دانشگاه صنعتی اصفهان

دانشکده علوم ریاضی

برنامه‌ریزی صف در سامانه‌هایی با چندین سرویس‌دهنده

پایان‌نامه کارشناسی ارشد (آمار ریاضی)

مسعود فاتحی پیکانی

استاد راهنما

دکتر علی رجالی

۱۳۸۹

دانشگاه صنعتی اصفهان

دانشکده علوم ریاضی

پایان‌نامه کارشناسی ارشد (آمار ریاضی) آقای مسعود فاتحی پیکانی

تحت عنوان

برنامه‌ریزی صف در سامانه‌هایی با چندین سرویس‌دهنده

در تاریخ توسط کمیته تخصصی زیر مورد بررسی و تصویب نهائی قرار گرفت.

۱— استاد راهنمای پایان‌نامه	دکتر علی رجالی
۲— استاد مشاور پایان‌نامه	دکتر افشین پرورده
۳— استاد داور ۱	دکتر محمدحسین علامت‌ساز (دانشگاه اصفهان)
۴— استاد داور ۲	دکتر صفیه محمودی

سرپرست تحصیلات تکمیلی دانشکده دکتر اعظم اعتماد

کلیه حقوق مادی مترتب بر نتایج مطالعات، ابتکارات و نوآوری‌های ناشی از
تحقیق موضوع این پایان‌نامه متعلق به دانشگاه صنعتی اصفهان است.

فهرست مطالب

فهرست نمادها ۱

فصل اول مقدمه ۳

فصل دوم پیش نیازها ۹

۲—۱ ویژگی‌های فرآیند صف ۹

۲—۲ فرآیندهای تصادفی ۱۲

۲—۳ انتگرال ایتو ۱۶

۲—۴ همگرایی فرآیندهای تصادفی ۲۰

۲—۵ گراف ۲۵

فصل سوم بهینه‌سازی هزینه‌ها در سامانه SBR ۲۷

۳—۱ مدل ریاضی ۲۷

۳—۱—۱ متعادل‌کردن در رژیم ترافیک‌سنگین چندین سرویس‌دهنده ۳۰

۳—۱—۲ هزینه‌ها و کنترل سامانه ۳۱

۳—۲ مدل A و خط‌مشی FSF ۳۴

۳—۲—۱ ارتباط یک سامانه SBR با مدل A متناظرش ۴۳

۳—۳ نتایج اصلی ۴۶

۳—۳—۱ هزینه نگهداری ۴۷

۳—۳—۲ قاعده $GC\mu$ ۵۴

۳—۳—۳ هزینه تاخیر ۵۷

فصل چهارم یک مرکز تلفن با دو دسته متفاوت کارمند ۶۲
۴—۱ مدل A ۶۳
۴—۲ مدل N ۶۴
۴—۲—۱ خط‌مشی QIR ۶۴
۴—۲—۲ قاعده $GC\mu$ ۷۰
۴—۲—۳ خط‌مشی WIR ۷۱
۴—۲—۴ قاعده $D-GC\mu$ ۷۹

پیوست (برنامه‌های شبیه‌سازی) ۸۵

مراجع ۱۰۷

چکیده:

در این پایان‌نامه یک سامانه صف با چند کلاس مشتری و چند بلوک سرویس‌دهنده مورد بررسی قرار گرفته است . برای برنامه‌ریزی در این سامانه ؛ خانواده QIR از قواعد واگذاری معرفی شده است . یک سرویس‌دهنده برای انتخاب مشتری بعدی خود ؛ از کلاسی مشتری انتخاب می‌کند (از بین کلاس‌هایی که می‌تواند به آن‌ها سرویس دهد) که طول صف آن بیشترین اختلاف را از نسبت مشخصی از طول صف کلی دارد . بررسی سامانه به صورت تحلیلی بسیار دشوار است و از رژیم ترافیک‌سنگین چندین سرویس‌دهنده برای تقریب فرآیندهای سامانه استفاده شده است (نرخ ورود و تعداد سرویس‌دهنده‌ها به بی‌نهایت میل می‌کند) . انتظار مشتریان در هر یک از کلاس‌ها ؛ هزینه‌ای متناسب با تعداد مشتریان در صف (هزینه نگهداری) و یا میزان انتظار مشتریان (هزینه انتظار) به سامانه تحمیل می‌کند . هدف این است که به گونه‌ای سامانه برنامه‌ریزی شود تا این هزینه‌ها به حداقل ممکن برسد . نشان داده می‌شود که در کنترل QIR (وقتی که نرخ‌های سرویس فقط به بلوک سرویس‌دهنده وابسته است) با انتخاب مناسب پارامترها این هدف محقق خواهد شد . با داشتن شرایط نظم اضافه روی توابع هزینه ؛ QIR به خط‌مشی ساده‌تری تبدیل می‌شود ؛ خطی بودن توابع هزینه منجر به قاعده ساده اولویت می‌شود که در آن مشتریان صفی که کمترین هزینه را دارد ؛ پایین‌ترین اولویت را دارند . توابع هزینه محدب اکید (به علاوه چند شرط نظم دیگر) ؛ قاعده GCr را نتیجه می‌دهند . در این قاعده یک سرویس‌دهنده که تازه بی‌کار شده است ؛ برای سرویس بعد از صفی مشتری انتخاب می‌کند که در آن لحظه سرعت رشد هزینه‌های آن بیشتر است . در پایان یک مثال خاص برای مدل N (که یک سامانه با دو کلاس مشتری و دو بلوک سرویس‌دهنده است ؛ یک بلوک اختصاصی است و کارمندان بلوک دیگر به هر دو کلاس سرویس می دهند) بیان می‌شود و برخی از نتایج شبیه‌سازی مطرح می‌شوند .

