Downloaded from jor.iranaict.ir at 9:07 +0330 on Wednesday November 25th 2026

فصلنامه علمي-پژوهشي
فنّاوری اطلاعات و ارتباطات ایران

سال اول، شماره‌هاي ۱ و ۲، پاییز و زمستان ۱۳۸۷
صص: ۱۹- ۳۰

یک روش چند عاملی جدید مبتنی بر یادگیری تقویتی برای شکل‌دهی ترافیک و تخصیص حافظه بافر در روترها

محمد طاهری تهرانی* **سید علی اکبر صفوی*[1]** **محمد رفیع خوارزمی****

*دانشکده مهندسی برق و کامپیوتر، دانشگاه شیراز

**دانشکده مهندسی برق، دانشگاه صنعتی شیراز

چکیده

دراین مقاله با توجه به ساختار توزیع شده شبکه‌های کامپیوتری و رفتار تصادفی موجود در آنها و از طرف دیگر محدودیت‌های زمانی که در الگوریتم‌های کنترلی برای اینگونه سیستمها وجود دارد، از مفاهیم سیستم‌های چند عاملی و تکنیک‌های یادگیری تقویتی برای شکل دهی ترافیک در روترها و تخصیص دینامیک حافظه بافر بین پورت های مختلف یک روتر استفاده شده است. در واقع با استفاده از این مفاهیم شکل دهنده ترافیک جدیدی بر مبنای یک الگوریتم سطل نشانه‌دار توسعه داده شده است که در آن به جای آنکه نرخ تولید نشانه‌ها به طور استاتیک تخصیص داده شود به طور دینامیک و هوشمند و بر مبنای وضعیت شبکه مشخص می‌شود. این پیاده‌سازی علاوه بر آنکه به استفاده بهینه و منطقی از پهنای باند منجر می‌شود باعث می‌گردد ترافیک در دیگر نقاط شبکه نیز کاهش یابد. همچنین از این مفاهیم استفاده می‌شود تا یک روش جدید برای تخصیص هوشمند و دینامیک حافظه بافر برای پورت‌های یک روتر توسعه داده شود. این پیاده‌سازی نیز باعث می‌گردد تا نرخ افت بسته‌ها در کل شبکه به ویژه در شرایطی که بار شبکه افزایش می‌یابد کاهش داده شود. نتایج شبیه‌سازی‌های انجام شده کار آمدی روش پیشنهادی را نشان می‌دهد.

کلید واژگان: شبکهٔ کامپیوتری، سیستم چند عاملی، حافظهٔ بافر، روتر، شکل‌دهی ترافیک

۱- مقدمه

امروزه با پیشرفت در شبکه‌های انتقال داده پرسرعت، کاربردهای زبان واقعی[1] و فعال مانند دورکنفرانس[2] و یا محاسبات توزیع شده[3] مطرح شده‌اند. با توجه به غیرهمسان بودن منابع این کاربردها (ویدئو، صدا یا داده)، کیفیت سرویس (QoS)[4] موجود در شبکه‌های انتقال نمی‌توانند به‌طور کافی انتقال این داده‌ها را به‌صورت پیوسته و مناسب تضمین کنند. بنابراین باید به نوعی به سمت روشهایی رفت که نیازهای (QoS) را به بهترین نحو ممکن برآورده کنند [۱و۱۴].

ازطرفی اکثر ابزارها و قالبهای مدیریتی که امروزه در شبکه‌های کامپیوتری استفاده می‌شود براساس روشهای متمرکز[5] بنا شده‌اند و از فنّاوریهایی که در بستر شبکه وجود دارد و می‌توان از آنها در راستای مدیریت شبکه استفاده کرد، چندان بهره‌ای نمی‌برند. به‌عنوان مثال، غالباً اپراتورها در سطوح مختلف سازمانی، به مدیریت شبکه‌ها بر اساس برنامه‌های ساده و غیر انعطاف‌پذیر می‌پردازند و لذا درتوسعه ابعاد شبکه، با مشکل توسعه‌پذیری[6] سیستم خود روبرو هستند [۲و۳].

الگوریتم‌های کنترل شبکه را می‌توان به دو دسته تقسیم کرد، یکی الگوریتم‌های کنترل جریان[7] و دیگر الگوریتم‌های

۱. نویسنده عهده‌دار مکاتبات (safavi@shirazu.ac.ir)

۱.Real Time
۲.Teleconference
۳.Distributed Computation
٤.Quality of Service
٥.Centralized
٦.Scalability
۷.Flow Control

Downloaded from jor.iranaict.ir at 9:07 +0330 on Wednesday November 25th 2026

کنترل ازدحام[۱]. در الگوریتم‌های کنترل جریان، شبکه به دنبال انجام کارهایی است که در آن از به وجود آمدن ازدحام جلوگیری شود در صورتی که در کنترل ازدحام، شبکه به دنبال آن است که زمان، گسترش و چگالی ازدحام به‌وجود آمده را حداقل کند. روشهای کنترل جریان به‌نوعی پیشگیرانه[۲] و الگوریتم‌های کنترل ازدحام واکنشی[۳] هستند[۳]. به‌عنوان مثال، روش شکل‌دهی ترافیک[۴] یکی از الگوریتم‌های کنترل ازدحام است که در کامپیوترها و روترها می‌تواند استفاده شود و به‌نوعی نرخ ارسال داده‌ها را بویژه در مواقعی که نرخ تولید این داده‌ها نامنظم و ضربه گونه[۵] است، منظم کند. این موضوع ممکن است ناشی از این نکته باشد که استفاده‌کنندگان این منابع، به‌طور سهوی یا به هر دلیل دیگر، سعی در افزایش نرخ ارسال مورد توافق درهنگام برقراری ارتباط داشته باشند.

امروزه با توجه به ابعاد زیاد داده‌ها و طبیعت ناهمسان آنها و محدودیت‌های زمانی مرتبط با کارهای کنترلی، بسیاری از جنبه‌های کنترل شبکه‌های مخابراتی را می‌توان برپایه روشهای هوش مصنوعی بنا نهاد و با توجه به اینکه شبکه ساختاری توزیع شده دارد، روشهای هوش مصنوعی توزیع شده (DAI)[۶] می‌توانند نامزد مناسبی باشند [۳و۴] . یکی از زیر شاخه‌های (DAI)، سیستم‌های چند عاملی[۷] هستند که با استفاده از مجموعه‌ای از واحدهای مستقل به نام عامل، می‌توانند رفتار یک سیستم را تحت کنترل درآورند.

از طرف دیگر برای اینکه سیستمهای چند عاملی بتوانند در تقابل با محیطی که درآن قرارگرفته‌اند، رفتار مناسب و بهینه‌ای نشان دهند، احتیاج به یادگیری دارند. در واقع یکی از خصوصیات اصلی سیستمهای هوشمند، نیاز به یادگیری است و سیستمهای چند عاملی هم از این قانون مستثنا نیستند [۵ و۶]. یکی ازاین روشهای یادگیری، یاد گیری تقویتی[۸] است. یادگیری تقویتی به معنای آن است که یک عامل بتواند با سعی وخطا، در تقابل با محیطی که در آن قرار گرفته است به یادگیری رفتار مناسب بپردازد. یادگیری تقویتی یکی از روشهای برنامه‌ریزی عاملهاست که در آن یک عامل بدون شناخت اولیه از محیط و با استفاده از تقابلی که با محیط دارد و پاداش‌هایی که از محیط دریافت می‌کند، به مرور زمان، رفتار بهینه خود را می‌آموزد. یکی از روشهایی که اصولاً برای حل مسائل مربوط به یادگیری تقویتی وجود دارد، استفاده از روشهای آماری وبرنامه‌ریزی پویا مانند روشهای بلمن[۱] برای تخمین بهترین عمل[۲]ها در هرحالتی از محیط است [۵-۷].

