



دانشکده مهندسی کامپیوتر

استراتژی تخلیه محاسباتی احتمالی برای وظایف ناهمگون

پروژه کارشناسی مهندسی کامپیوتر

محمد مبین داریوش همدانی

استاد راهنما

رضا انتظاری ملکی

خرداد ۱۴۰۱



تأییدی هیأت داوران جلسه دفاع از پروژه

نام دانشکده: دانشکده مهندسی کامپیوتر

نام دانشجو: محمدمبین داریوش همدانی

عنوان پروژه : استراتژی تخلیه محاسباتی احتمالی برای وظایف ناهمگون

تاریخ دفاع: خرداد ۱۴۰۱

رشته: مهندسی کامپیوتر

ردیف	سمت	نام و نام خانوادگی	مرتبۀ دانشگاهی	دانشگاه یا مؤسسه	امضاء
۱	استاد راهنما	دکتر رضا منتظاری ملکی	استادیار	دانشگاه علم و صنعت ایران	
۲	استاد داور داخلی	دکتر	دانشگاه علم و صنعت ایران	

تأییدی صحت و اصالت نتایج

باسمه تعالی

اینجانب محمدمبین داریوش همدانی به شماره دانشجویی ۹۶۵۲۱۱۹۱ دانشجوی رشته مهندسی کامپیوتر مقطع تحصیلی کارشناسی تأیید می‌نمایم که کلیه‌ی نتایج این پروژه حاصل کار اینجانب و بدون هرگونه دخل و تصرف است و موارد نسخه‌برداری شده از آثار دیگران را با ذکر کامل مشخصات منبع ذکر کرده‌ام. در صورت اثبات خلاف مندرجات فوق، به تشخیص دانشگاه مطابق با ضوابط و مقررات حاکم (قانون حمایت از حقوق مؤلفان و مصنفان و قانون ترجمه و تکثیر کتب و نشریات و آثار صوتی، ضوابط و مقررات آموزشی، پژوهشی و انضباطی) با اینجانب رفتار خواهد شد و حق هرگونه اعتراض در خصوص احقاق حقوق مکتسب و تشخیص و تعیین تخلف و مجازات را از خویش سلب می‌نمایم. در ضمن، مسئولیت هرگونه پاسخگویی به اشخاص اعم از حقیقی و حقوقی و مراجع ذیصلاح (اعم از اداری و قضایی) به عهده‌ی اینجانب خواهد بود و دانشگاه هیچ‌گونه مسئولیتی در این خصوص نخواهد داشت.

نام و نام خانوادگی: محمدمبین داریوش همدانی

تاریخ و امضا:

مجوز بهره‌برداری از پایان‌نامه

- بهره‌برداری از این پایان‌نامه در چهارچوب مقررات کتابخانه و با توجه به محدودیتی که توسط استاد راهنما به شرح زیر تعیین می‌شود، بلامانع است:
- ☐ بهره‌برداری از این پایان‌نامه برای همگان بلامانع است.
 - ☐ بهره‌برداری از این پایان‌نامه با اخذ مجوز از استاد راهنما، بلامانع است.
 - ☐ بهره‌برداری از این پایان‌نامه تا تاریخ ممنوع است.

استاد راهنما: رضا انتظاری ملکی

تاریخ:

امضا:

چکیده

رایانش لبه‌ای یک الگوی محاسبات توزیع شده است که با نزدیک کردن منابع پردازشی به لبه شبکه، سعی دارد تا مشکلاتی مانند زمان پاسخگویی، محدودیت باتری و تحرک پذیری را برای کاربران برطرف کند. از زمان معرفی رایانش لبه‌ای و استانداردهای معروف آن مانند رایانش لبه‌ای چند دسترسی^۱، یکی از چالش‌های مهم این حوزه طراحی استراتژی‌های کارآمد برای تخلیه وظایف بوده است.

علاوه بر این، با رشد روز افزون صنعت تلفن همراه و اینترنت اشیا، انواع زیادی از کاربردهای نرم‌افزاری جدید با نیازمندی‌های پردازشی متفاوت در سطح شبکه به وجود آمده است. بنابراین یک ویژگی مهم در طراحی استراتژی تخلیه وظایف در رایانش لبه‌ای، در نظر گرفتن ناهمگونی کاربردها از نظر میزان منابع مورد نیاز است.

در این مقاله روشی برای بدست آوردن استراتژی تخلیه وظایف با تاخیر کمینه تحت محدودیت توان مصرفی ارائه شده است. روش پیشنهادی شامل دو قسمت می‌باشد. در قسمت اول، سیستم تخلیه وظایف با کمک زنجیره مارکوف گسسته-زمان مدل‌سازی می‌شود و در قسمت دوم، با استفاده از الگوریتمی مبتنی بر برنامه‌ریزی خطی^۲ استراتژی تخلیه بهینه برای مدل ساخته شده محاسبه می‌شود.

علاوه بر تشریح و حل مسئله به صورت تئوری، ساختار نرم‌افزاری جدیدی در زبان Kotlin ارائه می‌شود که می‌توان با استفاده از آن استراتژی بهینه را برای سیستم مورد نظر بدست آورد و عملکرد آن را با کمک شبیه‌سازی بررسی کرد. مقاله فعلی گسترشی بر پژوهش [۱] است و از روشی مشابه با روش ارائه شده در این پژوهش استفاده می‌کند.

واژگان کلیدی: تخلیه وظیفه، رایانش لبه‌ای، زنجیره مارکوف، برنامه‌ریزی خطی، رایانش ابری

^۱ Multi-access Edge Computing

^۲ Linear Programming

فهرست مطالب

چ	فهرست تصاویر
ح	فهرست جداول
خ	فهرست علائم اختصاری
۱	فصل ۱: مقدمه
۳	فصل ۲: مروری بر ادبیات و کارهای انجام شده
۴	فصل ۳: شرح مسئله
۵	۱-۳ مدل وظایف
۶	۲-۳ مدل دستگاه کاربر
۸	۳-۳ مدل زمان
۸	۴-۳ مدل کانال بیسیم
۹	۵-۳ مفهوم کنش
۱۰	۶-۳ استراتژی تخلیه
۱۱	۷-۳ روند فعالیت سیستم تخلیه وظایف
۱۲	فصل ۴: روش پیشنهادی
۱۲	۱-۴ استراتژی تخلیه تصادفی
۱۳	۲-۴ مدل زنجیره مارکوف دستگاه کاربر

۳-۴	محاسبه تاخیر میانگین	۱۷
۴-۴	توان مصرفی میانگین	۲۰
۵-۴	استراتژی تخلیه وظیفه بهینه	۲۰

فصل ۵: آزمایش و نتایج

۱-۵	ساختار نرم‌افزاری Kompute	۲۳
۱-۱-۵	ساخت و حل یک مسئله تخلیه وظیفه نمونه در Kompute	۲۵
۲-۵	دو بهینه‌سازی برای الگوریتم جستجوی استراتژی	۲۶
۱-۲-۵	کاهش تعداد متغیرها	۲۶
۲-۲-۵	موازی‌سازی	۲۶
۳-۵	نتایج شبیه‌سازی	۲۷
۱-۳-۵	شبیه‌سازی تک صف	۲۸

فصل ۶: جمع‌بندی و پیشنهادها

۳۷	مراجع
----	-------

فهرست تصاویر

۱-۳	ساختار کلی سیستم تخلیه پردازش	۴
۲-۳	روند فعالیت دستگاه کاربر	۱۱
۱-۴	یک مارکوف نمونه برای مسئله پاکبختگی	۱۳
۲-۴	زنجیره مارکوف سیستم تخلیه در قالب گراف جهت دار	۱۴
۱-۵	کلاس دیاگرام فریمورک Kompute	۲۴
۲-۵	تاخیر سرویس به ازای نرخ ورود در حالت تک صف	۲۸

فهرست جداول

لیست کنش‌ها در سیستم با یک صف وظیفه	۱-۳
دسته‌بندی کنش‌ها در سیستم با k صف	۲-۳
پارامترهای محیط پردازش لبه‌ای در سناریو تک صف	۱-۵

فهرست علایم اختصاری

τ	حالت دستگاه کاربر
q_i	تعداد وظایف موجود در صف i -ام
α_i	نرخ ورود وظیفه به صف i -ام
β	احتمال ارسال موفق بسته
S	مجموعه تمام حالت‌های دستگاه کاربر
A	مجموعه تمام کنش‌های ممکن
η_i	نسبت وظایف نوع i که محلی اجرا می‌شوند
P_{tx}	توان مصرفی برای ارسال یک بسته
P_{loc}	توان مصرفی برای اجرای محلی به اندازه یک بازه زمانی
P_{max}	حداکثر توان مصرفی قابل قبول
L_i	تعداد بازه زمانی لازم برای پردازش محلی وظایف نوع i
M_i	تعداد بازه زمانی لازم برای تخلیه وظایف نوع i
C_i	تعداد بازه زمانی لازم برای پردازش لبه‌ای وظایف نوع i
t_{rx}	زمان اضافه لازم برای بازدریافت وظیفه از سرور لبه‌ای
a_G	میانگین تعداد یال‌های هر گره گراف G
Q	ظرفیت هر صف وظیفه
C_L	تعداد قسمت اجرا شده از وظیفه تخصیص داده شده به پردازنده محلی
C_R	تعداد قسمت ارسال شده از وظیفه تخصیص داده شده به واحد ارسال
T_L	نوع وظیفه تخصیص داده شده به پردازنده محلی

T_R	نوع وظیفه تخصیص داده شده به واحد ارسال
Δ	طول هر بازه زمانی
π_τ	احتمال حضور در حالت τ در توزیع پایدار زنجیره مارکوف
$\chi_{\tau',\tau}$	احتمال گذر از حالت τ' به τ در زنجیره مارکوف
g_τ^a	احتمال انتخاب کنش a در حالت τ در استراتژی g

فصل ۱

مقدمه

افزایش روز افزون تعداد دستگاه‌های موجود در لبه شبکه در سال‌های اخیر، و همچنین معرفی کاربردهای نرم افزاری جدید که نیازمند منابع محاسباتی بالا هستند باعث شده است که تقاضای زیادی برای خدمات پردازش ابری بوجود بیاید. پردازش ابری این امکان را به دستگاه‌های هوشمند از جمله تلفن همراه و اینترنت اشیا می‌دهد که بخشی از پردازش‌های سنگین خود را به سرورهای قدرتمند «تخلیه» کنند تا بر محدودیت‌های پردازشی خود غلبه کنند و کاربردهای نرم افزاری پیچیده‌ای مانند واقعیت افزوده و خودروهای هوشمند را برای کاربران فراهم کنند.