فهرست نمادها

$C = \{1, ..., I\}$: مجموعه کلاس‌های مشتری در سامانه صف

$P = \{1, ..., J\}$: مجموعه بلوک‌های سرویس‌دهی در سامانه صف

$E \subset \{(i, j) \in C \times P\}$: گراف متناظر با سامانه صف (اگر کارمندان بلوک j بتوانند مشتریان کلاس i را سرویس دهند ، آن‌گاه $(i, j) \in E$)

$I(j) = \{i \in C : (i, j) \in E\}$: مجموعه کلاس‌هایی که می توانند از سرویس دهندگان بلوک j سرویس دریافت کنند

$J(i) = \{j \in P : (i, j) \in E\}$: مجموعه بلوک‌های سرویس‌دهی که می‌توانند به مشتریان کلاس i سرویس دهند

λ_i : نرخ ورود کلاس i

$\lambda = \sum_{i \in C} \lambda_i$: نرخ ورود کلی سامانه

$a_i = \frac{\lambda_i}{\lambda}$

N_j^λ : تعداد سرویس دهندگان (کارمندان) بلوک j در سامانه با نرخ ورود کلی λ

$N_\Sigma^\lambda = \sum_{j \in P} N_j^\lambda$: تعداد همه کارمندان (سرویس دهندگان) در سامانه با نرخ ورود کلی λ

μ_j : نرخ سرویس‌دهی کارمندان بلوک j

$A_i^\lambda(t)$: تعداد ورودها در کلاس i تا لحظه t در سامانه با نرخ ورود کلی λ

$A^\lambda(t) = \sum_{i \in C} A_i^\lambda(t)$: مجموع ورودها در همه کلاس‌ها تا زمان t در سامانه با نرخ ورود کلی λ

$Q_i^\lambda(t)$: طول صف کلاس i در زمان t در سامانه با نرخ ورود کلی λ

$W_{Q,i}(t)$: زمان انتظار انباشته شده مشتری سر صف کلاس i تا لحظه t

$W_{Q,i}^{\lambda,\pi}$: زمان انتظار k امین مشتری کلاس i در سامانه با نرخ ورود کلی λ و تحت خط‌مشی π

$I_j^\lambda(t)$: تعداد کارمندان بی‌کار بلوک سرویس‌دهی j در زمان t در سامانه با نرخ ورود کلی λ

$Z_{ij}^\lambda(t)$: تعداد کارمندانی از بلوک j که در زمان t (در سامانه با نرخ ورود کلی λ) مشغول سرویس‌دهی

مشتریان کلاس i هستند .

$T_{ij}^{\lambda}(t) = \int_0^t Z_{ij}^{\lambda}(\sigma)\, d\sigma$: مجموع زمان‌هایی (تا لحظه t) که کارمندان بلوک j به مشتریان کلاس i سرویس داده‌اند

$A_{ij}(t)$: تعداد مشتریان کلاس i (تا زمان t) که به‌محض رسیدن (بدون انتظار در صف) ، به بلوک j واگذار شده‌اند

$B_{ij}(t)$: تعداد مشتریان کلاس i (تا زمان t) که بعداز انتظار در صف توسط کارمندان بلوک j انتخاب شده‌اند

$Q_{\Sigma}^{\lambda}(t) = \sum_{i \in C} Q_i^{\lambda}(t)$: طول صف کلی در زمان t در سامانه با نرخ ورود λ