در این مقاله با ارائه راهکارهای مبتنی بر یادگیری تقویتی و مفاهیم چند عاملی، سعی می‌شود تا در حد امکان علاوه بر استفاده بهینه از پهنای باند موجود، ترافیک و ازدحام در دیگر نقاط شبکه کاهش داده شود. در این راستا دو قالب چند عاملی ارائه می‌شود. یک قالب شکل دهنده ترافیک که ترافیک آمده به سمت روتر را به طور هوشمند و بر اساس وضعیت شبکه در جلو شکل دهی می‌کند و یک قالب تخصیص دهنده حافظه که نیز بر اساس شرایط شبکه به اشتراک و اختصاص حافظه بافر به هر پورت روتر می‌پردازد. برای عاملهای شکل دهنده ترافیک دو حالت تعریف شده است: یکی درصد افت بسته[۳] های اطلاعاتی که روتر جلوی خود می‌بیند و دیگری درصد بافر استفاده شده در پورت خروجی. عامل درهر لحظه با پسخور گرفتن از این دو حالت به عنوان ورودی، به‌صورت هوشمند و با استفاده از الگوریتم یادگیری تقویتی عمل مناسب را به عنوان خروجی به شبکه اعمال و در مقابل یک پاداش دریافت می‌کند. این پاداش بر این اساس تعریف شده که شبکه با گرفتن این عمل به چه حالت دیگری رفته است. در واقع این پاداش میزان مفید بودن عمل را در راستای نزدیک کردن شبکه به یک حالت مناسب نشان می‌دهد. طبیعی است که حالت مناسب طبق این تعاریف، کاهش درصد افت بسته‌ها و همچنین کاهش درصد بافر استفاده شده در پورت است. به این دلیل که هرچه این بافر خالی‌تر باشد، میزان افت برای بسته‌های ورودی کاهش می‌یابد.

علاوه بر آن که این عاملها بر روی تمام پورتهای خروجی یک روتر در قالب چند عاملی در نظر گرفته می‌شود و به کمک آنها شکل دهی هوشمند ترافیک در جلوی روتر و بر اساس وضعیت شبکه انجام می‌شود، برای هر پورت عامل دیگری هم در نظر گرفته می‌شود تا با توجه به وضعیت شبکه در پشت روتر و در قالب یک سیستم چند عاملی دیگرکار تخصیص بافر[۴] را برای هر پورت انجام دهد. اضافه‌کردن این قالب باعث کاهش درصد افت بسته‌هایی می‌شود که به سمت پورتهای شلوغ می‌روند. برای این عاملهای تخصیص دهنده حافظه نیز دو حالت

۱.Congestion Control
۲.Preventive
۳.Reactive
٤.Traffic Shaping
٥.Burst
٦.Distributed Artificial Intelligence
۷.Multi Agent
۸.Reinforcement Learning

۱.Bellman Methods
۲.Action
۳.Packet Drop
٤.Buffer Allocation

Downloaded from jor.iranaict.ir at 9:07 +0330 on Wednesday November 25th 2026

تعریف می‌شود: یکی میزان افت بسته‌ها در ورودی هر روتر و دیگری میزان پر یا خالی بودن بافر مربوط به هر پورت.

ساختار این مقاله به صورت زیر است: در بخش بعدی به اجمال به روشهای شکل‌دهی ترافیک، به ویژه الگوریتم سطل نشانه‌دار، اشاره می‌شود. پس از آن در بخش سوم به مروری درمورد یادگیری تقویتی پرداخته می‌شود. از این ایده در بخش چهارم برای طراحی و پیاده‌سازی راهکارهای پیشنهادی درساختار قالبهای چند عاملی استفاده می‌شود. در بخش پنجم به شبیه‌سازی و ارزیابی نتایج مدلهای توسعه داده شده پرداخته می‌شود و در نهایت در بخش آخر نتایج و پیشنهادهایی برای کارهای آینده آورده می‌شود.

۲- شکل‌دهی ترافیک

شکل دهی ترافیک (یا شکل‌دهی بسته‌ها) فرایندی است به منظور کنترل ترافیک شبکه‌های کامپیوتری که طی آن واحد شکل دهنده با تأخیر دادن به بسته‌ها و کنترل بر روی زمان و میزان ارسال بسته‌ها سعی می‌کند کارایی شبکه و همچنین پهنای باند شبکه را به طور بهینه استفاده کند و یا اهداف خاص مدیریتی را در پیاده‌سازی QoS تضمین نماید. به عبارت دیگر شکل دهی ترافیک به هر عملی گفته می‌شود که بر روی یک جریان از بسته‌ها اعمال شده و سعی می‌کند با تحمیل کردن یک زمان تأخیر اضافی بر روی آن بسته‌ها، آنها را در یک پروفایل[۱] از قبل تعریف شده ترافیکی قرار دهد. شکل‌دهی ترافیک وسیله ای است برای کنترل حجم ترافیکی که در یک بازه زمانی به داخل شبکه تزریق می‌شود و یا کنترل حداکثر نرخی که ترافیک به داخل شبکه فرستاده می‌شود. شکل‌دهی ترافیک اصولاً در لبه‌های شبکه مثل روترهای منبع[۲] (SR) برای کنترل ترافیک تزریقی به شبکه به کار برده می‌شود ولی با این حال می‌توان آن را در منابع ترافیکی مثل خود کامپیوترها یا در داخل شبکه مثل روترهای شبکه[۳] (NR) مورد استفاده قرار داد[۸،۹ و ۱۰].

دو روش اصلی که عمدتاً برای پیاده‌سازی شکل‌دهی ترافیک وجود دارد الگوریتم سطل سوراخ دار(LB)[۴] و الگوریتم الگوریتم سطل نشانه دار (TB)[۵] است. الگوریتم LB یک محدودیت سخت بر روی نرخ ارسال داده‌ها می‌گذارد و آنها را مجبور می‌کند که با یک نرخ ثابت ارسال شوند در حالی که الگوریتم TB اجازه می‌دهد ارسال داده‌ها به صورت متغیر و بعضاً همراه ضربه‌های قابل قبول باشد و فقط بر روی میانگین داده‌های ارسالی محدودیت می‌گذارد. این الگوریتم که در این تحقیق هم مورد توجه بوده است درواقع یک مکانیزم کنترلی است که بر اساس تعداد نشانه موجود در سطل، مشخص می‌کند چه موقع ترافیک آمده به سمت جلو ارسال شود. واحد شمارش نشانه بایت یا تعداد بسته است و این واحد بدین معناست که هر نشانه واحد می‌تواند به چه تعداد بایت داده و یا بسته اطلاعاتی اجازه ارسال دهد. مقدار دقیق این تعداد توسط مدیر شبکه مشخص می‌شود. با این تفاسیر در صورتی جریان ترافیکی اجازه ارسال خواهد یافت که به تعداد کافی نشانه برای ارسال آن وجود داشته باشد.

۳- یادگیری تقویتی

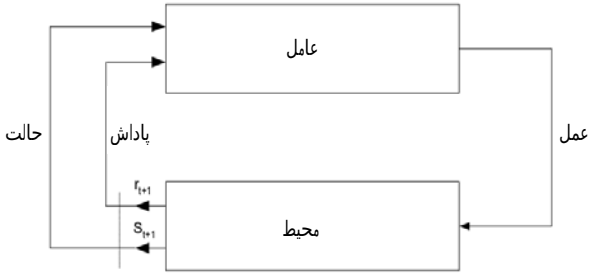
درسال‌های اخیر یادگیری تقویتی به عنوان یک روش یادگیری که در آن احتیاجی به استفاده از مدل محیط نیست و می‌توان به طور زمان واقعی از آن استفاده کرد مورد توجه قرار گرفته و بر روی آن مطالعات گسترده انجام شده است. از طرف دیگر همان طور که قبلاً ذکر شد برای اینکه سیستمهای چند عاملی بتوانند در تقابل با محیطی که در آن قرار گرفته‌اند، رفتار مناسب و بهینه‌ای نشان دهند، احتیاج به یادگیری دارند و از آنجا که هر عامل اطلاعات کمی از وضعیت دیگر عاملها دارد و توجه به این موضوع که محیط دائماً در حال تغییر است، یادگیری تقویتی می‌تواند به عنوان یک روش یادگیری مناسب در کنار آن استفاده شود.

یادگیری تقویتی به فرایندی گفته می‌شود که در طی آن یک عامل می‌آموزد چه اعمالی باید انجام دهد تا پاداش بیشتری از محیط دریافت کند. نکته ای که در این یادگیری قابل توجه است آن است که انجام یک عمل توسط عامل نه تنها شامل یک پاداش آنی می‌شود بلکه بر موقعیت های بعدی و همچنین مجموع پاداشهای به دست آمده طی فرایند نیز تأثیرگذار است.