با این وجود، پیاده‌سازی‌های سنتی پردازش ابری یک ایراد ذاتی دارند، و آن فاصله زیاد سرورهای ابری با دستگاه‌های پایانی است. معماری رایانش لبه‌ای و پیاده‌سازی‌های استاندارد آن مانند رایانش لبه‌ای دسترسی-چندگانه که توسط سازمان ETSI ارائه شده است، سعی دارند تا با آوردن بخشی از منابع محاسباتی به نزدیکی لبه شبکه، این مشکل را تا حدی برطرف کنند. علاوه بر تمایل دستگاه‌های لبه شبکه به کمتر شدن این فاصله و به عبارتی «کشش» منابع محاسباتی توسط آن‌ها به منظور افزایش کیفیت سرویس، شرکت‌های ارائه‌دهنده خدمات ابری نیز تمایل دارند تا با «فشردن» بخشی از منابع محاسباتی خود به لبه شبکه، بار محاسباتی و هزینه‌های تجهیزاتی خود را کاهش دهند. [۲]

یک امر مهم در پیاده‌سازی کارآمد رایانش لبه‌ای، طراحی استراتژی‌های تخلیه وظایف هوشمند و موثر است. این استراتژی‌ها نحوه تخصیص منابع توسط دستگاه کاربر^۱ را مشخص می‌کنند و این امکان را به دستگاه کاربر می‌دهند تا درباره تخلیه یا عدم تخلیه وظایف محاسباتی در طول زمان تصمیم بگیرد.

در این مقاله روشی برای بدست آوردن استراتژی تخلیه بهینه ارائه می‌دهیم که مبتنی بر زنجیره مارکوف گسسته-زمان و برنامه‌ریزی خطی می‌باشد. روش پیشنهادی گسترشی بر روش ارائه شده در [۱] می‌باشد. نوآوری و مزیت اصلی روش پیشنهادی ما نسبت به مقاله ذکر شده قابلیت پشتیبانی از وظایف با نیازمندی‌های پردازشی و شبکه‌ای متفاوت (وظایف ناهمگون) می‌باشد. انگیزه اصلی از این گسترش، تنوع محاسباتی وظایف در محیط‌های اینترنت اشیا بوده است. به طور مثال در بسیاری از پژوهش‌های حوزه تخلیه وظیفه در اینترنت اشیا، وظایف به دو دسته «سبک» و «سنگین» تقسیم می‌شوند. [۳] [۴] برای درک مفهوم وظایف سبک و سنگین می‌توان مثال اتومبیل خودران را در نظر گرفت. در این کاربرد، وظیفه پردازش اطلاعات تصاویر به منظور راندن خودرو یک وظیفه سنگین محسوب می‌شود، در حالی که وظیفه روشن کردن سیستم گرمایشی خودرو بر حسب داده‌ی سنسور دما، یک وظیفه سبک محسوب می‌شود.

ادامه این مقاله به پنج فصل تقسیم شده است. در فصل ۲ پژوهش‌های مرتبط انجام شده را مرور می‌کنیم. در فصل ۳ به شرح مسئله تخلیه وظیفه و ساختار رایانش لبه‌ای می‌پردازیم. در فصل ۴ روش پیشنهادی برای بدست آوردن استراتژی تخلیه بهینه را شرح می‌دهیم. در فصل ۵ ابتدا ساختار نرم‌افزاری^۲ جدیدی مبتنی بر زبان Kotlin با نام Kompute ارائه می‌دهیم که این امکان را به کاربران و پژوهشگران می‌دهد تا استراتژی تخلیه بهینه را به ازای سیستم دلخواه خود محاسبه کنند و آن استراتژی را با سایر استراتژی‌های پایه^۳ مقایسه کنند. در بخش دوم از فصل ۴ با استفاده از Kompute به طور جامع به آزمایش و شبیه‌سازی روش ارائه شده در بخش ۴ می‌پردازیم. در انتها در فصل ۶ یک جمع‌بندی کلی از تمامی مطالب ارائه می‌دهیم و پیشنهاداتی نیز برای گسترش روش پیشنهادی ارائه می‌کنیم.

¹User Equipment

²Framework

³Baseline

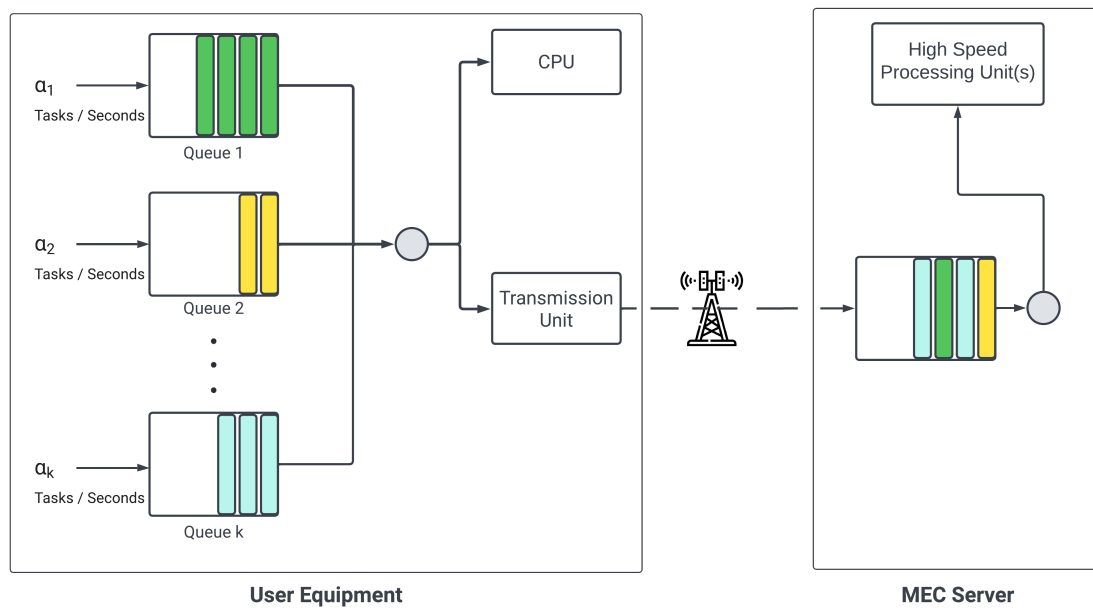
فصل ۲

مروری بر ادبیات و کارهای انجام شده

فصل ۳

شرح مسئله

در این مقاله قصد داریم در یک سامانه رایانش لبه‌ای مطابق با شکل ۳-۱، استراتژی تخلیه‌ای بیابیم که تاخیر سرویس میانگین \bar{T} را تحت محدودیت توان مصرفی P_{max} در درازمدت کمینه کند.



شکل ۳-۱: ساختار کلی سیستم تخلیه پردازش

همانطور که در شکل ۱-۳ مشاهده می‌شود، در سامانه مد نظر سه مولفه اصلی وجود دارد:

۱. دستگاه کاربر (User Equipment)

۲. سرور پردازش لبه‌ای چند-دسترسی (Multi-access Edge Computing Server)

۳. کانال بیسیم

در فصل جاری نحوه عملکرد هر کدام از این مولفه‌ها در قالب مدل‌های تئوری شرح داده می‌شود.

۱-۳ مدل وظایف

فرض می‌شود که k نوع وظیفه مختلف در سیستم رایانش لبه‌ای وجود دارد و به ازای هر نوع وظیفه دقیقاً یک صف در سیستم وجود دارد. وظایف نوع i -ام برای اجرا به صورت محلی^۱ احتیاج به L_i بازه زمانی پردازش توسط واحد پردازنده دارند و به منظور تخلیه به سرور رایانش لبه‌ای احتیاج به M_i واحد زمانی ارسال توسط واحد ارسال^۲ دارند. همچنین فرض می‌شود که وظایف نوع i -ام در سرور رایانش لبه‌ای به C_i بازه زمانی پردازش توسط سرور نیاز دارند. برای سادگی بیشتر در ادامه این مقاله برای اشاره به یک واحد زمانی اجرا توسط پردازنده از عبارت «قسمت»^۳ استفاده می‌کنیم که انتزاعی از قسمت‌های کد اجرایی است. و برای اشاره به یک واحد زمانی ارسال توسط واحد ارسال از عبارت «بسته» استفاده می‌شود.

^۱ Local

^۲ Transmission Unit

^۳ Section

۲-۳ مدل دستگاه کاربر

دستگاه کاربر مطابق با شکل ۱-۳ شامل دو مولفه پردازنده و واحد ارسال می‌باشد. همچنین همانطور که اشاره شد k صف مختلف به ازای هر کدام از انواع وظایف در سیستم وجود دارد. ظرفیت هر صف را برابر با مقدار ثابت Q در نظر می‌گیریم.

در هر بازه زمانی، واحد پردازنده یا به اندازه یک قسمت پردازش انجام می‌دهد و یا بیکار^۴ است. اجرای هر قسمت پردازش توسط پردازنده به میزان P_{loc} وات توان مصرف می‌کند. به طور مشابه واحد ارسال در هر بازه زمانی یا یک بسته را به شبکه ارسال می‌کند یا بیکار است. نکته قابل توجه در مورد واحد ارسال این است که با توجه به شرایط کانال بیسیم، در یک بازه زمانی خاص ممکن است ارسال موفقیت آمیز باشد یا نباشد. فرض می‌شود که ارسال موفقیت آمیز هر بسته به میزان P_{tx} وات توان مصرف می‌کند. توضیحات بیشتر در مورد نحوه کارکرد کانال بی‌سیم در بخش ۳-۴ آورده شده است.