$I_{\Sigma}^{\lambda}(t) = \sum_{j \in P} I_j^{\lambda}(t)$: تعداد همه کارمندان بی‌کار در زمان t در سامانه با نرخ ورود λ

$X_{\Sigma}^{\lambda}(t) = Q_{\Sigma}^{\lambda}(t) + \sum_{j \in P} (N_j^{\lambda} - I_j^{\lambda}(t))$: مجموع همه مشتریان درسامانه صف در زمان t

$$\hat{I}_i^{\lambda}(t) = \frac{I_i^{\lambda}(t)}{\sqrt{\lambda}}$$

$$\hat{Q}_i^{\lambda}(t) = \frac{Q_i^{\lambda}(t)}{\sqrt{\lambda}}$$

$$\hat{I}_{\Sigma}^{\lambda}(t) = \frac{I_{\Sigma}^{\lambda}(t)}{\sqrt{\lambda}}$$

$$\hat{Q}_{\Sigma}^{\lambda}(t) = \frac{Q_{\Sigma}^{\lambda}(t)}{\sqrt{\lambda}}$$

$$\hat{W}_{Q_i}^{\lambda}(t) = \sqrt{\lambda} W_{Q_i}^{\lambda}(t)$$

$$\hat{X}_{\Sigma}^{\lambda}(t) = \frac{X_{\Sigma}^{\lambda}(t) - N_{\Sigma}^{\lambda}}{\sqrt{\lambda}}$$

$J_1^{\lambda}(\pi, T) = \int_0^T \sum_{i=1}^{I} C_i(Q_i^{\pi,\lambda}(t))dt$: هزینه نگهداری تا زمان T و با تابع هزینه نگهداری $C(.) = (C_1(.), ..., C_I(.))$ در سامانه با نرخ ورود کلی λ

$J_2^{\lambda}(\pi, T) = \sum_{i=1}^{I} \frac{1}{\lambda_i} \sum_{k=1}^{A_i^{\lambda}(T)} C_i(\sqrt{\lambda} W_{Q_i}^{\lambda}(\pi))$: هزینه تاخیر تا زمان T و با تابع هزینه تاخیر $C(.) = (C_1(.), ..., C_I(.))$ درسامانه با نرخ ورود کلی λ

Π_1: مجموعه همه خط‌مشی‌های بدون‌پیش‌بینی

Π_2 : مجموعه خط‌مشی‌هایی از Π_1 است که در آن‌ها مشتریان در هر کلاس به صورت FCFS سرویس می‌شوند

Π_4 : مجموعه خط مشی‌های بدون پیش‌بینی مجانباً کارا

از اندیس SBR برای فرآیندهای سامانه صف SBR و از اندیس A برای فرآیندهای سامانه صف A استفاده شده‌است . همچنین بالانویس λ برای فرآیندها در سامانه با نرخ ورود کلی λ و از بالانویس π برای فرآیندهای سامانه صف تحت خط‌مشی π استفاده شده‌است .

فصل ۱

مقدمه

مطالب این قسمت بر گرفته از مقاله [۱۱] است .

امروزه خدمات و سرویس‌دهی ، بخش بزرگی از اقتصاد جهان را تشکیل می‌دهد که در دو روش مستقیم و غیرمستقیم انجام می‌شود . مثال‌های شاخص از نوع اول شعبه‌های بانک و خدمات بیمارستانی است و از نوع دوم می‌توان به پرداخت قبوض (از طریق تلفن یا اینترنت) یا گرفتن اطلاعات حساب بانکی اشاره کرد . خدمات غیرمستقیم از طریق مراکز تماسی که از کانال‌های مختلف شامل وب‌چت ، ایمیل و تلفن با مشتریان در تماس‌اند ، انجام می‌شود . در گذشته خدمات غیرمستقیم به وظایف ساده محدود بود، اما اخیراً وظایف خدماتی پیچیده از حالت مستقیم به خدمات غیر مستقیم در حال انتقال هستند . به عنوان مثال بانک‌ها امروزه خدمات مشاوره سرمایه‌گذاری را از طریق چت و تلفن ارائه می‌دهند.بحث این پایان‌نامه بر روی مراکز تماسی متمرکز است که از طریق تلفن خدمات ارائه می دهند، بنابراین عبارت مرکز تلفن به جای عبارت کلی‌تر مرکز تماس بکار برده می‌شود .