درشکل۱ مدلی از فرایند یادگیری تقویتی ارائه شده است. در این مدل در لحظه t عامل یک عمل a_t انجام می‌دهد در مقابل این عمل، محیط عامل را از حالت S_t به حالت S_{t+1} می‌برد و یک پاداش لحظه ای r_t به آن می‌دهد. در واقع این مقدار پاداش به نوعی عامل را در راستای پیدا کردن سیاست بهینه تقویت می‌کند و به همین دلیل است که به این روش یادگیری تقویتی گفته می‌شود.

۱.Profile
۲.Source Routers
۳.Network Routers
۴.Leaky Bucket
۵.Token Bucket

Downloaded from jor.iranaict.ir at 9:07 +0330 on Wednesday November 25th 2026



شکل۱: مدل ساختار یادگیری تقویتی

برای یافتن سیاست بهینه و برای اینکه مشخص شود یک عامل چگونه آینده را در تصمیم‌گیری‌های فعلی خود در نظر می‌گیرد و چگونه در لحظهٔ حال عمل می‌کند باید یک مدل رفتار بهینه تعریف شود. مدل رفتار بهینه‌ای که بیشترین کاربرد را در مدلسازی‌ها وحل مسایل داشته است مدل کاهش یافته با افق نامحدود است. در این مدل عامل به جای توجه کردن به h مرحله به یک بازهٔ نامحدود فکر می‌کند و به پاداشهایی که در بلندمدت به دست می‌آورد توجه می‌کند. در این مدل، عامل پاداشهایی را که در آینده به دست می آورد مطابق رابطهٔ۱ با یک نرخ کاهش یافتهٔ هندسی با ضریب با γ دریافت می‌کند:

$$E(\sum_{t=0}^{\infty}\gamma^t r_t)$$
$$0 \le \gamma \le 1 \qquad (۱)$$

محبوبترین روش درحل مسائل یادگیری تقویتی روش تفاوت موقتی است. این روش علاوه بر آنکه نیازی به مدل محیط یادگیری ندارد قابلیت بالایی در محاسبات مرحله به مرحله افزایشی دارد. در این روش عامل سعی می‌کند به تقریب تابع مقدار بر اساس پاداش لحظه‌ای و مقدار تخمین زده شده برای حالت بعدی که در آن قرار خواهد گرفت بپردازد و در این راستا از روش تکرار مقدار نیز کمک می‌گیرد.

ساده‌ترین و پرکاربردترین روش یادگیری تفاوت موقتی روش یادگیری Q است. این روش تجربه بدست آمده از هر گذر حالت در محیط را استفاده کرده و مقادیر مربوط به هر حالت را در یک جدول به نام جدول Q به روز می‌کند. این جدول برای هر زوج حالت و عمل (a و S) یک مقدار Q(S,a) دارد و به ازای هر گذر حالت از S_t به S_{t+1} و دریافت پاداش r_{t+1} مقادیر Q را در جدول طبق رابطه۲ به روز می‌کند:

$$Q(s_t,a_t) := Q(s_t,a_t) + \alpha\left(r_{t+1} + \gamma \max_a Q(s_{t+1},a_t) - Q(s_t,a_t)\right) \qquad (۲)$$

اگر دقت شود در این الگوریتم ،عامل موقعی که عمل a را در حالت s انتخاب می‌کند و به حالت s' می‌رود برای به روز کردن مقدار Q خود در آن لحظه یعنی $Q(s,a)$ از ماکزیمم مقدار Q حالت بعدی یعنی $\max Q(s',a')$ نیز استفاده می‌کند. در این الگوریتم اگر هر عمل a در هر حالت، در بی نهایت اجرای برنامه بی نهایت بار انتخاب شود و پارامتر α به خوبی تنظیم شود مقادیر Q با احتمال صدر صد به Q* همگرا خواهند شد .

۴- مدل سیستم ارائه شده

همچنان که ذکر شد در این مقاله با ارائه راهکارهای مبتنی بر یادگیری تقویتی و مفهوم چند عاملی، سعی می‌شود تا در حد امکان علاوه بر استفاده بهینه از پهنای باند موجود، ترافیک و ازدحام در دیگر نقاط شبکه کاهش داده شود. در این راستا دو قالب چند عاملی ارائه می‌شود. یک قالب شکل دهنده ترافیک که ترافیک آمده به سمت روتر را به طور هوشمند و بر اساس وضعیت شبکه در جلو شکل دهی می‌کند و یک قالب تخصیص دهنده حافظه که بر اساس شرایط شبکه به اشتراک و اختصاص حافظه بافر به هر پورت روتر می‌پردازد. در ادامه ابتدا به مدل سازی قالب شکل‌دهنده ترافیک پرداخته می‌شود و در نهایت مدل قالب تخصیص دهنده حافظه ارائه می‌گردد.

درساختارهای ارائه شده در این مقاله، روترها در دو دسته اصلی روترهای شبکه[۱] (NR) و روترهای منبع[۲] (SR) طبقه‌بندی می‌شوند. روترهای منبع آن دسته از روترها هستند که در سمت کاربران قرار دارند و مستقیماً با آنها در ارتباط هستند و روترهای شبکه آن دسته از روترها هستند که کاربران مستقیماً با آنها سروکار ندارند و به عنوان متصل کننده بین چند زیر شبکه عمل می‌کنند. هر SR با یک سطل نشان دار (TB) به طول l بیت و نرخ تولید نشانه g بیت بر ثانیه مشخص می‌شود. در این حالت روتر در صورتی می‌تواند یک بسته اطلاعاتی کامل را به داخل شبکه تزریق کند که به اندازهٔ کافی نشان برای ارسال آن وجود داشته باشد و در مقابل موقعی که تعداد نشانه‌ها بیشتر از حد TB شود، از بین می‌روند. درصورتی

۱.Network Router
۲.Source Router

Downloaded from jor.iranaict.ir at 9:07 +0330 on Wednesday November 25th 2026

که یک موج ضربه‌ای از بسته‌های اطلاعاتی به روتر برسند و تعداد بیت های آن بیشتر از تعداد نشانه‌های موجود در سطل باشد، تنها آن تعداد از بسته‌هایی که تعداد بیت آنها کمتر یا برابر نشانه‌ها باشند اجازه ارسال خواهند یافت و بقیه تا زمانی که نشان برای آنها تولید شود، منتظر خواهند ماند.

روترهای شبکه (NR) هم مسؤول تولید اطلاعات لازم در مورد وضعیت شبکه برای SR ها هستند که آنها با توجه به این اطلاعات بتوانند پارامترهای ذکر شده را به نحوی تنظیم کنند که احتمال حذف بسته‌ها را در یک شبکه حداقل کنند. اصولاً روشهایی که در عمل برای تنظیم این پارامترها به ویژه نرخ تولید نشانه g استفاده می‌شود، روشهای با نرخ ثابت هستند. به عنوان مثال از یک الگوریتم سطل سوراخ (LB) قبل از آن استفاده می‌شود که با یک نرخ ثابت و بدون توجه به بقیه شرایط به تولید نشانه می‌پردازد. این نرخ ثابت معمولاً توسط مشخصه‌های ترافیکی و یا رزرو منابع تعیین می‌شود.

در این مقاله به دنبال یک مکانیزم انعطاف پذیر در SR برای انتخاب پارامترهای الگوریتم سطل نشانه‌دار بویژه پارامتر g هستیم تا بتواند در هر لحظه با توجه به وضعیت شبکه در جلوی خود که در این حالت نرخ حذف بسته‌ها در NR است و همچنین وضعیت پر یا خالی بودن بافرهای روتر، بسته‌های اطلاعاتی را به سمت جلو ارسال کند. بدین منظور یک سیستم هوشمند مبتنی بر یادگیری تقویتی طراحی می‌شود که بتواند در هر حالت شبکه بهترین g ممکن را که در واقع همان نرخ ارسال داده‌ها به سمت جلو است برای سطل نشان دار بیاموزد. در این طرح، عمل[1] مربوط به عامل یعنی a یک عدد بین ۰ تا ۱۰۰ است و با رابطه۳ مقدار g را مشخص می‌کند:

$$g = a \times g_{\max} \qquad (۳)$$

که در این رابطه $g_{\max}$ با رابطه زیر مشخص می‌شود:

$$g_{\max\,,\,j} = w_j \qquad (۴)$$

که w_j به‌عنوان پهنای باند لینک ارتباطی j امین پورت خروجی روتر که در واقع همان پورت مورد نظر ما است در نظر گرفته می‌شود. به عبارت دیگر عمل a در هر لحظه میزان درصد پهنای باندی را که عامل برای ارسال بسته‌های اطلاعاتی به سمت مقابل می‌تواند استفاده کند، مشخص می‌کند.