با توجه به توضیحات داده شده می‌توان مدلی برای «حالت دستگاه کاربر»^۵ تعریف کرد. در [۱] برای مشخص کردن حالت دستگاه در زمان t از یک سه تایی مانند $\tau[t] = (q[t], c_T[t], c_L[t])$ استفاده شده است، که در آن $q[t]$ مشخص کننده تعداد وظایف موجود در صف وظایف، $c_T[t]$ مشخص کننده تعداد بسته ارسال شده از وظیفه تخصیص داده شده به واحد ارسال است، و $c_L[t]$ مشخص کننده تعداد قسمت اجرا شده از وظیفه تخصیص داده شده به پردازنده است. همچنین حالت $c_T[t] = 0$ معادل با بیکار بودن واحد ارسال و $c_L[t] = 0$ معادل با بیکار بودن واحد پردازنده تعریف می‌شود. به عنوان نمونه سه تایی $(4, 2, 1)$ به این معنی است که ۴ وظیفه در صف وظایف وجود دارد، واحد پردازش در حال تخلیه وظیفه‌ای است و تا کنون یک بسته از آن وظیفه را ارسال کرده و به عنوان قدم بعدی باید بسته شماره ۲ را ارسال کند. واحد پردازنده نیز در حال اجرای وظیفه‌ای به صورت محلی است و تا کنون یک قسمت از آن وظیفه را اجرا کرده است.

^۴Idle

^۵User Equipment State

با این حال مدل فوق در مسئله تخلیه وظیفه با چند نوع وظیفه قابل استفاده نیست و نیاز به تغییر دارد. ما در این مقاله برای تعیین حالت دستگاه کاربر از یک چندتایی^۶ به طول $k + 4$ مطابق با رابطه ۱-۳ استفاده می‌کنیم. در این رابطه متغیرهای $q_1[t]$ تا $q_k[t]$ تعداد وظایف موجود از هر نوع در صف مربوطه را مشخص می‌کنند. متغیرهای $c_R[t]$ و $c_L[t]$ مشابه با مقاله [۱] تعریف می‌شوند و به ترتیب وضعیت واحد ارسال و واحد پردازنده را مشخص می‌کنند. دو متغیر جدید $T_R[t]$ و $T_L[t]$ به ترتیب مشخص کننده نوع وظیفه در حال ارسال توسط واحد ارسال و نوع وظیفه در حال اجرا توسط پردازنده‌اند.

$$\tau[t] = (q_1[t], q_2[t], \dots, q_k[t], c_R[t], c_L[t], T_R[t], T_L[t]) \quad (۱-۳)$$

رابطه ۲-۳ شروط حاکم بر متغیرهای فضای حالت مسئله را عنوان می‌کند و به عبارتی توصیف‌گر فضای حالت مسئله است. (نکته: در رابطه ۲-۳ و سراسر این مقاله منظور از $\tau\{X\}$ مقدار متغیر X در حالت τ است.)

$$\begin{aligned} \forall \tau \in S, i \in \{1, 2, \dots, k\} \quad 0 \leq \tau\{q_i\} \leq Q \\ \forall \tau \in S \quad \tau\{T_L\}, \tau\{T_R\} \in \{0, 1, 2, \dots, k\} \\ \forall \tau \in \{\tau' \in S \mid \tau'\{T_R\} = 0\} \quad \tau\{C_R\} = 0 \\ \forall \tau \in \{\tau' \in S \mid \tau'\{T_R\} \neq 0\} \quad 1 \leq \tau\{C_R\} \leq M_{\tau\{T_R\}} \\ \forall \tau \in \{\tau' \in S \mid \tau'\{T_L\} = 0\} \quad \tau\{C_L\} = 0 \\ \forall \tau \in \{\tau' \in S \mid \tau'\{T_L\} \neq 0\} \quad 1 \leq \tau\{C_L\} \leq L_{\tau\{T_L\}} - 1 \end{aligned} \quad (۲-۳)$$

^۶Tuple

۳-۳ مدل زمان

وضعیت سیستم تخلیه وظیفه در فواصل زمانی^۷ با طول ثابت Δ میلی ثانیه بررسی می‌شود. برای مثال حالت دستگاه کاربر را در بازه زمانی t -ام با $\tau[t]$ مشخص می‌کنیم، و حالت دستگاه در بازه زمانی $t + 1$ را با $\tau[t + 1]$ مشخص می‌کنیم و فاصله بین این دو بازه زمانی Δ میلی ثانیه است.

بررسی زمان به صورت واحدهای گسسته به منظور ساده‌سازی مسئله و همچنین گسترش پذیری آن به شرایط محیطی مختلف صورت گرفته است. در عمل، یک مقدار قابل استفاده برای Δ طول بازه‌های زمانی شبکه دسترسی^۸ مورد نظر است. برای مثال در شبکه‌های LTE طول هر بازه زمانی ۵/۰ میلی‌ثانیه می‌باشد. [۵]

۴-۳ مدل کانال بیسیم

در این مقاله مشابه با [۱] کانال بی‌سیم را به صورت تصادفی مدل می‌کنیم^۹ یکی از دلایل اصلی برای مدل‌سازی کانال به صورت احتمالی، وجود نویز و ناپایداری در ارتباطات بیسیم است. کانال بی‌سیم را با یک مدل ساده احتمالی دوجمله‌ای مدل می‌کنیم. به این صورت که ارسال هر بسته در بازه زمانی t توسط واحد ارسال با احتمال β موفقیت آمیز خواهد بود و با احتمال $1 - \beta$ ناموفق خواهد بود. در عمل مقدار β با توجه به رابطه ۳-۳ (رابطه شنون) محاسبه می‌شود، که در آن R مشخص کننده سبب هر بسته است، $r(t)$ مشخص کننده نرخ ارسال در زمان t ، B پهنای باند سیستم، $\gamma[t]$ مقدار بهره کانال^{۱۰} و N_0 مشخص کننده اندازه نویز کانال است.

$$\beta = P(r(t) \geq R)$$

$$r(t) = B \log_r \left(1 + \frac{\gamma[t] P_{tx}}{N_0 B} \right) \quad (3-3)$$

⁷Time Slot

⁸Access Network

⁹Stochastic Channel

¹⁰Channel Gain

۳-۵ مفهوم کنش

یک استراتژی تخلیه در هر بازه زمانی مانند t می‌بایست یک کنش^{۱۱} مانند $A[t]$ را برای اجرا توسط دستگاه کاربر انتخاب کند. برای درک مفهوم کنش، ابتدا مشابه [۱] حالتی را در نظر می‌گیریم که تنها یک صف (یعنی یک نوع وظیفه) در سیستم وجود داشته باشد. در این حالت می‌توانیم مجموعه کنش‌ها را با چهار عضو مطابق جدول ۳-۱ مشخص کنیم.

ID	Transmit	Local Execution	Description
1	False	False	No operation
2	False	True	Add to CPU
3	True	False	Add to TU
4	True	True	Add to both units

جدول ۳-۱: لیست کنش‌ها در سیستم با یک صف وظیفه

به طور مشابه در حالتی که بیش از یک نوع وظیفه در سیستم وجود دارد کنش‌های ممکن مطابق با جدول ۳-۲ بدست می‌آید.

ID	Transmit	Local Execution	Description	Count
$\{1\}$	False	False	No operation	1
$\{2, \dots, k+1\}$	False	True	Add to CPU	k
$\{k+2, \dots, 2k+1\}$	True	False	Add to TU	k
$\{2k+2, \dots, 2k+k*k-1\}$	True	True	Add to both units	k^2

جدول ۳-۲: دسته‌بندی کنش‌ها در سیستم با k صف

اجرای هر کنش طبعا ممکن است که حالت سیستم را تغییر دهد. به طور مثال اجرای هر کنش نوع Add To CPU یک وظیفه را از صف مربوطه بر می‌دارد، بنابراین طول صف مطابق $q_i[t+1] = q_i[t] - 1$ تغییر می‌کند. با اجرای این کنش همچنین وضعیت پردازنده از $c_L[t] = 0$ یعنی حالت بیکار به $c_L[t+1] = 1$ تغییر خواهد کرد زیرا قسمت اول وظیفه مربوطه در بازه زمانی t انجام خواهد شد. به طور مشابه برای سایر کنش‌ها نیز میتوان توابع انتقال^{۱۲} مشخص تعریف کرد که با گرفتن یک حالت

¹¹ Action

¹² Transition Function

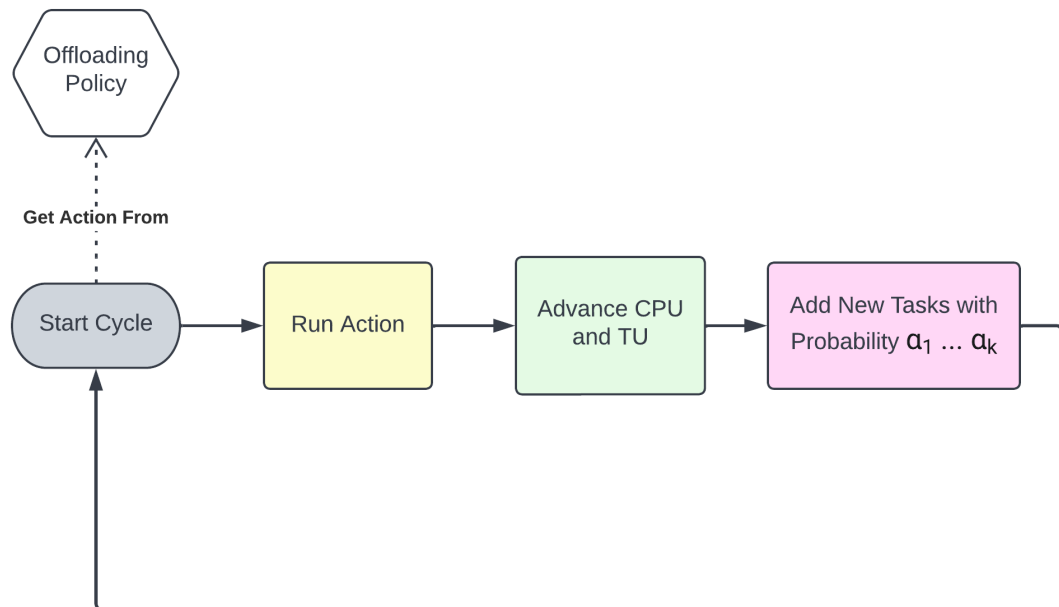
ورودی، حالت خروجی را محاسبه نماید. به دلیل پیچیدگی و حجم زیاد معادلات این توابع از توضیح بیشتر در این بخش صرف نظر شده است. برای مشاهده منطق دقیق این توابع در قالب کد، به پیوست ۱ مراجعه شود.