مرکز تلفن که همان ترجمه عبارت Call Center است به نهادهایی اطلاق می‌شود که خدمات بسیار گسترده‌ای را از طریق تلفن ارائه می‌دهند . باتوجه به‌تنوع خدمات و محدودیت‌هایی که در آموزش کارمندان وجود دارد ، معمولا در مراکز تلفن بزرگ ، کارمندان با توجه به مهارت‌هایشان در دسته‌های مختلف قرار می‌گیرند که آن دسته‌ها را بلوک‌های کارمندی می‌نامند . مشتریان نیز بر اساس معیارهای مورد نظر مدیریت مرکز تلفن در کلاس‌های مختلف دسته‌بندی می‌شوند . برخی از مراکز تلفن مشتریانی از دیگر کشورها را سرویس می‌دهند که در این مورد مشتریان به طور طبیعی بر اساس زبانی که صحبت

می‌کنند کلاس‌بندی می‌شوند؛ ویا ممکن است مشتریان براساس ارزش تجارت آنها کلاس‌بندی شوند. در نتیجه داشتن کلاس‌های‌مشتری و بلوک‌های‌کارمندی چندگانه بسیار معمول است و این چندگانگی است که سامانه سرویس‌دهی موازی (مانند شکل ۱ درقسمت ۲—۱) را به عنوان یک مدل طبیعی از مرکز تلفن مجاز می‌سازد.

بسیاری از مراکز تلفن بزرگ هستند و صدها یا حتی هزاران کارمند را به‌کار گرفته‌اند. پس عجیب نیست که بیش از ۷۰٪ هزینه‌های اداره‌ی این مراکز مربوط به هزینه‌های کارمندان است. مدیران مرکز تلفن با این چالش غیر بدیهی مواجهند که بین بهره‌وری کارمندان و خدمات کارا تعادل بهینه برقرار کنند. چالش در این است که سطح کیفی خدمات یا هدف فروش از پیش تعیین شده را برآورده نمایند و همزمان هزینه‌های کارمندی را حد اقل کنند واز منابع به‌صورت موثر بهره‌برداری نمایند.

کارایی خدمات به چند روش اندازه‌گیری می‌شود. در مراکز تلفن با خدمات محدود کارایی به‌وسیله زمان‌های انتظار، نرخ ترک‌کردن یا کمیت‌های دیگر که برای تعیین کیفیت سرویس به‌کار می‌روند، اندازه‌گیری می‌شود. در مراکز تلفنی که خدمات و فعالیت‌های تجاری با هم آمیخته شده‌اند، منطقی است که کارایی همچنین با عباراتی چون درآمد یا دیگر اندازه‌های مربوط به فروش اندازه‌گیری شود.

مدیریت توام بهره‌وری و کارایی یک مرکز تلفن مسئله ساده‌ای نیست و به مدل‌های اطلاعاتی [۱]، منابع انسانی[۲] و همچنین تحقیق در عملیات[۳] نیاز دارد. قسمتی که در ادامه می‌آید سیمایی از مدیریت مرکز تلفن است که بیشتر به این بحث مربوط می‌شود، مسئله اصلی این‌است که در طول زمان، مشتری کدام کلاس به کدام بلوک واگذار شود. برای دیدن منظرهای دیگر از مدیریت مرکز تلفن و جزئیات بیشتر به [۲] و [۹] می‌توان مراجعه کرد.

پیچیدگی اصلی در مدیریت مرکزتلفن مدرن، نتیجه چندگانگی کلاس‌های مشتری و بلوک‌های کارمندان و چگونگی ارتباط آنها است. توزیع‌کننده‌های اتوماتیک تماس(ACD's)[۴] مدرن در ارتباط با این چندگانگی به اندازه کافی پیشرفته هستند. ACD می‌تواند طوری برنامه‌ریزی شود که مشتریان از هر کلاس را فقط به کارمندانی که مهارت‌های مربوطه را دارند واگذار کند. به این قابلیت ACD ها تحت عنوان واگذاری مبتنی بر مهارت (SBR)[۵] اشاره می‌شود. ACD به قواعد ورودی از طرف مدیریت مرکز تلفن نیاز دارد یعنی باید قواعد تصمیم برنامه‌ریزی و واگذاری، هر دو مشخص شوند. قواعد واگذاری تعیین می کند که هنگام ورود یک مشتری، اگر در بیش از یک بلوک کارمندانی وجود دارند که می توانند به آن مشتری سرویس دهند، ACD بایستی مشتری را در اختیار کدام سرویس دهنده قرار دهد؛ قواعد

[۱] Finate-Information Systems
[۲] Human Resources
[۳] Operation Research
[۴] Automatic Call Distribution
[۵] Skill-Based Routing