برای هرعامل شکل دهنده ترافیک، حالت شبکه با دو پارامتر مشخص می‌شود، یکی درصد حذف بسته‌ها در مقابل SR که توسط پورت مورد نظر و یا به عبارتی i امین پورت SR در لحظه t دیده می‌شود ($p_{t,i}$) و دیگری درصد پر بودن بافر خروجی پورت مورد نظر در لحظه t که به m امین NR متصل است ($b_{t,im}$).

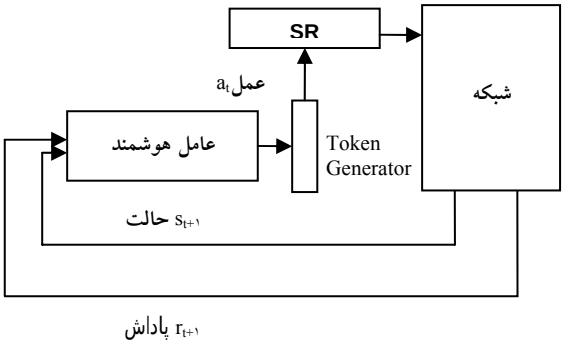
پاداشی که این عامل در مقابل انجام عمل a در لحظه t دریافت می‌کند یعنی r_{t+1}، مقدار مؤثر بودن عمل فوق را در راستای بردن حالت شبکه از یک وضعیت به یک وضعیت بهتر نشان می‌دهد. این پاداش مقداری بین ۰ تا ۱ می‌گیرد. برای محاسبه این پاداش در هر لحظه از رابطه (۵) استفاده می‌شود:

$$r_{t+1} = \begin{cases} \dfrac{d_t - d_{t+1}}{d_t} & d_{t+1} < d_t \\ 0 & d_{t+1} \geq d_t \end{cases} \qquad (۵)$$

که در آن d_t از رابطه (۶) محاسبه می‌شود:

$$d_t = \sqrt{p_{t,i}^2 + b_{t,im}^2} \qquad (۶)$$

با درنظرگرفتن اینکه $p_{t,i}$ وضعیت حذف بسته‌ها در شبکه و $b_{t,im}$ درصد پر یا خالی بودن بافر پورت خروجی را نشان می‌دهد و با توجه به این مطلب که حالت ایده آل رسیدن به حالتی است که مقدار این دو پارامتر در آن کم باشد، می‌توان مقدار این پاداش را توصیف کرد. به عبارت دیگر با تعریف این پاداش سعی بر آن است که عامل به سمت حالات بهتر ترغیب شود و یا به عبارت دیگر در صفحه حالت از هر نقطه به سمت مبدأ نزدیک شود. درشکل۲ مدلی از سیستم ارایه شده برای هر عامل شکل دهنده ترافیک آمده است. همان طور که در این شکل دیده می‌شود عامل با پسخورگرفتن از محیط، پارامترهای مورد نظر را در SR تنظیم می‌کند.



شکل۲: مدل سیستم پیشنهادی برای هر یک از عاملهای شکل‌دهنده ترافیک.

۱.Action

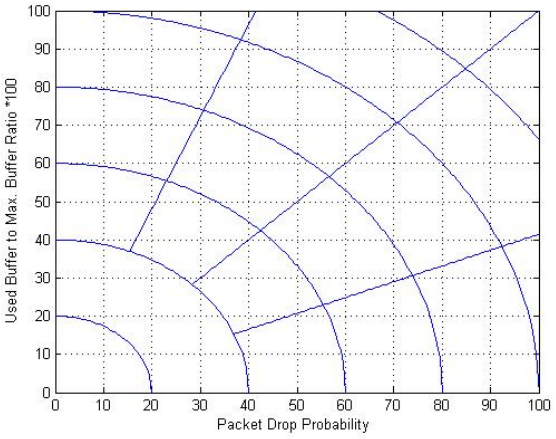
Downloaded from jor.iranaict.ir at 9:07 +0330 on Wednesday November 25th 2026

برای محاسبه پارامترهای ذکر شده در بالا، $b_{t,im}$ که همان درصد پربودن بافر خروجی پورت مورد نظر است به سادگی در داخل SR قابل محاسبه است. اما در مورد محاسبه $p_{t,i}$ دو مسأله مطرح می‌شود. یکی اینکه چگونه مقدار این پارامتر اندازه‌گیری شود و ثانیاً اینکه چگونه مقدار اندازه‌گیری شده را به محل استفاده از آن یعنی SR منتقل کنیم. برای رفع این مسائل از مکانیزم ECN استفاده شده است که می‌تواند به طور سریع و راحت، عملیات سیگنالینگ لازم برای انتقال اطلاعات لازم به SR را انجام دهد. این مکانیزم بجای آنکه بسته‌های اطلاعاتی را متناسب با ازدحام به‌وجود آمده حذف کند، آنها را با احتمالی متناسب با ازدحام در شبکه علامت گذاری می‌کند. به طور ایده آل، امکان اینکه اطلاعات لازم را از NR، یعنی جایی که این علامت گذاری انجام می‌شود، به سمت SR، یعنی جایی که این اطلاعات قرار است استفاده شود، به صورت پسخور مستقیم ارسال کنیم، وجود ندارد. به همین دلیل بسته ACK[1] مربوط به پروتکل TCP می‌تواند در این زمینه کمک مناسبی کند. اگر گیرنده‌های TCP یک بسته اطلاعاتی را با علامت دریافت کنند، می‌توانند این علامت را از سرایندهای IP این بسته (این علامت در دو بیت قرار دارد که در IPv۴ در فیلد TOS و در IPv۶ در فیلد TC وجود دارد) کپی و در داخل سرایند TCP بسته ACK ارسالی به سمت فرستنده قرار دهند و به این طریق، وضعیت ازدحام و حذف بسته‌ها را به SR برسانند. به دلیل آنکه این اطلاعات پسخور شده تنها در دو بیت قرار دارند، در ترافیک شبکه نقش زیادی ندارند[۱۲و۱۳].

SR پس از دریافت بسته‌های ACK و شمردن اینکه چه تعداد از آنها حاوی اطلاعات علامت گذاری شده هستند، در پایان هر چرخه شبکه به محاسبه $p_{t,i}$ طبق رابطه۷ می‌پردازد:

$$p_{t,i} = \frac{\text{تعداد بسته‌های علامت گذاری شده}}{\text{تعداد کل بسته‌های دریافت شده}} \qquad (۷)$$

با داشتن این دو پارامتر و تزریق آن به یک آشکار ساز حالت دو بعدی، در هر لحظه می‌توان حالت فعلی شبکه را مشخص کرد. آشکار ساز حالت حاصل گسسته سازی[2] صفحه حالت[3] است که این گسسته سازی کاملاً به نظر طراح و شرایط سیستم بستگی دارد. در این مقاله آشکار ساز حالت انتخاب شده



شکل۳: آشکار ساز حالت ارائه شده برای عاملهای شکل دهنده ترافیک

مطابق شکل۳ است که در آن محور افقی $p_{t,i}$ و محور عمودی $b_{t,im}$ را نشان می‌دهد. در این شکل تعداد کل حالات در نظر گرفته شده ۲۰ عدد است که توسط خطوط شعاعی و قطری درکل صفحه حالت تقسیم بندی شده اند. انتخاب این تعداد حالت و طرح گسسته سازی صفحه حالت بر اساس تجارب و اهداف مورد نظر انجام شده است. بدیهی است هر چه تعداد این حالات بیشتر شود، دقت بالاترمی‌رود ولی حجم محاسبات لازم افزایش می‌یابد.