۳-۶ استراتژی تخلیه

استراتژی تخلیه در هر بازه زمانی تصمیم می‌گیرد که دستگاه کاربر چه کنشی را اجرا کند. بنابراین استراتژی تخلیه یک تابع مانند $G(\tau)$ می‌باشد که با گرفتن حالت دستگاه کاربر $\tau[t]$ به عنوان ورودی، یک کنش مانند a را به عنوان خروجی می‌دهد. لازم به ذکر است که در اینجا این تابع را به صورت مفهومی انتزاعی در نظر می‌گیریم و در فصل‌های آتی به طور دقیق به نحوه بدست آوردن تابع بهینه $g(\tau)^*$ خواهیم پرداخت.

۷-۳ روند فعالیت سیستم تخلیه وظایف

نحوه عملکرد دستگاه کاربر در هر بازه زمانی مطابق با فرآیند مشخص شده در شکل ۲-۳ می باشد. در هر بازه، دستگاه کاربر ابتدا کنش اجرایی را از یک استراتژی تخلیه دریافت می کند. سپس کنش انتخاب شده توسط دستگاه کاربر اجرا خواهد شد که ممکن است منجر به تغییر حالت دستگاه شود. سپس پردازنده و واحد ارسال هر کدام در صورت فعال بودن به اندازه یک بازه زمانی فعالیت خواهند کرد. در انتها وظایف جدید با احتمالات $\alpha_1, \dots, \alpha_k$ به صف های وظایف اضافه خواهند شد.



شکل ۲-۳: روند فعالیت دستگاه کاربر

فصل ۴

روش پیشنهادی

۱-۴ استراتژی تخلیه تصادفی

با استفاده از مدل‌های توصیف شده در بخش‌های قبل، حال می‌توانیم یک تعریف ریاضی از «استراتژی تخلیه تصادفی» داشته باشیم. مشابه با مقاله [۱] یک استراتژی تخلیه تصادفی را به صورت توزیع احتمالی مانند g_T^a بر روی مجموعه $S \times A$ تعریف می‌کنیم. در اینجا عبارت $S \times A$ نمایانگر ضرب دکارتی مجموعه تمام حالت‌های سیستم در مجموعه تمام کنش‌های ممکن در سیستم است. یک نکته قابل توجه این است که برخی از دو تایی‌های حاصل از این ضرب دکارتی هیچ گاه در واقعیت امکان‌پذیر نیست. برای مثال در حالتی که صف خالی باشد تنها یک کنش امکان‌پذیر است و آن هم کنش شماره ۱ (No Operation) است. با این حال برای سادگی در توضیح تئوری روش حل مسئله، این دو تایی‌ها را نیز در دامنه تابع توزیع احتمالی استراتژی تخلیه در نظر می‌گیریم تا همواره اندازه دامنه تابع احتمال برابر با مقدار ثابت $|S| \cdot |A|$ باشد اما در پیاده‌سازی عملی چنین دوتایی‌هایی از دامنه حذف می‌شوند و مقدار احتمال ثابتی برابر صفر می‌گیرند تا با کاهش فضای حالت مسئله، سرعت الگوریتم اجرایی بهبود یابد.

همچنین طبق تعریف توزیع احتمال، رابطه ۴-۱ باید برای هر استراتژی تخلیه تصادفی برقرار باشد.

$$\sum_{\tau \in S} \sum_{a \in A} g_{\tau}^a = 1 \quad (4-1)$$

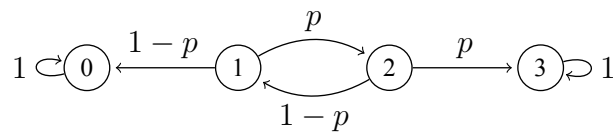
۲-۴ مدل زنجیره مارکوف دستگاه کاربر

در این قسمت ابتدا مدل آماری زنجیره مارکوف گسسته-زمان را معرفی می‌کنیم و سپس توضیح می‌دهیم که چگونه می‌توان با استفاده از این مدل معیارهای تاخیر و توان میانگین را برای یک سیستم تخلیه وظیفه محاسبه کرد.

تعریف ۱.۴. دنباله‌ای از متغیرهای تصادفی X_1, X_2, \dots را که احتمال تغییر وضعیت از زمان t به $t+1$ مستقل از وضعیت‌های قبلی باشد را یک **زنجیره مارکوف گسسته-زمان** می‌نامند. این گزاره را به بیان متغیرهای تصادفی و تابع احتمال به صورت رابطه زیر نشان می‌دهیم.

$$\Pr(X_{t+1} = x \mid X_1 = x_1, X_2 = x_2, \dots, X_n = x_t) = \Pr(X_{t+1} = x \mid X_t = x_t)$$

زنجیره مارکوف گسسته-زمان را می‌توان با گراف جهت‌دار نیز نمایش داد. در شکل ۴-۱ یک زنجیره نمونه مشاهده می‌شود.



شکل ۴-۱: یک زنجیره مارکوف نمونه برای مسئله پاکبختی قمارباز^۱

^۱The Gambler's ruin

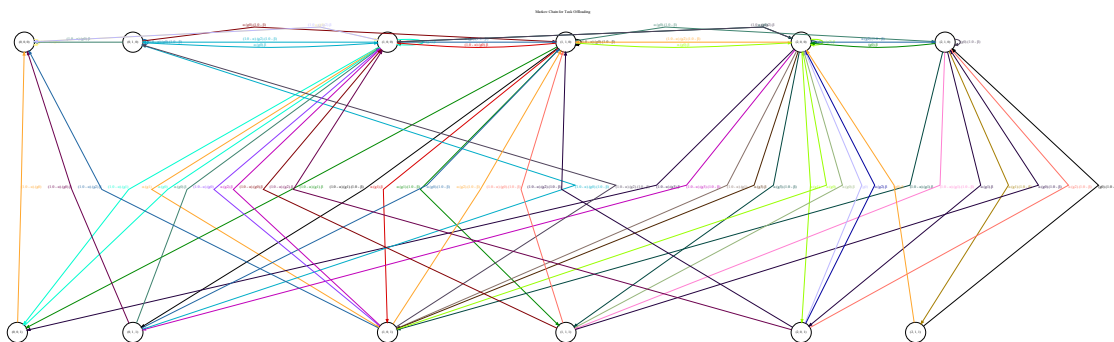
تعریف ۲.۴. زنجیره مارکوف گسسته زمان $X(t)$ را همگن-زمان می‌گوییم اگر شرط زیر همواره برقرار باشد:

$$P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i)$$

به عبارت دیگر یعنی احتمالات مربوط به انتقال بین حالت‌ها به زمان t وابسته نیستند. در این حالت احتمال انتقال زنجیره از حالت i به j را با عبارت $p_{ij} = P(X_1 = j | X_0 = i)$ نمایش می‌دهیم و همچنین ماتریس انتقال را با $P = (p_{ij})$ نمایش می‌دهیم. ماتریس انتقال را می‌توان به صورت یک گراف جهت‌دار نیز توصیف کرد به طوری که درایه $p_{i,j}$ در ماتریس معادل یک یال جهت‌دار از راس i به راس j با وزن $p_{i,j}$ است.

طبق تعاریف ۱.۴ و ۲.۴ می‌توان زنجیره مارکوف گسسته‌زمانی برای حالت دستگاه کاربر در طی زمان در نظر گرفت که در آن حالت دستگاه کاربر $\tau[t]$ حالت زنجیره در زمان t را مشخص می‌کند. همچنین ماتریس انتقال χ تعریف می‌شود به طوری که $\chi_{\tau, \tau'}$ احتمال انتقال از حالت τ به τ' را مشخص می‌کند.

ماتریس انتقال را می‌توان به ازای یک استراتژی تخلیه داده شده و پارامترهای سیستمی مانند $\alpha_1, \dots, \alpha_k$ بدست آورد. در شکل ۲-۴ گراف جهت‌دار معادل زنجیره مارکوف برای یک سیستم با یک صف وظیفه و $Q = 2$ و تعداد ۲ قسمت به ازای هر وظیفه و تعداد ۱ بسته به ازای هر وظیفه رسم شده است.^۲



شکل ۲-۴: زنجیره مارکوف سیستم تخلیه در قالب گراف جهت‌دار (برای مشاهده جزئیات زوم کنید)

^۲ کد استفاده شده برای رسم این گراف در آدرس <https://github.com/dalisyron/OffloadingVisualizer> موجود می‌باشد

به منظور محاسبه معیارهایی مانند توان مصرفی میانگین و تاخیر سرویس میانگین لازم است که بتوانیم درباره وضعیت سیستم تخلیه وظیفه در طولانی مدت استنتاج کنیم. در همین راستا مفهوم توزیع پایدار را تعریف می‌کنیم.

تعریف ۳.۴. به ازای یک زنجیره مارکوف داده شده با ماتریس انتقال P یک **توزیع پایدار** توزیع احتمالی مانند π است که شرط زیر برای آن برقرار باشد:

$$\pi = \pi P \iff \pi_j = \sum_i \pi_i P_{ij} \quad \forall j.$$

یک سوالی که ممکن است بوجود بیاید این است که آیا لزوماً هر زنجیره مارکوف گسسته زمانی توزیع پایدار دارد؟ برای پاسخ به این سوال لازم است دو مفهوم زنجیره مارکوف تقلیل‌ناپذیر و غیرمتناوب را تعریف کنیم.

تعریف ۴.۴. اگر رسیدن از هر نقطه به نقطه دیگر از فضای حالت با احتمال مثبت در زنجیره مارکوف میسر باشد، زنجیره را **تقلیل‌ناپذیر** گویند. به بیان ریاضی می‌توان تقلیل‌ناپذیر بودن زنجیره مارکوف را به صورت زیر نشان داد.

$$\Pr(X_{n_{ij}} = j \mid X_0 = i) = p_{ij}^{(n_{ij})} > 0$$

تعریف ۵.۴. تناوب $d(i)$ برای حالت i به صورت $d(i) = \gcd\{n : P_{ii}^n > 0\}$ تعریف می‌شود، که به معنی بزرگ‌ترین مقسوم علیه مشترک تعداد مراحل ممکن است به صورتی که از i شروع کرده و به i برگردیم. یک زنجیره مارکوف تقلیل‌ناپذیر را متناوب با تناوب d می‌گوییم اگر تمامی حالت‌ها تناوب $d > 1$ را داشته باشند. یک زنجیره مارکوف تقلیل‌ناپذیر را **غیرمتناوب** می‌گوییم اگر تمامی حالت‌ها تناوب برابر ۱ داشته باشند.