برنامه ریزی نیز مشخص می سازد یک کاربرد که نیازمند یک کارشده است از بین کلاس‌های مشتریانی که می تواند به آنها سرویس دهد، بایستی کدام مشتری را برای سرویس بعدی انتخاب کند. ACD های مدرن می‌توانند طوری پیکربندی شوند که قواعد تصمیم بسیار متنوعی را به‌کار برند ؛ اما آنها نمی‌توانند بهینه‌سازی بهترین قاعده از میان سایر برنامه‌ها را انجام دهند . در حقیقت پیدا کردن قواعد بهینه، به‌جز برای مدل‌های بسیار محدود، در حال حاضر مسئله‌ای حل نشده‌است . پیدا کردن قواعد "خوب" به آنالیز پیشرفته در سامانه های صف و تعیین مجموعه قواعدی نیاز دارد که‌با مد نظر قراردادن اهداف اقتصادی بهینه یا نزدیک به بهینه‌باشند . عنا سفانه این سیستم های چندکلاسی ، چندبلوکی با تحلیل واقعی کارایی قابل مهار نیستند و بایستی برای یافتن جواب های بهینه تقریبی از رویکردهای مختلف استفاده کرد . رویکرد در اینجاتقریب زدن با استفاده از رژیم ترافیک سنگین است که با استفاده از آن بهینگی مجانبی خانواده‌QIR[6] از قواعد واگذاری نشان داده می‌شود .

کنترل بهینه سامانه‌های صف برای مینیمم کردن هزینه‌های نگهداری یا ماکزیمم کردن در آمدها، موضوع بسیاری از مقاله‌هاست . بسیاری از این مقالات از فرآیندهای تصمیم مارکوف (MDP[7]) استفاده کرده‌اند . اما حوزه دیدگاه MDP لزوما به مدل‌هایی که دارای خاصیت مارکفی‌اند محدود می‌شود . با استفاده از این روش جواب‌های بهینه تحلیلی فقط برای مدل‌های ساده می‌توانند پیدا شوند .

دیدگاه دیگری که مدل‌های پیچیده را قابل بررسی می‌کند ،استفاده از تقریب ترافیک‌سنگین است . این تقریب بادر نظرگرفتن حد یک دنباله از فرآیندهای صف که به‌طور مناسب مقیاس شده‌اند ، برای توصیف و کنترل سامانه به‌کار می‌رود . هدف این است که از بین روش‌هایی که به‌طور مجانبی بهینه‌اند ، روش مناسب انتخاب شود . در رژیم ترافیک‌سنگین سنتی با ثابت نگه داشتن تعداد سرویس‌دهنده‌ها ، حجم تقاضا افزایش می‌یابد . اما روشی که در این پایان‌نامه استفاده شده است رژیم ترافیک‌سنگین چندین سرویس‌دهنده است که در آن تعداد سرویس دهنده‌ها همراه با حجم تقاضا افزایش می‌یابد . از کنترل‌های مجانبی به‌دست آمده در این روش برای کنترل سامانه‌های خدمات با تعداد زیاد سرویس‌دهنده (مانند مراکز تلفن بزرگ) استفاده می‌شود .

یک مقاله کلیدی در رژیم ترافیک‌سنگین سنتی نوشته مندلبوم[8] و استولیر[9] [۱۴] است که در آن مقاله ، کار اولیه وان میگم[10] [۱۷] (با آرایش چند کلاس و یک سرویس‌دهنده) توسعه داده شده است . در این مقالات و همچنین در این پایان‌نامه سامانه صف دارای چند کلاس مشتری است و مشتری‌ها پس از اتمام یک سرویس سامانه را ترک می‌کنند . این ویژگی متفاوت است با شبکه‌های صف که در آن

[6] Queue-and-Idleness Ratio
[7] Markov Decision Processes
[8] Mandelbaum
[9] Stolyar
[10] Van Mieghem

مشتری‌ها قبل از ترک سامانه از چندین جایگاه ، سرویس دریافت می‌کنند . مثال‌هایی از سامانه‌های ساده با سرویس‌دهنده‌های مولتی در شکل ۲ (در بخش ۳—۱) نشان داده شده‌اند . هر مستطیل در بالا یک صف برای مشتری‌های یک کلاس را نشان می‌دهد و هر دایره در پایین یک بلوک از سرویس‌دهندگان با مهارت‌های مشترک است .