برای قالب چند عاملی تخصیص حافظه بافر نیز SR با یک حافظه بافر با حجم M بیت مشخص می‌شود که بین پورتهای روتر به طور مساوی تقسیم می‌شود و هر عامل می‌تواند در اختیار بگیرد. راهکار استفاده شده در این قسمت بر این مبناست که هرعامل تخصیص دهنده حافظه در هر پورت می‌تواند مقداری از حافظه بافر در اختیار خود را در صورت احتیاج نداشتن در اختیار عاملهای تخصیص دهنده دیگری که به آن نیاز دارد بگذارد. میزان حافظه‌ای که هر عامل می‌تواند با دیگران به اشتراک بگذارد کاملاً به نظر طراح بستگی دارد. در این مقاله فرض برآن است که هر عامل می‌تواند نیمی از حافظه در اختیار خود را در اختیار دیگران گذارد. به عبارت دیگر هر عامل در ابتدا می‌تواند نیمی از آنچه را می‌تواند به طور ایستا دراختیار بگیرد، را استفاده کند و نیم دیگر را در۵ مرحله ۱۰درصدی استفاده کند و یا به دیگران قرض دهد. برای پیاده‌سازی این روش توسط عاملها، عمل هر عامل یعنی a یک عدد بین ۰ تا ۸ در نظر گرفته شده است. $a=0$ یعنی این که عامل احتیاجی به حافظه بیش از آنچه در ابتدا دراختیار قرار گرفته است ندارد. $a=1$ یعنی این که ده درصد از بقیه

۱.Acknowledgement Packet
۲.Discretization
۳.State Plain

Downloaded from jor.iranaict.ir at 9:07 +0330 on Wednesday November 25th 2026

حافظه خود را در اختیار می‌گیرد و به همین ترتیب $a = 8$ یعنی این که عامل۳ واحد۱۰ درصدی از بقیه پورتها قرض می‌گیرد .طراحی این عملها و همچنین نحوه اشتراک گذاشتن کاملاً به نظر طراح بستگی دارد. در این طرح به دلیل آنکه هر عامل بتواند در یک شرایط پایدار کار کند و همچنین این فرض که معمولاً به طور میانگین یک بار ترافیکی در شبکه وجود دارد تحت هر شرایطی ۵۰٪ از حافظه بافر هر پورت در اختیار آن قرار گرفته است. از طرف دیگر برای آنکه به دلیل بافر شدن زیاد بارهای ترافیکی سنگین به Time Out شدن و ارسال مجدد در فنّاوری TCP/IP منجر نشود، نهایت حافظه ای که هر عامل توانسته قرض بگیرد به سه واحد۱۰ درصدی یعنی $a = 8$ محدود شده است.

همانطور که قبلاً نیز ذکر شد، حالتهای در نظر گرفته شده برای عاملهای تخصیص دهنده دو پارامتر UBS[1] و SR-DP است. UBS همانند قبل همان درصد پربودن بافر خروجی پورت مورد نظر و به عبارتی i امین پورت در لحظه t است که به mامین NR متصل است و با $b_{t,im}$ نمایش داده می‌شود و SR-DP احتمال حذف بسته‌ها در ورودی پورتهای SR و یا به‌عبارتی iامین پورت SR در لحظه t است که با $sp_{t,i}$ نمایش داده می‌شود و در داخل SR قابل محاسبه است.

با داشتن این دو پارامتر و تزریق آن به آشکار ساز حالت مشابه قسمت قبل می‌توان درهر لحظه حالت فعلی شبکه را مشخص کرد. این آشکار ساز شبیه آشکار ساز استفاده شده برای عاملهای شکل دهنده ترافیک است با این تفاوت که در آن محور افقی از $p_{t,i}$ و یا به عبارتی NR-DP به $sp_{t,i}$ و یا به همان SR-DP تغییر یافته است.

محاسبه پاداش داده شده به هر عامل که در مقابل انجام عمل a درلحظه t دریافت می‌کند، یعنی r_{t+1} نیز شبیه عاملهای شکل دهنده ترافیک طبق روابط ۸ و ۹ محاسبه می‌شود و عددی بین ۰ و ۱ است.

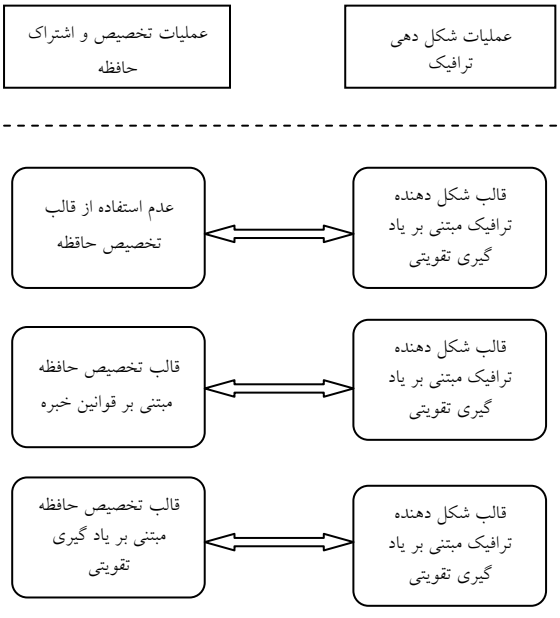
$$r_{t+1} = \begin{cases} \dfrac{d_t - d_{t+1}}{d_t} & d_{t+1} < d_t \\ 0 & d_{t+1} \geq d_t \end{cases} \qquad (۸)$$

$$d_t = \sqrt{sp_{t,i}^2 + b_{t,im}^2} \qquad (۹)$$

باز همچنان که در محاسبه این پاداشها دیده می‌شود، هدف ترغیب عاملها به سمت حالاتی بوده است که نرخ حذف بسته‌ها را در ورودی SR تا حد امکان کاهش دهند و در مقابل یک تخصیص بهینه و مفید بین عاملها انجام دهند.

۵- شبیه‌سازی

برای شبیه‌سازی و ارزیابی مدلهای ارائه شده در این مقاله ، قبل از آنکه به ارزیابی قالبهای چند عاملی شکل دهنده ترافیک و تخصیص دهنده حافظه در کنار یکدیگر پرداخته شود به ارزیابی مدلهای ساده تر آنها می‌پردازیم. در ابتدا مدل چند عاملی شکل دهنده ترافیک مبتنی بر یادگیری تقویتی بدون قالب تخصیص حافظه بر روی چهار پورت یک روتر ارزیابی می‌شود تا اثر تخصیص پویا و هوشمند نرخ تولید نشانه در الگوریتم سطل نشانه دار دیده شود. سپس در مرحله بعد یک قالب تخصیص حافظه که بر مبنای قوانین تجربی و خبره[1] عمل می کند در کنار آن در نظر گرفته می‌شود تا معیار خوبی برای مقایسه با قالب تخصیص دهنده حافظه مبتنی بر یادگیری تقویتی باشد و در نهایت به جای قالب تخصیص حافظه خبره، یک قالب چند عاملی مبتنی بر یاد گیری تقویتی در کنار قالب شکل دهنده ترافیک در نظر گرفته می‌شود تا اثر تخصیص هوشمند و پویای حافظه بافر دیده شود و سپس به بررسی و ارزیابی آنها پرداخته می‌شود. روند نمای شکل۴ نمایی کلی از مراحل مختلف این راهکارها را به ترتیب نشان می‌دهد.



شکل۴: روند نمای ترتیب انجام مراحل راهکارهای پیشنهادی.

۱.Used Buffer Size

۱.Expert

Downloaded from jor.iranaict.ir at 9:07 +0330 on Wednesday November 25th 2026

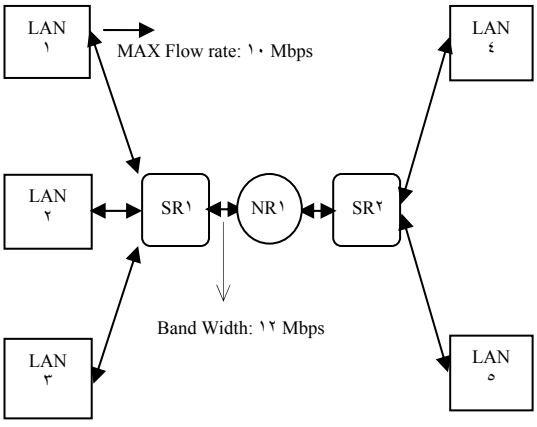
۵-۱- قالب چند عاملی مبتنی بر یادگیری تقویتی برای شکل‌دهی ترافیک بدون قالب تخصیص دهنده حافظه

در این مدل فقط از قالب شکل دهنده ترافیک استفاده شده است. در این قالب عاملهای شکل دهنده ترافیک بر روی چهار پورت یک روتر قرار می‌گیرند و مستقل از یکدیگر باتوجه به شرایط شبکه به ارسال داده‌ها به سمت جلو می‌پردازند. در این مدل هر عامل در هر پورت تنها یک مقدار مشخص از حافظه بافر روتر را می‌تواند به طور ثابت برای این پورت در نظر بگیرد و در نتیجه احتیاجی به قالب چند عاملی تخصیص دهنده حافظه در عقب روتر نیست. به عبارت دیگربافر موجود به طور ایستا و مساوی بین چهار پورت تقسیم می‌شود.