قضیه ۱.۴ (همگرایی) هر زنجیره مارکوف تقلیل ناپذیر و غیر متناوب دارای توزیع پایدار منحصر به فردی مانند π می باشد.

حال با استفاده از قضیه ۱.۳ ثابت می کنیم که زنجیره مارکوف مربوط به سیستم تخلیه وظیفه دارای توزیع پایدار منحصر به فرد است. برای سادگی فرض می کنیم که سامانه یک صف دارد و سپس نحوه بسط نتیجه به چندین صف را توضیح می دهیم.

قضیه ۲.۴. زنجیره مارکوف مربوط به سیستم تخلیه تک صف تقلیل ناپذیر است.

اثبات:

قسمت الف) با توجه به تعریف سیستم تخلیه می دانیم که از هر حالت غیر شروع مانند $(0, 0, 0) \neq (x, y, z)$ می توان به حالت شروع رفت. به این منظور کافی است که تمام وظایف داخل صف به نحوی (اجرا یا ارسال) به اتمام برسند و وظیفه جدیدی نیز در این حین وارد سیستم نشود.

قسمت ب) همچنین می توان ثابت کرد که از حالت شروع $(0, 0, 0)$ می توان به هر حالت دیگر (x, y, z) رفت. به این منظور دنباله رخدادهای زیر را در نظر بگیرید:

۱. ورود x وظیفه جدید

۲. انتقال یک وظیفه به واحد ارسال و ورود یک وظیفه جدید، هر دو در صورتی که $y > 0$

۳. پیشرفت واحد ارسال به مدت y سیکل و عدم ورود وظیفه جدید در این حین

۴. انتقال یک وظیفه به پردازنده و ورود یک وظیفه جدید، هر دو در صورتی که $z > 0$

۵. پیشرفت واحد ارسال به مدت z سیکل و عدم ورود وظیفه جدید در این حین

با توجه به نتایج بخش الف و ب می توان نتیجه گرفت که از گشتی با احتمال مثبت از هر حالت به حالت دیگر وجود دارد بنابراین طبق تعریف زنجیره تقلیل ناپذیر است.

قضیه ۳.۴. زنجیره مارکوف مربوط به سیستم تخلیه تک صف غیر متناوب است.

اثبات:

به منظور اثبات این قضیه فقط کافی است که به این نکته توجه کنیم که در زنجیره حالت $(0, 0, 0)$ دارای تناوب یک می باشد. زیرا با احتمالی مثبت (متناظر با رخداد عدم ورود وظیفه و کنش No Operation) می توان در همان حالت ماند. با توجه به همین نکته و تقلیل ناپذیر بودن زنجیره میتوانیم نتیجه بگیریم که سایر حالت ها نیز باید تناوب یک داشته باشند. بنابراین زنجیره غیرمتناوب است.

با توجه به قضایای ۲.۴ و ۳.۴ و قضیه همگرایی می توان نتیجه گرفت که زنجیره مارکوف سیستم تخلیه تک صف دارای توزیع پایدار منحصر به فرد می مطابق با رابطه ۲-۴ می باشد. برای بسط این اثبات به حالت چند صف اثبات غیرمتناوب بودن یکسان خواهد بود و در اثبات تقلیل ناپذیر بودن، رخداد اول به ورود x_1, \dots, x_k وظیفه از انواع مختلف تغییر پیدا می کند.

$$\begin{cases} \sum_{\tau' \in \mathcal{S}} \chi_{\tau', \tau} \pi_{\tau'} = \pi_{\tau}, \forall \tau \in \mathcal{S} \\ \sum_{\tau \in \mathcal{S}} \pi_{\tau} = 1 \end{cases} \quad (۲-۴)$$

۳-۴ محاسبه تاخیر میانگین

تأخیر هر وظیفه شامل تأخیر انتظار در صف وظایف و تأخیر پردازش می باشد. به منظور بدست آوردن تأخیر میانگین سیستم ابتدا θ_i را به عنوان کسری از وظایف سیستم در طولانی مدت که از نوع i هستند تعریف می کنیم. اگر طول صف ها به مقدار کافی بزرگ باشد و همچنین استراتژی تخلیه ای داشته باشیم که منجر به پر شدن صف و اتلاف وظیفه^۳ نشود مقدار θ_i طبق رابطه ۳-۴ بدست می آید.

$$\theta_i = \frac{\alpha_i}{\sum_{j=1}^k \alpha_j} \quad (۳-۴)$$

^۳Task Loss

پارامتر t_q^i را برابر با مقدار میانگین تاخیر انتظار در صف مربوط به وظایف نوع i تعریف می‌کنیم. طبق قانون Little می‌توان مقدار این تاخیر را بر اساس رابطه ۴-۴ بدست آورد. همانطور که پیش‌تر ذکر شد برای برقراری این رابطه لازم است که اتلاف وظیفه در صف هیچ‌گاه رخ ندهد. به عبارت دیگر با فرض اینکه استراتژی تخلیه‌ی ارائه شده «کارآمد» باشد این رابطه برقرار است. در پیاده‌سازی عملی، محدودیت «کارآمد» بودن یک استراتژی بدین گونه تعریف شده است که احتمال پر بودن صف حداکثر مقداری ناچیز باشد.

$$t_q^i = \frac{\theta_i}{\alpha_i} \sum_{j=0}^Q i \cdot \Pr\{q_i[t] = i\} = \frac{1}{\alpha} \sum_{\tau \in S} \tau\{q_i\} \cdot \pi_\tau \quad (۴-۴)$$

همچنین t_{tx}^i را به عنوان تاخیر ارسال میانگین یک وظیفه از نوع i توسط واحد ارسال تعریف می‌کنیم که مقدار آن بر اساس امید ریاضی موفقیت در فرآیند برنولی مطابق با رابطه ۴-۵ بدست می‌آید.

$$t_{tx}^i = M_i \sum_{j=1}^{\infty} j(1-\beta)^{(j-1)}\beta \quad (۵-۴)$$

به یاد داریم که مقدار تاخیر در صورت پردازش محلی برای وظایف نوع i برابر L_i می‌باشد. تاخیر اجرا در صورت تخلیه وظیفه به صورت مجموع زمان ارسال وظیفه t_{tx}^i زمان اجرا در سرور لبه‌ای C_i و تاخیر دریافت نتیجه از سرور t_{rx}^i محاسبه می‌شود.

$$t_c^i = t_{tx}^i + C_i + t_{rx}^i \quad (۶-۴)$$

در نتیجه می‌توان تاخیر اجرای میانگین وظایف نوع i را نیز مطابق رابطه ۴-۷ بیان کرد.

$$t_p^i = \eta_i L_i + (1 - \eta_i) t_c^i \quad (۷-۴)$$

که در آن η_i بیانگر کسری از وظایف نوع i می‌باشد که در طولانی‌مدت به صورت محلی اجرا می‌شوند

و مطابق با رابطه ۸-۴ بدست می آید.

$$\eta_i = \frac{\sum_{\tau, a \in S_1^i \cup S_3^i \cup S_5^i} \pi_{\tau} g_{\tau}^a}{\sum_{\tau, a \in S_1^i \cup S_2^i \cup S_3^i \cup S_4^i} \pi_{\tau} g_{\tau}^a + 2 \sum_{\tau, a \in S_5^i} \pi_{\tau} g_{\tau}^a} \quad (۸-۴)$$

که در آن S_1^i, \dots, S_5^i به صورت زیر تعریف می شوند:

$$(۹-۴)$$

$$S_1^i = \{\tau, a \in \mathcal{S} \times A | type(a) = AddToCPU \wedge locType(a) = i\}$$

$$S_2^i = \{\tau, a \in \mathcal{S} \times A | type(a) = AddToTU \wedge ofloadType(a) = i\}$$

$$S_3^i = \{\tau, a \in \mathcal{S} \times A | type(a) = AddToBoth \wedge locType(a) = i \wedge ofloadType(a) \neq i\}$$

$$S_4^i = \{\tau, a \in \mathcal{S} \times A | type(a) = AddToBoth \wedge locType(a) \neq i \wedge ofloadType(a) = i\}$$

$$S_5^i = \{\tau, a \in \mathcal{S} \times A | type(a) = AddToBoth \wedge locType(a) = i \wedge ofloadType(a) = i\}$$

در رابطه فوق تابع $type(a)$ نوع کنش را مشخص می کند و یکی از چهار نوع بیان شده در بخش ۳-۵ می باشد. توابع $locType(a)$ و $ofloadType(a)$ نیز نوع وظیفه مربوط به کنش a را مشخص می کنند.

با استفاده از روابط بالا همچنین می توانیم میانگین تاخیر سرویس هر وظیفه در سیستم را طبق رابطه ۴-۱۰ محاسبه کنیم. نکته مهم این است که این تابع همان تابع هدف برای کمینه سازی در مسئله پیدا کردن استراتژی تخلیه بهینه است.

$$\bar{T} = \sum_{i=1}^k \theta_i (t_q^i + t_p^i) \quad (۱۰-۴)$$

۴-۴ توان مصرفی میانگین

اگر پارامتر μ_{τ}^{loc} و μ_{τ}^{tx} را که به ترتیب به عنوان احتمال فعالیت پردازنده در حالت τ و احتمال فعالیت واحد ارسال در حالت τ تعریف کنیم، آنگاه توان مصرفی میانگین طبق رابطه زیر بدست می آید:

$$\bar{P} = \sum_{\tau \in S} \pi_{\tau} (\mu_{\tau}^{loc} P_{loc} + \mu_{\tau}^{tx} P_{tx}) \quad (۱۱-۴)$$