مندلبوم و استولیر [۱۴] نشان دادند که قاعده $GC\mu$ به‌طور مجانبی هزینه‌های نگهداری برای توابع محدب را حداقل می‌کند . فرض کنید μ_{ij} نرخ سرویس مشتری کلاس i توسط سرویس‌دهنده بلوک j و $Q_i(t)$ طول صف کلاس i در زمان t باشند و نیز صف کلاس i موجب هزینه‌ای با نرخ $C_i(Q_i(t))$ باشد (وقتی‌که $C_i(.)$ تابعی محدب اکید ، صعودی اکید ، دو بار پیوسته مشتق پذیر با $C_i(0) = C_i'(0) = 0$ است) آن گاه قاعده $GC\mu$ بیان می‌کند که هرگاه یک سرویس‌دهنده از بلوک j در زمان t بی‌کار شد ، مشتری بعدی برای سرویس از کلاسی انتخاب می‌شود که $\mu_{ij}C_i'(Q_i(t))$ را بیشینه کند ؛ یعنی مشتری کلاس i^* توسط سرویس‌دهنده بلوک j سرویس می‌شود هرگاه

$$i^* \in \arg\max_i \ \mu_{ij}C_i'(Q_i(t))$$

در قاعده $GC\mu$ برای تصمیم گرفتن در مورد مشتری بعدی برای کارمند بلوک j کافی است طول صف کلاس‌هایی که آن کارمند می‌تواند به آنها سرویس دهد ، مشخص باشد . در نتایج بهینگی مجانبی در [۱۴] از این واقعیت استفاده شده است که وقتی شرایط مناسب برقرار باشد ، فرآیند طول صف کلی از پایین به یک حرکت براونی بازگشتی (RBM[11]) کراندار خواهد بود ؛ بنابراین برای حداقل کردن هزینه‌ها به‌طور مجانبی کافی‌است مطمئن شد که

۱) طول صف کلی به‌طور مجانبی برابر با RBM است (یعنی به حداقل ممکن رسیده است) ،

۲) این طول صف به‌طوری بین کلاس‌های مختلف توزیع شده است که هزینه‌ها حداقل می‌شود .

مندلبوم و استولیر نشان دادند که $GC\mu$ به هر دو شرط دست می‌یابد .

گرویچ [12] و وایت [13] [۱۰] نتایج مندلبوم و استولیر [۱۴] را به رژیم ترافیک‌سنگین چندین سرویس‌دهنده توسعه داده‌اند (در این پایان‌نامه نیز نتایج مربوط به این مقاله بررسی می‌شود) . برخلاف مقاله [۱۴] که فقط توابع هزینه محدب اکید را مورد بررسی قرار می‌دهد ، نتایج مقاله [۱۰] توابع هزینه خطی را نیز پوشش می‌دهد . این از تفاوت‌های مهم بین دو رژیم ترافیک سنگین سنتی و چندین سرویس‌دهنده است . تفاوت دیگر این است که نتایج بهینگی برای رژیم ترافیک‌سنگین چندین سرویس‌دهنده محدود به سامانه‌هایی است که در آن نرخ سرویس فقط به بلوک سرویس بستگی دارد ، یعنی $\mu_{ij} = \mu_{kj}$ و این در

[11] reflected Brownian Motion

[12] Gurvich

[13] Whitt

عمل پارامتر μ را از قاعده GC_μ حذف می‌کند . گریچ و ولیت [۱۱] خانواده کنترل‌های QIR (تعریف ۱۹.۳) را معرفی کرده‌اند و نشان دادند که نوع خاصی از QIR (با انتخاب مناسب پارامترها) به‌طور مجانبی هزینه‌های نگهداری را حداقل می‌کند و این نتایج ، هزینه‌های خطی را نیز در بر می‌گیرد . به‌علاوه زمانی‌که توابع هزینه محدب اکید باشند (با شرایط نظم اضافی) مولفه برنامه‌ریزی QIR به قاعده ساده GC_μ تبدیل می‌شود . (البته همان‌طور که گفته شد عنصر μ نقشی ندارد) .

مقاله گرویچ [۱۰] اولین مقاله‌ای نیست که حداقل کردن هزینه‌ها با استفاده از رژیم ترافیک‌سنگین چندین سرویس‌دهنده را بررسی می‌کند . هریسون [۱۴] و زیوی [۱۵] [۱۳] یک مدل چند کلاسی با یک بلوک سرویس را(مدل V) ، وقتی که توابع هزینه‌های نگهداری و ترک سامانه خطی هستند ، مورد بررسی قرار داده‌اند . آرمونی در [۱] طول صف را در یک مدل A (یک مدل با یک کلاس مشتری و چند بلوک سرویس‌دهنده) حداقل می‌کند . در این پایان‌نامه علاوه بر این که برخی از نتایج او بیان می‌شود ، نتایجی که گرویچ [۱۰] با توسعه نتایج آرمونی گرفته است ، نیز بیان می‌شود .

