همچنین افزایش شناسی انتخاب کانال‌های کمتر در دسترس می‌شود.



(شکل ۵-۲) تعداد سوئیچینگ کانال، در محیط ثابت با ۳۰۰۰ اسلات زمانی

۶ نتیجه‌گیری

رادپوشناختی مبتنی بر اتوماتای یادگیر، یک تکنولوژی بسیار بالقوه برای مقابله با چالش کمبود طیف در شبکه‌های رادیو شناختی می‌باشد. با بکارگیری روش پیشنهادی توانستیم بطور قابل توجهی استفاده مفید از طیف فرکانسی را افزایش دهیم و همچنین منجر به بهبود بعضی از پارامترهای کیفیت سرویس شویم. ما از اتوماتای یادگیری در سناریوی کاربر ثانویه SU و کاربر اولیه PU برای دسترسی طیف پویا در شبکه‌های رادپوشناختی استفاده کردیم و نشان دادیم که الگوریتم دسترسی طیف پویای مبتنی بر اتوماتای یادگیری پاسخ چندگانه می‌تواند رقابت بین کاربران ثانویه و همچنین استفاده از کانال‌های در دسترس را کنترل کند. این امر منجر به کاهش نرخ برخوردها و افزایش توان گذردهی کلی می‌شود. هزینه سوئیچینگ و همگرایی همه الگوریتم‌ها را بررسی و با طرح‌های مطرح شده، مقایسه و ارزیابی می‌شود. همچنین نشان دادیم که الگوریتم تعقیب Pursuit Estimator از تاریخچه عملیات انتخاب شده و فیدبک‌های بدست آمده برای محاسبه احتمال پاداش‌های عملیات استفاده می‌کند. در این الگوریتم؛ ما از لیست کانال‌ها، کانالی را براساس مقدار توزیع احتمالش انتخاب و سپس پاداش آنها تخمین زده خواهد شد. که بروزسانی احتمال عملیات در الگوریتم تعقیب برخلاف الگوریتم LR-I، به شناسایی کانال انتخاب شده بستگی ندارد. با بکارگیری از شبکه رادپوشناختی مبتنی بر اتوماتای یادگیر، با مشکل کمبود طیف در شبکه‌ها مقابله کردیم.

منابع و مراجع

- [1] I.F. Akyildiz, W.-Y. Lee, K.R. Chowdhury, Cognitive radio ad hoc networks: research challenges, to appear in *Ad Hoc Networks Journal*, Elsevier, 2009.
- [2] H. Kim, K.G. Shin, Adaptive MAC-layer sensing of spectrum availability in cognitive radio networks, Technical Report, CSE-TR-518-06, University of Michigan, 2006.
- [3] H. Kim, K.G. Shin, Efficient discovery of spectrum opportunities with MAC-layer sensing in cognitive radio networks, *IEEE Trans. Mobile Comp.* 7 (2008) 533–545. May.
- [4] W.Y. Lee, I.F. Akyildiz, Optimal spectrum sensing framework for cognitive radio networks, *IEEE Trans. Wireless Commun.* 7 (10) (2008) 845–3857.
- [5] P. Wang, L. Xiao, S. Zhou, J. Wang, Optimization of detection time for channel efficiency in cognitive radio systems, in: *Proceedings of Wireless Communications and Networking Conference*, March 2008, pp. 111–115.
- [6] A. Ghasemi, E.S. Sousa, Optimization of spectrum sensing for opportunistic spectrum access in cognitive radio networks, in: *Proceedings of IEEE Consumer Communications and Networking Conference*, January 2007, pp. 1022–1026.
- [7] Y. Pei, A.T. Hoang, Y.-C. Liang, Sensing throughput tradeoff in cognitive radio networks: how frequently should spectrum sensing be carried out? in: *Proceedings of the 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, September 2007, pp. 5330–5335.
- [8] W.Y. Lee, I.F. Akyildiz, Optimal spectrum sensing framework for cognitive radio networks, *IEEE Trans. Wireless Commun.* 7 (10) (2008) 845–3857.
- [9] L. Luo, S. Roy, Analysis of search schemes in cognitive radio, in: *Proceedings of IEEE Workshop on Networking Technologies for Software Defined Radio Networks*, June 2007, pp. 647–654.
- [10] S. Srinivasa, S.A. Jafer, Soft sensing and optimal power control for cognitive radio, in: *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM)*, November 2007, pp. 1380–1384.
- [11] S.-Y. Lien, C.-C. Tseng, K.-C. Chen, Carrier sensing based multiple access protocols for cognitive radio networks, in: *Proceedings of IEEE International Conference on Communications (ICC)*, May 2008, pp.3208–3214.
- [12] IEEE 802.22 Working Group on Wireless Regional Area Networks, <<http://www.ieee802.org/22/>>.
- [13] C. Zhou, C. Chigan, A game theoretic DSA-driven MAC framework for cognitive radio networks, in: *Proceedings of IEEE International Conference on Communications (ICC)*, May 2008, pp. 4165–4169.
- [14] P. Pawelczak, R. Venkatesha Prasad, Liang Xia, Ignas G.M.M. Niemegeers, Cognitive radio emergency networks – requirements and design, in: *Proceedings of IEEE DySPAN*, November 2005, pp.601–606.
- [15] L. Ma, X. Han, C.-C. Shen, Dynamic open spectrum sharing for wireless ad hoc networks, in: *Proceedings of IEEE DySPAN*, November 2005, pp. 203–213.
- [16] H. Su, X. Zhang, CREAM-MAC: An efficient cognitive radio-enabled multi-channel MAC protocol for wireless networks, in: *Proceedings of IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, June 2008, pp. 1–8.
- [17] L. Ma, C.-C. Shen, B. Ryu, Single-radio adaptive channel algorithm for spectrum agile wireless ad hoc networks, in: *Proceedings of IEEE DySPAN*, April 2007, pp. 547–558.
- [18] C. Cordeiro, K. Challapali, C-MAC: A cognitive MAC protocol for multichannel wireless networks, in: *Proceedings of IEEE DySPAN*, April 2007, pp. 147–157.
- [19] J. Zhao, H. Zheng, G.-H. Yang, Spectrum sharing through distributed coordination in dynamic spectrum access networks, *Wireless Commun. Mobile Comp.* 7 (9) (2007) 1061–1075.
- [20] Q. Zhao, L. Tong, A. Swami, Y. Chen, Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework, *IEEE J. Selected Areas Commun.* 25 (3) (2007) 589–600.