۵-۴ استراتژی تخلیه وظیفه بهینه

با توجه به توابع بدست آمده برای تاخیر و توان مصرفی میانگین در بخش های پیشین، حال می توانیم مسئله پیدا کردن استراتژی تخلیه بهینه را به صورت یک مسئله بهینه سازی مانند P_1 بیان کنیم:

$$\begin{aligned} \mathcal{P}_1 : \min_{\{g_{\tau}^a\}} \bar{T} &= \left(\sum_{i=1}^k \frac{1}{\alpha_i} \sum_{\tau \in S} \tau\{q_i\} \cdot \pi_{\tau} \right) + T_p^0 \\ \text{s.t.} \quad &\begin{cases} \bar{P} \leq \bar{P}_{\max} \\ \sum_{\tau' \in S} \chi_{\tau', \tau} \pi_{\tau'} = \pi_{\tau}, \tau \in S, \\ \sum_{\tau \in S} \pi_{\tau} = 1, \\ \sum_{\alpha \in A} g_{\tau}^{\alpha} = 1, \forall \tau \in S \\ g_{\tau}^a \geq 0, \forall \tau \in S, a \in A \end{cases} \end{aligned} \quad (۱۲-۴)$$

که در آن T_p^0 برابر با تاخیر اجرای میانگین است که به ازای مقادیر داده شده از η_0, \dots, η_k مقداری ثابت دارد و از رابطه زیر بدست می آید:

$$T_p^0 = \sum_{i=1}^k (\eta_i L_i + (1 - \eta_i) t_c^i) \quad (۱۳-۴)$$

با توجه به وجود پارامتر η_i در تابع هدف \mathcal{P}_1 این مسئله بهینه‌سازی، یک مسئله خطی نیست. با این حال می‌توانیم با استفاده از تغییری کوچک مسئله را به دنباله‌ای از مسائل برنامه‌ریزی خطی تبدیل کنیم. به این منظور مشابه [۱] از تعریف «معیار احاطه^۴» در زنجیره مارکوف استفاده می‌کنیم. به طور دقیق‌تر مجموعه متغیرهای جدید $\{x_\tau^a\}$ را تعریف می‌کنیم به طوری که $x_\tau^a = \pi_\tau g_\tau^a$ تعریف می‌کنیم. به عبارتی x_τ^a برابر با احتمال حضور در حالت τ و انتخاب کنش a می‌باشد. همچنین طبق

$$\text{تعریف } 1 \sum_{a \in A} g_\tau^a = 1 \text{ و بنابراین } \pi_\tau = \sum_{a \in A} x_\tau^a$$

حال با جایگذاری $\{x_\tau^a\}$ به جای $\{\pi_\tau\}$ در \mathcal{P}_1 خواهیم داشت:

$$\mathcal{P}_2 : \min_{\mathbf{x}, \boldsymbol{\eta}} \bar{T} = \left(\sum_{i=1}^k \frac{1}{\alpha_i} \sum_{\tau \in \mathcal{S}} \sum_{a \in A} \tau \{q_i\} \cdot x_\tau^a \right) + T_p^0$$

$$\text{s.t.} \begin{cases} \nu_{loc}(\mathbf{x}) P_{loc} + \beta \nu_{tx}(\mathbf{x}) P_{tx} \leq \bar{P}_{\max} \\ \Gamma(\mathbf{x}, \eta_i) =, \forall i \in \{1, \dots, k\} \\ F_\tau(\mathbf{x}) = 0, \forall \tau = (i, m, n) \in \mathcal{S} \\ \sum_{\tau \in \mathcal{S}} \sum_{a \in A} x_\tau^a = 1 \\ \eta_i \in [0, 1], \forall i \in \{1, \dots, k\} \\ x_\tau^a \geq 0, \forall \tau \in \mathcal{S}, a \in A \end{cases} \quad (14-4)$$

که در آن ν_{tx} و ν_{loc} به ترتیب احتمال فعالیت پردازنده و واحد ارسال را در یک واحد زمانی دلخواه مشخص می‌کنند و به ازای یک استراتژی داده شده قابل محاسبه اند.^۵ تابع $\Gamma(\mathbf{x}, \eta_i)$ بر اساس رابطه ۸-۴ می‌باشد و به صورت زیر محاسبه می‌شود:

$$\Gamma(\mathbf{x}, \eta) = \eta \sum_{\tau, \mathbf{a} \in \mathcal{S}_1^i \cup \mathcal{S}_2^i \cup \mathcal{S}_3^i \cup \mathcal{S}_4^i} x_\tau^a + 2\eta \sum_{\tau, \mathbf{a} \in \mathcal{S}_5^i} x_\tau^a - \eta \sum_{\tau, \mathbf{a} \in \mathcal{S}_1^i \cup \mathcal{S}_3^i \cup \mathcal{S}_5^i} x_\tau^a \quad (15-4)$$

⁴Occupation Measure

^۵برای مشاهده روش محاسبه این دو پارامتر در قالب کد به پیوست ۲ مراجعه شود.

و تابع $F_\tau(x)$ به صورت زیر تعریف می‌شود:

$$F_\tau(x) = \sum_{\tau' \in S} \sum_{a \in A} \tilde{\chi}_{\tau', \tau, a} x_{\tau'}^a - \sum_{a \in A} x_\tau^a \quad (۱۶-۴)$$

در رابطه فوق منظور از $\tilde{\chi}_{\tau', \tau, a}$ احتمال این است که به شرط اینکه در حالت τ' باشیم و کنش a انتخاب شده باشد، آنگاه به حالت τ' برویم.

در صورتی که مقادیر η_0, \dots, η_k معلوم باشد آنگاه مسئله \mathcal{P}_2 تبدیل به یک مسئله برنامه‌ریزی خطی می‌شود. با یافتن مقادیر جواب بهینه $\{x_\tau^a\}$ می‌توان استراتژی بهینه را طبق رابطه زیر بدست آورد:

$$g_\tau^{a*} = \frac{x_\tau^{a*}}{\sum_{a \in A} x_\tau^{a*}}, \forall \tau \in S, a \in A \quad (۱۷-۴)$$

بنابراین جهت یافتن استراتژی بهینه برای یک سیستم تخلیه وظیفه کافی است که مسئله برنامه‌ریزی خطی حاصل از \mathcal{P}_2 را به ازای مقادیر مختلف η_0, \dots, η_k حل کرده تا استراتژی بهینه بدست بیاید. مراحل این فرآیند جستجو در الگوریتم ۱ به صورت خلاصه آمده است.

الگوریتم ۱.۴ جستجوی استراتژی تخلیه وظیفه بهینه

Require: $precision \geq 2$

- 1: $\etaSettings \leftarrow splitRange([0, 1], precision)^k$
 - 2: $optimalPolicy = null$
 - 3: **for each** $s \in \etaSettings$ **do**
 - 4: $(\eta_0, \dots, \eta_k) \leftarrow s$
 - 5: $solution \leftarrow solveLP(\eta_0, \dots, \eta_k)$
 - 6: **if** $optimalPolicy = null$ **or** $solution.delay < optimalPolicy.delay$ **then**
 - 7: $optimalPolicy \leftarrow solution.policy$
 - 8: **end if**
 - 9: **end for**
 - 10: **return** $optimalPolicy$
-

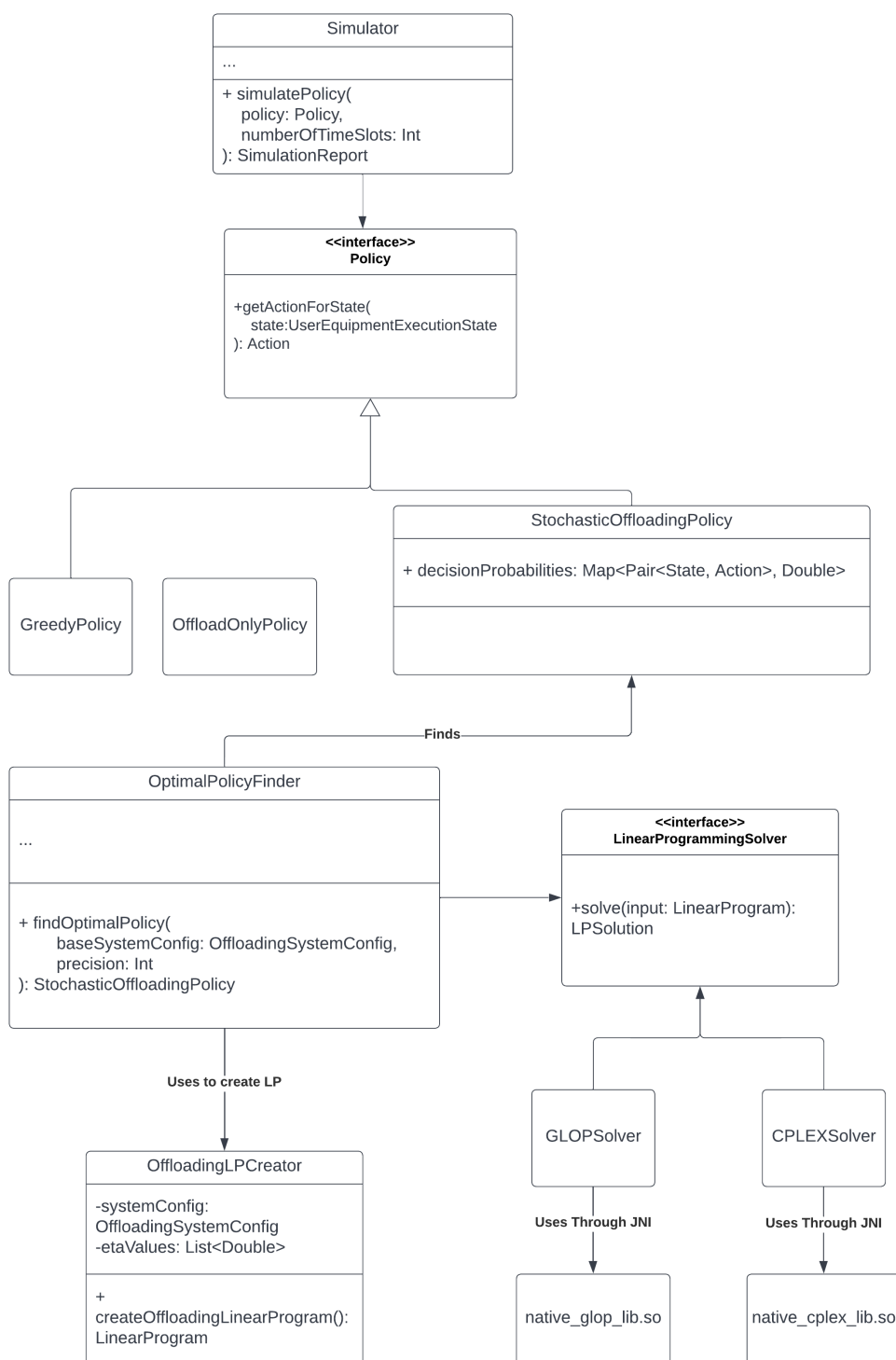
فصل ۵

آزمایش و نتایج

۵-۱ ساختار نرم‌افزاری Kompute

به منظور تست الگوریتمی که در بخش پیشین ارائه گردید، یک ساختار نرم‌افزاری (فریم‌ورک) با نام Kompute در زبان Kotlin تعبیه شده است. با استفاده از این فریم‌ورک می‌توان الگوریتم جستجوی استراتژی تخلیه بهینه را به ازای پارامترهای محیطی مختلف اجرا کرد و عملکرد استراتژی بدست آمده از الگوریتم را با کمک شبیه‌سازی با سایر استراتژی‌ها مقایسه کرد. این فریم‌ورک طبق یافته‌های ما اولین پیاده‌سازی متن‌باز در زمینه استراتژی تخلیه وظایف ناهمگون است.