یک سامانه SBR در حالت کلی (سامانه با چند کلاس مشتری و چند بلوک سرویس‌دهنده) توسط اتار[۱۶] [۳] بررسی شده‌است . او با استفاده از رژیم ترافیک‌سنگین چندین سرویس‌دهنده هزینه‌های نگهداری را به‌طور مجانبی حداقل می‌کند . نتایج او بسیار کلی است و این کلی بودن بینش کمی در مورد حالات خاص ایجاد می‌کند و همچنین کنترل‌های پیشنهادی او به روشنی قاعده GC_μ نیستند .

نتایج گرویچ و ولیت در [۱۱،۱۰] ارتباط نزدیکی دارد با نتایجی که به‌طور همزمان اما مستقل از آن ها توسط دای [۱۷] و تزکن [۱۸] [۸،۷] ارائه شد . مقاله اولشان [۷] جواب‌های روشنی را برای موارد خاص با فرض خطی بودن توابع هزینه نگهداری در بر می‌گیرد . در این مقاله مدل N (یک مدل با دو کلاس مشتری و دو بلوک سرویس‌دهنده که یک بلوک فقط به یکی از کلاس‌ها سرویس می‌دهد و بلوک دیگر می‌تواند به مشتریان هردو کلاس سرویس دهد) مورد بررسی قرار می‌گیرد . مقاله دوم [۸] یک سامانه SBR عمومی با نرخ‌های سرویس وابسته به بلوک سرویس را بررسی می‌کند ، تعداد کلاس‌های مشتری و بلوک‌های سرویس دلخواه است اماهنوز محدودیت خطی بودن توابع هزینه وجود دارد . گرویچ [۱۰] (که در این پایان‌نامه بحث شده است) ، نتایج آنها را به حالت کلی‌تر وقتی توابع هزینه نگهداری محدب هستند ، توسعه می‌دهد . به‌علاوه نشان داده می‌شود که خطی‌بودن توابع هزینه برای روشی که بهینه بودن آن در [۸] نشان داده شده‌است ، یک شرط لازم نیست و یک شرط ضعیف‌تر روی مشتق توابع هزینه برای

[۱۴] Harrison
[۱۵] Zeevi
[۱۶] Atar
[۱۷] Dai
[۱۸] Tezcan

بهینگی آن روش کافی است .

روش‌های تحلیل در [۸] و [۱۰] شبیه به هم است اما تفاوت‌هایی نیز وجود دارد ؛ دای و تزکن [۸]کارشان را بر اساس کار قبلیشان که تعمیمی از مقاله براسون [۹] [۵] است ، بنا کرده‌اند . اما گرویج و وایت [۱۰] از نتایج مقاله قبلی خود [۱۱] و نتایج آرمونی [۱] استفاده کرده‌اند .

در این پایان‌نامه در فصل ۲ مفاهیم و اصطلاحات مورد نیاز در طول پایان‌نامه توضیح داده شده‌اند . در فصل ۳، ابتدا مدل ریاضی برای یک مرکز تلفن ارائه می‌شود و فرض‌هایی درمورد سامانه صف و همچنین فرض‌های ترافیک سنگین چنین سرویس‌دهنده بیان می‌شوند . سپس مدل A به عنوان یک حالت خاص از سامانه‌های SBR بررسی می‌شود و یک خط‌مشی بهینه‌مجانبی برای این مدل معرفی می‌شود . بعد از آن ارتباط یک سامانه SBR عمومی با مدل A بیان می‌شود و از این ارتباط برای نشان دادن بهینگی مجانبی خط‌مشی QIR استفاده می‌شود . در ادامه فصل چند خط‌مشی دیگر معرفی می‌شوند و بهینگی آن‌ها و نیز ارتباطشان با یکدیگر بررسی می‌شود . در فصل ۴ ، قضایا و مفاهیم بیان شده در فصل ۳ توسط یک مثال عملی توضیح داده می‌شوند .

[۹] Brumelle

فصل ۲

پیش نیازها

۲–۱ ویژگی‌های فرآیند صف

در این پایان‌نامه ، یک سامانه صف را به‌طور کلی می‌توان به این صورت توصیف کرد که مشتریانی جهت استفاده از خدمات وارد سامانه می‌شوند ، اگر سرویس فوراً در اختیار نباشد منتظر می‌مانند (و یا سامانه را ترک می‌کنند) و پس از دریافت سرویس سامانه را ترک می‌کنند . (کلمه مشتری در یک مفهوم عام و کلی به کار می‌رود و لزوما یک مشتری انسانی را بیان نمی‌کند ؛ به عنوان مثال یک مشتری می‌تواند یک تماس تلفنی باشد که منتظر است پاسخ داده شود.) آشکار است که توصیف دقیق و معقولی از چنین سیستمی نیاز به تعیین ویژگی فرآیندهای مربوط دارد . ویژگی‌های اساسی فرآیندهای صف در این قسمت مورد بحث قرار می‌گیرد . این ویژگی‌ها به قرار زیرند([۱۱] و گراس [۲۱]) :