۵-۱-۱- نتایج شبیه‌سازی

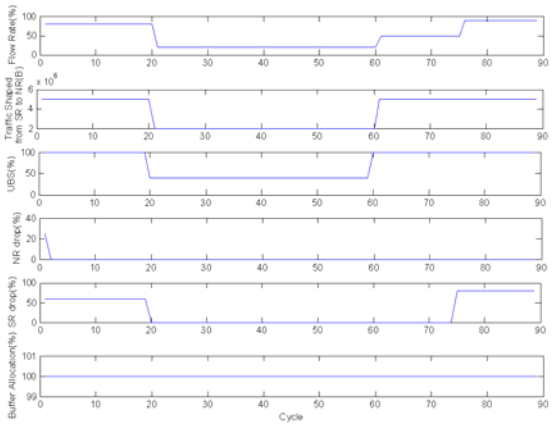
در این قسمت به شبیه‌سازی و ارزیابی مدل ارائه شده در محیط شبیه ساز توسعه داده شده می‌پردازیم. شرایط شبیه‌سازی مطابق شکل۴ است که در آن چهار عامل شکل دهنده ترافیک بر روی چهار پورت SR قرار گرفته اند. برای اینکه مدل توسعه داده شده در شرایط پویا و نزدیک به واقعیت آزمایش شود و قابلیتهای آنها به ویژه در فرایند ارسال و شکل دهی داده‌ها بهتر دیده شود، فلوهای ورودی مختلفی به آنها اعمال می‌شود.

فلوهای ورودی که در این شبیه‌سازی‌ها در نظر گرفته شده به نحوی بوده است که بتواند کارایی قالبهای چند عاملی را در مقابل دینامیک های مختلف اعمالی مشخص کند. نتایج شبیه‌سازی برای پورت اول این SkR در قسمتهای اول تا ششم شکل۶ آورده شده است.



شکل۵: ساختار شبکه در نظر گرفته شده برای شبیه‌سازی راهکارهای پیشنهادی.

همچنان که درقسمت اول این شکل دیده می‌شود فلو ورودی به این پورت به صورت درصدی از پهنای باند ورودی به SR که در این آزمایش ۱۰Mb در نظر گرفته شده نشان داده می‌شود. این فلو تا چرخه حدود ۲۰ در حدود ۸۰ درصد پهنای باند ورودی و پس از آن تا چرخه حدود ۶۰ در حدود ۲۰ درصد پهنای باند ورودی و پس از آن در دو مرحله باز افزایش داشته است. به تبع آن می‌توان نرخ ارسال عامل شکل دهنده ترافیک را به سمت NR درقسمت دوم شکل۶ و دو پارامتر تصمیم گیرنده این عامل یعنی درصد پر و خالی بودن بافر مربوط به هر پورت (UBS)[۱] ومیزان حذف بسته‌ها در NR (NR- DP)[۲] را به ترتیب درقسمتهای سوم و چهارم مشاهده کرد. نرخ حذف بسته‌ها در ورودی SR (SR-DP)[۳] که به عنوان یکی از پارامترهای عامل تخصیص دهنده حافظه در قسمتهای بعدی از آن استفاده می‌شود در قسمت پنجم و عمل تخصیص حافظه در قسمت ششم این شکل نمایش داده شده است که در این قسمت به دلیل ثابت بودن حافظه بافر برای هر پورت در کل مسیر فرایند، عامل در همه شرایط ۱۰۰ درصد بافرخود را در اختیار داشته است. در این شکل دیده می‌شود که با اعمال این فلو به ورودی پورت، عامل بدون در نظر گرفتن شرایط شبکه در پشت روتر و حذف بسته‌ها که در ورودی آن در حال اتفاق است (SR-DP) و همچنین مستقل از بقیه عاملها، ترافیک آمده را به نحوی به سمت NR ارسال کرده است که حذف بسته‌ها در مقابل خود یعنی NR-DP را و همچنین میزان استفاده از بافر خود یعنی UBS را پایین نگه دارد. در این شکل دیده می‌شود که در کل مسیر، به جز چرخه‌های اولیه، NR-DP صفر بوده ولی UBS بالا بوده است که این به دلیل ثابت بودن بافر پورت و زیاد بودن نسبی حجم ترافیک ورودی بوده است.



شکل۶: نتایج شبیه‌سازی عاملهای شکل دهنده ترافیک بدون تخصیص حافظه برای پورت اول روتر.

۱.Used Buffer Size
۲.NR-Drop
۳.SR-Drop

Downloaded from jor.iranaict.ir at 9:07 +0330 on Wednesday November 25th 2026

۵-۲- قالب چند عاملی مبتنی بر یادگیری تقویتی برای شکل‌دهی ترافیک در کنار قالب چندعاملی تخصیص دهنده حافظه مبتنی بر قوانین خبره

دراین مدل از دو قالب چند عاملی مستقل در یک SR با چهار پورت خروجی استفاده می‌شود. فرض می‌شود یکی از این قالبها همانند قبل در جلوی روتر نقش شکل دهی ترافیک به سمت جلو و دیگری به طور همزمان براساس قوانین خبره به تخصیص حافظه بین این پورتها و بر اساس نیاز هر پورت می‌پردازد. به دلیل آنکه فرض می‌شود SR دارای چهار پورت خروجی است، هر کدام از این قالبها دارای ۴ عامل است به نحوی که هر کدام از آنها در یک پورت SR قرار می‌گیرد. در واقع هر پورت یک عامل برای شکل‌دهی ترافیک و یک عامل برای تخصیص حافظه بافر به آن بر اساس شرایط شبکه دارد. عامل شکل دهنده ترافیک بر اساس مدل ارائه شده در قسمت قبل و بر پایه یادگیری تقویتی به شکل دهی ترافیک می‌پردازد با این تفاوت که در این مدل، این عامل یک حافظه بافر مشخص و ثابت را در اختیار ندارد بلکه به کمک عامل تخصیص دهنده حافظه، در هر لحظه و بر اساس نیاز خود و همچنین شرایط دیگر پورتها این مقدار حافظه را برای خود مشخص می‌سازد.

همانطور که درمدل قبلی نیز مشخص بود، برای عاملهای شکل دهنده ترافیک ارسال داده‌ها با حداکثر نرخ ارسال ممکن تا جایی اهمیت داشت که احتمال نرخ حذف بسته‌ها در NR پایین بماند. به همین منظور این عاملها به نحوی آموزش داده شدند که با درنظرگرفتن این پارامتر وهمچنین میزان پر یا خالی بودن بافر خود، نرخ ارسال خود را تنظیم کنند. به عبارتی قالب شکل‌دهنده ترافیک در این مدل تنها با در نظرگرفتن شرایط در مقابل SR به‌عنوان پارامترهای تصمیم‌گیری خود، به ارسال داده‌ها می‌پرداخت.

اما در این مدل با در نظرگرفتن یک قالب تخصیص دهنده حافظه در کنار قالب شکل دهنده ترافیک نه تنها استفاده از حافظه بافر SR به طور موثرتر و انعطاف پذیر صورت می‌پذیرد، بلکه به نوعی می‌توان شرایط ترافیک آمده به سمت SR را نیز در نظر گرفت. به عبارت دیگر در قالب شکل‌دهنده ترافیک، عاملها به صورت مستقل و تنها بر اساس شرایط جلوی SR عمل می‌کردند و هیچ توجهی به شرایط در عقب SR مانند وضعیت ترافیک ورودی به آن نداشتند.

با درنظرگرفتن نرخ حذف بسته‌ها در ورودی SR به‌عنوان یکی از پارامترها و همچنین میزان درصد پر یا خالی بودن بافرخروجی SR به‌عنوان متغیرهای تصمیم گیرنده برای عاملهای تخصیص دهنده و استفاده از قوانین خبره در این مدل به تخصیص دینامیک حافظه بافر بین این چهارپورت پرداخته می‌شود. این چهار عامل در این چهار پورت، با درنظرگرفتن وضعیت یکدیگر در مقدار این پارامترها، و همچنین انتخاب یک راه بهینه بر اساس قوانین خبره، باعث می‌گردند تا هر پورت بر اساس نیاز خود به استفاده از حافظه به صورت مشترک بپردازند. این قوانین خبره به صورت دستورهای اما و اگر و بر اساس اهداف طراح و تجربه‌های قبلی پیاده‌سازی می‌شوند. در شرایطی که ترافیک‌های موج گونه لحظه‌ای به سمت بعضی از پورتها می‌رود این شانس به آنها داده می‌شود تا درصورت نیاز نداشتن دیگر پورتها به حافظه خود، از آن به صورت لحظه‌ای استفاده کنند و به نوعی نرخ حذف بسته‌ها را در شبکه پایین آورند.