معماری کلی این فریم‌ورک در قالب یک کلاس دیاگرام در صفحه بعد آورده شده است.



شکل ۵-۱: کلاس دیاگرام فریم‌ورک Kompute

۱-۱-۵ ساخت و حل یک مسئله تخلیه وظیفه نمونه در Kompute

در کد نمونه زیر مسئله تخلیه وظیفه برای محیط پردازش لبه‌ای با دو صف^۱ حل شده است.

```
fun main(args: Array<String>) {
    val systemConfig = OffloadingSystemConfig(
        userEquipmentConfig = UserEquipmentConfig(
            stateConfig = UserEquipmentStateConfig(
                taskQueueCapacity = 5,
                tuNumberOfPackets = listOf(1, 3),
                cpuNumberOfSections = listOf(7, 2),
                numberOfQueues = 2
            ),
            componentsConfig = UserEquipmentComponentsConfig(
                alpha = listOf(0.4, 0.9),
                beta = 0.90,
                etaConfig = null,
                pTx = 1.0,
                pLocal = 0.8,
                pMax = 1.7
            )
        ),
        environmentParameters = EnvironmentParameters(
            nCloud = listOf(1, 1),
            tRx = 0.5,
        )
    )

    val optimalPolicy = RangedOptimalPolicyFinder.findOptimalPolicy(
        baseSystemConfig = systemConfig,
        precision = 10
    )
    /*
    // For multi-threaded execution use this instead:

    val optimalPolicy = ConcurrentRangedOptimalPolicyFinder(
        baseSystemConfig = systemConfig
    ).findOptimalPolicy(precision = 10, numberOfThreads = 8)
    */

    val decisionProbabilities: Map<StateAction, Double>
    = optimalPolicy.stochasticPolicyConfig.decisionProbabilities

    println(decisionProbabilities)
}
```

^۱شرایط بر اساس تقسیم‌بندی وظایف به Heavy و Light در اینترنت اشیا

۲-۵ دو بهینه‌سازی برای الگوریتم جستجوی استراتژی

در این قسمت دو بهینه‌سازی مختلف را به منظور بهبود عملکرد الگوریتم ۱ معرفی می‌کنیم.

۱-۲-۵ کاهش تعداد متغیرها

در مسئله بهینه‌سازی P_2 که در فصل قبل ارائه شد (رابطه ۴-۱۴) تعداد $|A| \cdot |S|$ متغیر وجود دارد. این مقدار برای تعداد صف‌های کم (برای مثال $k \leq 4$) قابل اجرا می‌باشد اما با افزایش تعداد صف‌ها اجرای الگوریتم را بسیار زمان‌بر و یا غیرممکن می‌کند. یک بهینه‌سازی خیلی ساده ولی کارآمد که در [۱] به آن اشاره‌ای نشده است این است که می‌توان تمام متغیرهای مانند x_τ^a که کنش a جزو کنش‌های ممکن در τ نباشد را حذف کرد زیرا مقدار آنها در جواب مسئله همواره برابر صفر می‌باشد. این بهینه‌سازی در فریم‌ورک Kompute اعمال شده است.

۲-۲-۵ موازی‌سازی

الگوریتم ۱.۴ به گونه‌ای تعریف شده است که امکان موازی‌سازی و مقیاس‌پذیری^۲ آن به صورت خطی وجود دارد. به عبارت دیگر می‌توان مسئله برنامه‌ریزی خطی متناظر با هر مقداردهی از η_0, \dots, η_k را به یک هسته/گره پردازشی خاص اختصاص داد. در فریم‌ورک Kompute این بهینه‌سازی نیز پیاده‌سازی شده است و هنگام اجرا بر روی سرور ابرآروان با ۲۴ هسته و تقسیم‌بندی به ۲۴ رشته^۳ عملکردی معادل ۲۰ برابر سریع‌تر از حالت تک‌رشته^۴ داشت.

^۲Scaling

^۳Thread

^۴Single-thread

۳-۵ نتایج شبیه‌سازی

در این بخش عملکرد استراتژی یافت شده توسط الگوریتم ۱۰۴ را با چهار الگوریتم پایه زیر مقایسه می‌کنیم:

۱. استراتژی «فقط تخلیه»^۵ که همه‌ی وظایف را تخلیه می‌کند

۲. استراتژی «حریصانه، تخلیه اول»^۶ که در هر بازه زمانی اگر واحد ارسال یا پردازنده بیکار باشند به هر کدام از آنها یک وظیفه از صفی رندوم تخصیص می‌دهد و در صورتی که تنها یک وظیفه در صف باشد و مجبور به انتخاب بین تخلیه و اجرای محلی باشد، تخلیه را انتخاب می‌کند.

۳. استراتژی «حریصانه، محلی اول»^۷ که در هر بازه زمانی اگر واحد ارسال یا پردازنده بیکار باشند به هر کدام از آنها یک وظیفه از صفی رندوم تخصیص می‌دهد و در صورتی که تنها یک وظیفه در صف باشد و مجبور به انتخاب بین تخلیه و اجرای محلی باشد، اجرای محلی را انتخاب می‌کند.

۴. استراتژی «فقط (اجرای) محلی»^۸

^۵Offload Only

^۶Greedy (Offload First)

^۷Greedy (Local First)

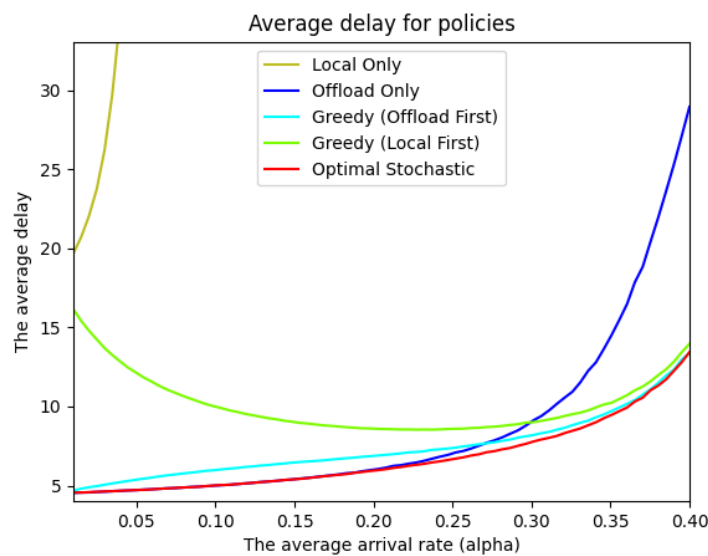
^۸Local Only

۱-۳-۵ شبیه‌سازی تک صف

با توجه به اینکه روش ارائه شده توسط ما حالت گسترش یافته [۱] است، ابتدا محیط تست ارائه شده در آن پژوهش را برای تست الگوریتم در نظر می‌گیریم. پارامترهای این محیط در جدول ۱-۵ خلاصه شده‌اند. نتیجه این آزمایش در شکل ۲-۵ مشاهده می‌شود.

پارامتر	P_1	L_1	β	P_{tx}	P_{loc}	P_{max}	C_1	t_{rx}
مقدار	۱	۱۷	۴.۰	۰.۱	۸.۰	۶.۱	۱	۰.۰

جدول ۱-۵: پارامترهای محیط پردازش لبه‌ای در سناریو تک صف

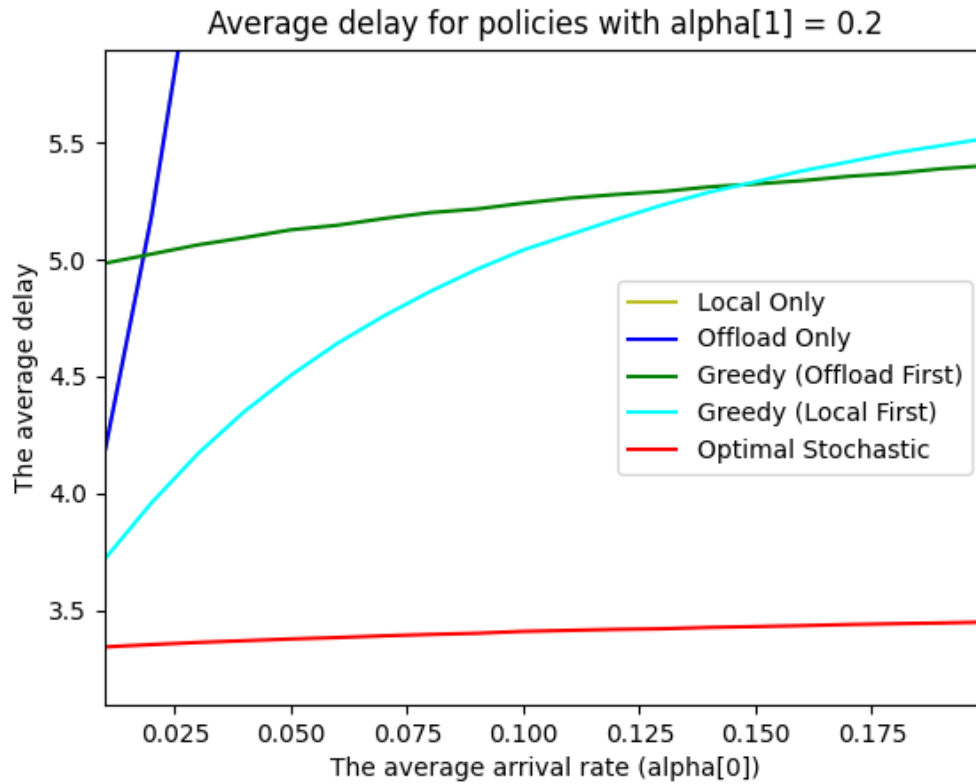


شکل ۲-۵: تاخیر سرویس به ازای نرخ ورود در حالت تک صف

همانطور که مشاهده می‌شود استراتژی تخلیه تصادفی یافت شده از تمام الگوریتم‌های پایه بهتر عمل می‌کند و شکل منحنی‌های نمودار با [۱] مطابقت دارد.