- [21] S. Geirhofer, L. Tong, B.M. Sadler, Cognitive medium access: constraining interference based on experimental models, *IEEE J. Selected Areas Commun.* 26 (1) (2008) 95–105.
- [22] Y.R. Kondareddy, P. Agrawal, Synchronized MAC protocol for multihop cognitive radio networks, in: *Proceedings of IEEE International Conference on Communications (ICC)*, May 2008, pp. 3198–3202.
- [23] H. Su, X. Zhang, Opportunistic MAC protocols for cognitive radio based wireless networks, in: *Proceedings of Annual Conference on Information Sciences and Systems*, March 2007, pp. 363–368.
- [24] J. Ansari, X. Zhang, P. Mõhõnen, *Computer Communications* 36 (2013) 1399–1410, : A decentralized MAC protocol for opportunistic spectrum access in cognitive wireless networks, Accepted 1 June 2013, pp. 1399-1410
- [25] M.A.L. Thathachar and P.S. Sastry. *Networks of Learning Automata Techniques for Online Stochastic Optimization*. Kluwer Academic Publishers, 2004.
- [26] Ya. Z. Tsypkin. *Adaptation and learning in automatic systems*. Academic Press, New York, 1971.
- [27] K. S. Narendra and M. A. L. Thathachar, *Learning Automata: An Introduction*. Prentice Hall, Englewood Cliffs. 1989.
- [28] B. J. Oommen and M. Agache. Continuous and discretized pursuit learning schemes: Various algorithms and their comparison. *IEEE Transactions on Systems, Man and Cybernetics: Part B*, 31:277-287,2001.
- [29] G. I. Papadimitriou, M. Sklira, and A. S. Pomportsis . A new class of foptimal learning automata. *IEEE Transactions on Systems. Man and Cybernetics: Part B*, 33, 2003. to appear.
- [30] J.K Lanctot and B.J. Oomrnen, Discretized estimator learning automata . *IEEE Transactions on Systems. Man and Cybernetics*, 22(6):1473-1483 . 1992.
- [31] V. Vapnik. *Nature of Statistical Learning Theory*. Springer-Verlag, New York,1997.
- [32] Hannaneh Bizhani, Abdorasoul Ghasemi, Joint Admission Control and Channel Selection Based on Multi Response Learning Automata (MRLA) in Cognitive Radio Networks. *Wireless Personal Communications*. July 2013, Volume 71, Issue 1, pp 629–649