روند تخصیص حافظه بین این عاملها همانند مدل توسعه داده شده در قسمتهای قبلی است و فرض می‌شود هر عامل در ابتدا نیمی از آنچه می‌تواند به طور ایستا و ثابت (منظور از ایستا آن است که مانند قسمت قبل از همان ابتدا حافظه موجود بین آنها به طور مساوی تقسیم شود) در اختیار داشته باشد را دراختیار می‌گیرد و نیم دیگر را در ۵ مرحله می‌تواند با بقیه عاملها (پورتها) به اشتراک بگذارد. به عبارت دیگر هر عامل می‌تواند درهر مرحله ۱۰٪ از حافظه خود را در اختیار دیگران بگذارد و یا در صورت نیاز خود علاوه بر ۵۰٪ اولیه، آن را در اختیار بگیرد. دراین مدل فرض می‌شود که هر عامل در صورت نیاز به حافظه قرض داده شده خود در هر لحظه می‌تواند آن را از دیگر عاملها پس بگیرد.

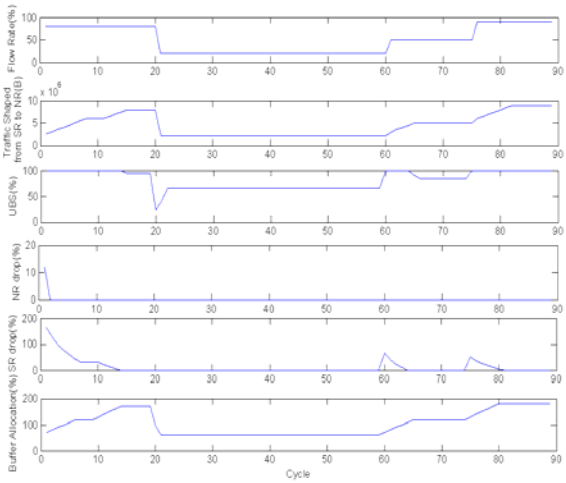
۵-۲-۱- نتایج شبیه‌سازی

دراین قسمت فرض می‌شود مدل ارائه شده، در داخل یک SR و مطابق شکل۵ شبیه‌سازی می‌شود. شرایط شبیه‌سازی دقیقاً مطابق با قسمت قبل درنظرگرفته می‌شود تا در هر مرحله مزایا و معایب مدلهای پیشنهادی به خوبی مشخص شود.

SR در نظر گرفته شده دارای چهار پورت خروجی است که دو قالب چند عاملی هر کدام با چهار عامل در داخل آن قرارمی‌گیرد. یکی از این قالبها برای شکل دهی هوشمند ترافیک در چهار پورت به صورت مستقل و یکی از این قالبها برای به اشتراک گذاشتن و اختصاص حافظه به هر پورت.

همانطور که در شکل۷ برای پورت اول این SR دیده می‌شود، برای این که این عاملها بتوانند درشرایط نزدیک به

Downloaded from jor.iranaict.ir at 9:07 +0330 on Wednesday November 25th 2026



شکل۷: نتایج شبیه‌سازی عاملهای شکل دهنده ترافیک درکنارعاملهای تخصیص دهنده حافظه خبره برای پورت اول روتر.

واقعیت آزمون شوند و قابلیتهای آنها به ویژه در فرایند اشتراک گذاری حافظه بهتر دیده شود، و ازطرف دیگر بین قسمتهای مختلف مقایسه صورت گیرد، فلوهای ورودی مطابق با قسمت قبل به آنها اعمال می‌شود. اگر به این نتایج دقت شود و با نتایج قسمت قبل درشکل۶ مقایسه شود اثر اضافه شدن قالب تخصیص دهنده حافظه در این مدل در مقایسه با مدل قبلی که حافظه عامل ثابت بود و هیچ گونه تعاملی بین عاملها در به اشتراک گذاشتن حافظه وجود نداشت بهتر دیده می‌شود. اگر دقت شود اضافه شدن این قالب در کنار قالب شکل دهنده ترافیک و پویا شدن عمل تخصیص حافظه، نه تنها باعث شده نرخ حذف بسته‌ها در ورودی SR کاهش یابد، بلکه باعث شده نرخ ارسال عاملهای شکل دهنده ترافیک به سمت NR نیز افزایش یابد در حالی که نرخ حذف در NR همچنان پایین است. این موضوع نشانگر آن است که این دو قالب با وجودی که مستقل از یکدیگر هستند با یکدیگردر تعاملند. به عبارت دیگر، موقعی که فلوورودی به این پورت SR افزایش داشته است، عاملهای تخصیص دهنده حافظه سعی کرده اند حافظه بافر بیشتری در اختیار آن قرار دهند و این باعث شده تا عامل شکل دهنده ترافیک، UBS کمتری داشته باشد و به حالتی در نقشه آشکار ساز حالت خود برود که باعث افزایش نرخ ارسال آن شود.

۵-۳- قالب چند عاملی مبتنی بر یادگیری تقویتی برای شکل دهی ترافیک درکنار قالب چند عاملی تخصیص دهنده حافظه مبتنی بر یادگیری تقویتی

در این مدل نیز مانند مدل قبلی از دو قالب چند عاملی که هر کدام دارای چهار عامل هستند استفاده می‌شود. این دو قالب در یک SR با چهارپورت خروجی قرار گرفته و به شکل دهی هوشمند ترافیک و همچنین تخصیص حافظه بین پورتها می‌پردازند.

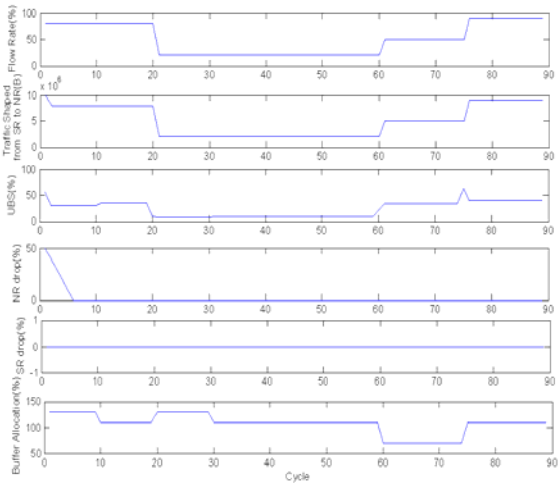
تفاوتی که در این مدل با مدل قبلی وجود دارد، استفاده از یادگیری تقویتی برای آموزش عاملها در قالب تخصیص دهنده حافظه است. در مدل قبلی، فقط قالب شکل دهنده ترافیک از این روش برای ارسال داده‌ها به سمت جلو استفاده می‌کرد و قالب تخصیص دهنده حافظه فقط بر اساس شرایط شبکه که نرخ افت بسته در SR و همچنین UBS در نظرگرفته شده بود و استفاده از یک سری قوانین تجربی به تخصیص و اشتراک حافظه بین پورتها می‌پرداخت. در این مدل، این عاملها با استفاده از این پارامترها برای هر پورت و روش یادگیری تقویتی می آموزند که چگونه این عمل را به طور پویا و هوشمند انجام دهند. نحوه طراحی ساختار یادگیری تقویتی برای این عاملها شبیه عاملهای شکل دهنده ترافیک است با این تفاوت که برای حالتهای در نظر گرفته شده برای آن به جای استفاده از NR- DP ، ازSR-DP استفاده می‌شود که مانند UBS در داخل SR محاسبه می‌شود.