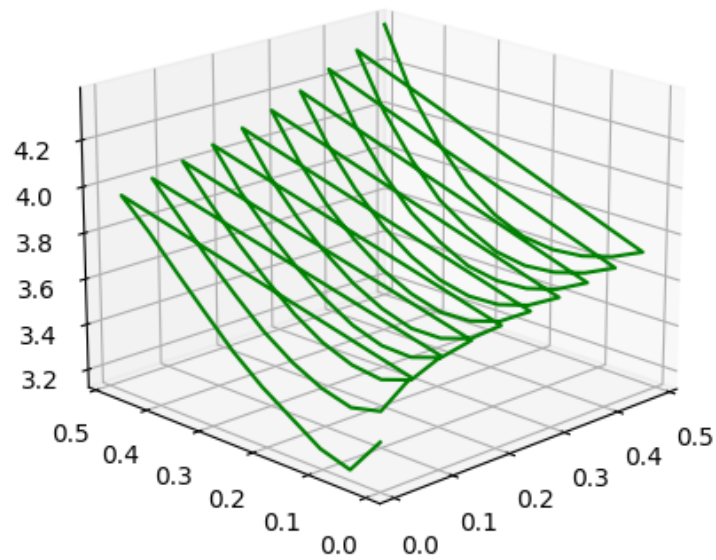
شبیه‌سازی دو صف با یک صف ثابت

در این قسمت سناریوی تست به این گونه است که به ازای مقادیر مختلف نرخ ورود برای صف شماره یک و مقدار ثابت صف شماره دو مشاهده می‌شود.



شبیه‌سازی دو صف متغیر

Average delay vs α_1 and α_2



تست کارآمدی

Policy	Optimal	Local Only	Greedy (Local First)	Greedy (Offload First)	Offload Only
Value	100.0	23.33	72.49	92.95	74.62

فصل ۶

جمع‌بندی و پیشنهادها

در این پژوهش روشی برای بدست آوردن استراتژی تخلیه وظیفه با تاخیر کمینه در شرایط حضور چندین نوع وظیفه در محیط رایانش لبه‌ای معرفی شد. عملکرد این الگوریتم نیز با کمک شبیه‌سازی بررسی شد. در طول انجام این پژوهش ایده‌های مختلفی برای بهبود روش ارائه شده به ذهن ما رسید که برخی از آنها مانند بهینه‌سازی‌های معرفی شده در بخش ۵-۲ پیاده‌سازی شدند. اما برخی از این ایده‌ها به مرحله پیاده‌سازی نرسیدند و پژوهش درباره آنها امکان بهبود روش فعلی را فراهم خواهد کرد.

یکی از این موارد کاهش تعداد متغیرهای مسئله برنامه‌ریزی خطی P_2 (رابطه ۴-۱۴) با استفاده از حذف «تک کنش» ها می‌باشد. پیشتر در بخش ۵-۲-۱ به این موضوع اشاره شد که می‌توان متغیرهایی که متناظر با کنش‌های غیر ممکن هستند را از مسئله بهینه‌سازی P_2 حذف کرد. با استدلالی مشابه این امکان وجود دارد که متغیرهایی که متناظر با تنها کنش ممکن در یک حالت هستند را از الگوریتم حذف کرد، زیرا مقدار این متغیرها در تعیین استراتژی بهینه نقشی ندارد چون احتمال انتخاب آنها همواره ۱ (قطعی) می‌باشد. با این حال حذف این متغیرها بر خلاف بهینه‌سازی ۵-۲-۱ ساختار زنجیره مارکوف را دگرگون خواهد، بنابراین احتمالاً نیازمند تغییر توابع انتقال و/یا تغییر شروط ۴-۱۴ خواهد شد.

پیوست ۱ – توابع انتقال حالت

تابع انتقال حالت به ازای کنش ورودی

```
fun getNextStateRunningAction(  
    sourceState: UserEquipmentState,  
    action: Action  
): UserEquipmentState {  
    return when (action) {  
        is Action.NoOperation → {  
            sourceState  
        }  
        is Action.AddToCPU → {  
            getNextStateAddingToCPU(sourceState, action.queueIndex)  
        }  
        is Action.AddToTransmissionUnit → {  
            getNextStateAddingToTU(sourceState, action.queueIndex)  
        }  
        is Action.AddToBothUnits → {  
            getNextStateAddingToBothUnits(  
                sourceState,  
                action.cpuTaskQueueIndex,  
                action.transmissionUnitTaskQueueIndex  
            )  
        }  
    }  
}
```

تابع انتقال حالت پایه

```

fun getNextStateAddingToCPU(
    sourceState: UserEquipmentState,
    queueIndex: Int
): UserEquipmentState {
    require(sourceState.cpuState == 0)
    require(sourceState.taskQueueLengths[queueIndex] > 0)

    val updatedLengths = sourceState.taskQueueLengths.decrementedAt(queueIndex)

    return sourceState.copy(
        taskQueueLengths = updatedLengths,
        cpuState = -1,
        cpuTaskTypeQueueIndex = queueIndex
    )
}

```

تابع انتقال حالت با کنش ارسال توسط واحد ارسال

```

fun getNextStateAddingToTU(
    sourceState: UserEquipmentState,
    queueIndex: Int
): UserEquipmentState {
    require(sourceState.tuState == 0)
    require(sourceState.taskQueueLengths[queueIndex] > 0)

    val updatedLengths = sourceState.taskQueueLengths.decrementedAt(queueIndex)

    return sourceState.copy(
        taskQueueLengths = updateLengths,
        tuState = 1,
        tuTaskTypeQueueIndex = queueIndex
    )
}

```

تابع انتقال حالت با کنش اجرا و ارسال به طور همزمان

```
fun getNextStateAddingToBothUnits(
    sourceState: UserEquipmentState,
    cpuQueueIndex: Int,
    tuTaskQueueIndex: Int
): UserEquipmentState {
    if (cpuQueueIndex == tuTaskQueueIndex) {
        require(sourceState.taskQueueLengths[cpuQueueIndex] > 1)
    } else {
        require(sourceState.taskQueueLengths[cpuQueueIndex] > 0)
        require(sourceState.taskQueueLengths[tuTaskQueueIndex] > 0)
    }
    return getNextStateAddingToCPU(
        getNextStateAddingToTU(sourceState, tuTaskQueueIndex),
        cpuQueueIndex
    )
}
```

پیوست ۲ – تابع ساخت شرط حداکثر توان مصرفی در برنامه خطی

تابع ساخت شرط حداکثر توان مصرفی

```
fun getEquation2(): EquationRow {
    val pLoc = systemConfig.pLoc
    val pTx = systemConfig.pTx
    val beta = systemConfig.beta
    val rhsEquation2 = systemConfig.pMax
    val coefficients = mutableListOfOfZeros(indexMapping.variableCount)

    indexMapping.coefficientIndexByStateAction.forEach { (stateAction, index) →
        val (state, action) = stateAction
        var coefficientValue = 0.0

        if (state.isTUActive())
            || (action is Action.AddToTransmissionUnit
            || action is Action.AddToBothUnits)) {
            coefficientValue += beta * pTx
        }

        if (state.isCPUActive()
            || (action is Action.AddToCPU
            || (action is Action.AddToBothUnits))) {
            coefficientValue += pLoc
        }

        coefficients[index] = coefficientValue
    }

    return EquationRow(
        coefficients = coefficients,
        rhs = rhsEquation2,
        type = EquationRow.Type.LessThan
    )
}
```


مراجع

- [1] J. Liu, Y. Mao, J. Zhang, and K. B. Letaief, "Delay-optimal computation task scheduling for mobile-edge computing systems," in 2016 IEEE International Symposium on Information Theory (ISIT), pp.1451–1455, 2016.
- [2] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," IEEE Internet of Things Journal, vol.3, no.5, pp.637–646, 2016.
- [3] A. Yousefpour, G. Ishigaki, R. Gour, and J. P. Jue, "On reducing iot service delay via fog offloading," IEEE Internet of Things Journal, vol.5, no.2, pp.998–1010, 2018.
- [4] H. Tran-Dang and D.-S. Kim, "Frato: Fog resource based adaptive task offloading for delay-minimizing iot service provisioning," IEEE Transactions on Parallel and Distributed Systems, vol.32, no.10, pp.2491–2508, 2021.
- [5] A.-E. M. Taha, N. A. Ali, and H. S. Hassanein, Frame-Structure and Node Identification, pp.147–160. 2011.

Abstract:

Exploring community structure is an appealing problem that has been drawing much attention in the recent years. One serious problem regarding many community detection methods is that the complete information of real-world networks usually may not be available most of the time, also considering the dynamic nature of such networks (e.g. web pages, collaboration networks and users friendships on social networks), it is most probable possibility that one could detect community structure from a certain source vertex with limited knowledge of the entire network. The existing approaches can do well in measuring the community quality, Nevertheless they are largely dependent on source vertex chosen for the process. Additionally, using unsuitable seed vertices may lead to finding of low quality or erroneous communities for output of many of the algorithms. This paper proposes a method to find better source vertices to be used as seeds to construct community structures locally. Inspired by the fact that many gargantuan real-world networks and respectively their graphs contain a myriad of lightly connected vertices, we explore community structure heuristically by giving priority to vertices which have a high number of links pertaining to the core structure of the network. Experimental results prove that our method can perform effectively for finding high quality seed vertices.

Keywords: Community Detection, Complex Networks, Graph Algorithms



Iran University of Science and Technology
Computer Engineering Department

Stochastic Offloading Policy for Heterogeneous Tasks

Bachelor of Science Thesis in Computer Engineering

By:

Mohammadmobin Dariushhamedani

Supervisor:

Dr. Reza Entezari-Maleki

June 2022