۵-۳-۱- نتایج شبیه‌سازی

در این قسمت به ارزیابی مدل چند عاملی ارائه شده در قسمت قبلی پرداخت می‌شود. سناریو و شرایط شبیه‌سازی دقیقاً مطابق با شرایط و سناریوی شبیه‌سازی مدلهای قبلی است تا بتوان درمجموع بین مدلهای ارائه شده مقایسه صورت گیرد و مزایا و معایب آنها مشخص گردد. دو قالب چند عاملی مبتنی بر یادگیری تقویتی، هر کدام با چهار عامل، در یک SR با چهار پورت مطابق با شکل۵ قرار می‌گیرند. در یک قالب عاملها با همکاری یکدیگر ترافیک آمده به سمت هر پورت را مطابق با وضعیت شبکه به سمت جلو شکل دهی می‌کنند تا در حدامکان، از پهنای باند جلو استفاده بهینه شود تا حدی که نرخ حذف بسته‌ها در NR پایین باشد. در قالب دیگر عاملها با در نظر گرفتن وضعیت یکدیگر در میزان ترافیک آمده به سمت آنها و همچنین میزان استفاده فعلی خود از حافظه بافر موجود، به تخصیص حافظه واشتراک آن بین یکدیگر می‌پردازند. فلو ورودی اعمالی به SR و برای هر پورت دقیقاً مطابق با فلو اعمالی در شبیه‌سازی مدلهای قبلی است و به گونه ای بوده است تا دینامیک استفاده از پهنای باند و همچنین تخصیص حافظه به خوبی نمایانگر شود.

همانطور که در شکل۸ برای یک پورت این SR دیده می‌شود، در این حالت فرایند تخصیص حافظه پایدار و منظم بوده و از طرف دیگر، عاملهای شکل دهنده ترافیک با پایین نگاه داشتن نرخ حذف بسته‌ها در NR بهتر توانسته اند از پهنای باند مقابل استفاده کنند.

Downloaded from jor.iranaict.ir at 9:07 +0330 on Wednesday November 25th 2026



شکل۸: نتایج شبیه‌سازی عاملهای شکل دهنده ترافیک درکنارعاملهای تخصیص دهنده حافظه مبتنی بر یادگیری تقویتی برای پورت اول روتر.

در این شکل دیده می‌شود که نرخ حذف بسته‌ها در ورودی SR نیز به صفر کاهش یافته است. اگر یک مقایسه متناظر بین این نتایج و نتایج به دست آمده در مدل قبلی در شکل۷ که در آن فرایند تخصیص حافظه بر اساس قوانین تجربی انجام می‌شد صورت گیرد مشخص است که در مدل اخیر علاوه بر بهبود وضعیت حذف بسته‌ها در SR و همچنین پایین تر بودن درصد UBS ، ترافیک شکل داده شده از SR به NR و یا به عبارتی میزان استفاده از پهنای باند مقابل بیشتر شده است. در ضمن آنکه عاملها با در نظر گرفتن شرایط پروتکلهای TCP/IP , علاوه برآن که منظم تر و پایدارتر به تخصیص حافظه پرداخته‌اند، از قرض گرفتن بیش ازحد مجاز حافظه پرهیز کرده‌اند.

۶- نتایج

در این مقاله به مسئله شکل دهی ترافیک در شبکه‌های کامپیوتری و با استفاده ار روشهای هوشمند پرداخته شد. با توجه به ساختار توزیع شده شبکه‌های کامپیوتری و رفتار تصادفی موجود در آنها و از طرف دیگر محدودیت‌های زمانی که در الگوریتم‌های کنترلی برای اینگونه سیستمها وجود دارد، از مفاهیم سیستم‌های چند عاملی و روشهای یادگیری تقویتی برای شکل دهی ترافیک در روترها وتخصیص دینامیک حافظه بافر بین پورتهای مختلف یک روتر استفاده گردید. در اینجا با استفاده از این مفاهیم شکل دهنده ترافیک جدیدی بر مبنای یک الگوریتم سطل نشانه دار توسعه داده شد که در آن به جای آنکه نرخ تولید نشانه‌ها به طور ایستا تخصیص داده شود به طور پویا و هوشمند و برمبنای وضعیت شبکه مشخص می‌شود. این پیاده‌سازی علاوه بر آنکه به استفاده بهینه و منطقی از پهنای باند منجر شد موجب کاهش ترافیک در دیگر نقاط شبکه نیز گردید. علاوه بر این یک روش جدید دیگر برای تخصیص هوشمند و پویای حافظه بافر برای پورتهای یک روتر توسعه داد شد. نتایج شبیه‌سازی‌های انجام شده کار آمدی روشهای پیشنهادی را نشان داد.

با توجه به ابعاد زیاد داده‌ها در شبکه و متنوع بودن متغیرهای موثر در ساختار توزیع شده آن، و از طرف دیگر در نظر گرفتن این موضوع که با افزایش پارامترها در الگوریتم یادگیری تقویتی حجم محاسبات افزایش و به تبع آن سرعت یادگیری کاهش می‌یابد، در ادامه می‌توان به دنبال راهکارهایی بود که بتواند پارامترهای مؤثر بیشتری را در تصمیم‌گیری عاملها در بستر شبکه در نظرگیرد و در عین حال از پیچیدگی و اضافه شدن حجم سیستم جلو گیری کند. این راهکارها نه تنها باعث می‌شود از پارامترهای موثر بیشتری در بستر شبکه برای شکل دهی ترافیک و تخصیص حافظه استفاده شود، بلکه از دید یادگیری تقویتی می‌توان آن را روشی برای کاهش تعداد حالات که یکی از معضلات این روش در محیطهای پیچیده است، درنظرگرفت.

سپاسگزاری

بخشی از هزینه‌های تحقیقاتی این پروژه با حمایت مالی مرکز تحقیقات مخابرات ایران تأمین شده است که از این حمایت قدردانی می‌شود.

مراجع

[۱] Park E., H. Choi, "Adaptive Token Bucket Algorithm for fair Bandwidth Allocation in Diffserve Networks". School of Electrical Engineering and computer science, Seoul National University, Seoul, Korea, ۲۰۰۲.

[۲] Radhakrishnan S., S.V. Raghavan, K.Agrawala, "Flexible Traffic Shaper for High Speed Networks Design and comparative study with leaky Bucket", Computer Networks and ISDN systems ۲۸, ۴۵۳-۴۶۹, ۱۹۹۶.

[۳] Davison R.G., J.J.Hardwick, , "Applying the agent paradigm to network management., BT Technology Journal., Vol ۱۶.,No۳.,۱۲۵-۱۳۶, ۱۹۹۸.

Downloaded from jor.iranaict.ir at 9:07 +0330 on Wednesday November 25th 2026

[۴] Harle D., P.vila, "A Multi Agent Approach to dynamic virtual path management in ATM networks", Communication Division, Department of Electronics & Electrical Engineering, university of strathclyde, UK,۲۰۰۴.

[۵] Kaebling L.P., M.L. Littman, A.W. Moore, "Reinforcement Learning: A Survey". Journal of Artificial Intelligence Research, ۴: ۲۳۷-۲۸۵, ۱۹۹۶.

[۶] Babuska R., L. Busoniu, "Reinforcement Learning for MultiAgent systems", Delft Center for Systems and control,۲۰۰۴.

[۷] Watkins C., "Q-Learning. Machine Learning",۸: ۲۷۹-۲۹۲, ۱۹۹۲.

[۸] Chen T.M., S. Liu,"ATM Switching Systems", Artech House Publishers, ۱۹۹۵.

[۹] Tanenbaum A.S, "Computer Networks" ,Prentice Hall ,۲۰۰۴.

[۱۰] Cisco TechNotes, "Comparing Traffic policing and Traffic Shaping for Bandwidth Limiting", Document ID: ۱۹۶۴۵

[۱۱] Sutton R., A.G Barto, "Reinforcement learning: An Introduction", MIT Press, Cambridge, MA, USA, ۱۹۹۶.

[۱۲] Ramakrishnan K., S. Floyd, "A Proposal to add Explicit congestion notification (ECN) to IP. "RFC ۲۴۸۱, ۱۹۹۹.

[۱۳] Shames I., N. Najmaei, M. Zamani, A.A. Safavi, "Application of Reinforcement learning in Development of a new Adaptive Intelligent Traffic Shaper", ICMLA'۰۶ , Orlando, Florida, December, ۲۰۰۶.

[۱۴] Tumer K., Agogino A., "Agent Reward shaping for Alleviating Traffic Congestion" AAMAS'۰۶ Hokkaido, Japan, May, ۲۰۰۶.

[۱۵] Buffet O.,A. Dutech,F. Charpillet, "Shaping Multi-Agent Systems with Gradient reinforcement Learning",AAMASJ۰'۷, ۲۰۰۷.