۱) الگوی ورودی مشتریان — مشتریان می‌توانند از یک یا چند ورودی وارد سامانه شوند. مشتریان ورودی‌های مختلف ممکن است از نظر خدماتی که متقاضی دریافت آن هستند متفاوت باشند. در این صورت هر ورودی را یک کلاس مشتری[1] می‌نامند. ورود مشتریان را بر حسب میانگین تعداد ورودها در واحد زمان و یا به‌طور دقیق‌تر در شکل توزیع احتمالی مربوط به فرآیند ورود تعیین می‌نمایند. عامل مورد توجه دیگر راجع به فرآیند ورود، این است که ورودی‌ها به‌طور یکی یکی وارد سامانه شوند و یا بتوانند به صورت دسته جمعی وارد شوند. همچنین لازم است واکنش مشتری را به محض ورود به سامانه دانست. مشتری ممکن است تصمیم بگیرد طول صف هر اندازه که باشد (و یا هر قدر منتظر بماند) صبر کند و یا شاید در صورت طولانی بودن بیش از حد صف (و یا طولانی شدن بیش از حد انتظارش) تصمیم بگیرد وارد صف نشود (و یا پس از کمی انتظار سامانه را ترک کند). در مواردی که دو یا چند صف موازی وجود دارد مشتریان ممکن است از یک صف انتظار به صف انتظار دیگری بروند. عامل دیگری که راجع به الگوی ورودی مورد بررسی قرار می‌گیرد نحوه‌ای است که در آن الگو با زمان تغییر می‌کند. الگوی ورودی که با زمان تغییر نمی‌کند، الگوی ایستا نامیده می‌شود. اگر $Q(t)$ تعداد افراد در صف در لحظه t را نشان دهد آن گاه فرآیند تصادفی $\{Q(t)\}_{t\geq 0}$ را فرآیند طول صف می گویند.

۲) الگوی کار سرویس‌دهندگان — اکثر مباحث راجع به الگوی ورودی، در بحث الگوهای سرویس نیز مناسب هستند، اما یک اختلاف مهم بین سرویس و ورودی‌ها وجود دارد و آن این‌است که موقعی که راجع به سرویس یا زمان سرویس صحبت می‌شود، عبارات بر روی خالی نبودن سامانه شرطی می‌شوند. زمان سرویس بر حسب توزیع احتمال آن بیان می‌شود. ممکن است نرخ سرویس به تعداد مشتریان در صف بستگی داشته باشد. همچنین زمان سرویس نیز مانند ورودی‌ها می‌تواند نسبت به زمان ایستا یا ناایستا باشد.

۳) تعداد کانال‌های سرویس — تعداد کانال‌های سرویس اشاره به تعداد ایستگاه‌های سرویس موازی دارد که می‌توانند به‌طور همزمان به مشتریان سرویس دهند. گاهی تعداد کانال‌های سرویس بسیار زیاد است و سرویس‌دهندگان همه (از لحاظ آماری) یکسان نیستند؛ در این حالت به هر دسته از سرویس‌دهندگان یکسان، یک بلوک سرویس‌دهنده[2] گفته می‌شود. به بیانی دیگر، هر بلوک سرویس شامل چندین سرویس‌دهنده (کانال سرویس) است که همگی از لحاظ آماری یکسان‌اند، یعنی زمان سرویس آنها دارای توزیع یکسان است و سرویس‌دهندگان مهارت‌های مشابه دارند.

[1] Customer Class
[2] Server Pool

۴) نظم صف — نظم صف اشاره به چگونگی انتخاب مشتریان برای ورود به سرویس دارد. در معمول‌ترین نظمی که در زندگی روزمره هم می‌توان مشاهده کرد اولین فرد ورودی، اولین فرد برای سرویس (FCFS)[۲] است. به هر حال این تنها نظم صف ممکن نیست. در مواردی که چند کلاس مشتری و چند بلوک سرویس وجود دارد، مسئله کمی پیچیده‌تر می‌شود. ممکن است یک بلوک از سرویس‌دهندگان بتوانند کلاس‌های خاصی از مشتریان را سرویس دهند و به تبع آن، هر مشتری برای سرویس می‌تواند وارد بلوک‌های معینی شود. مسئله‌ای که در این موارد مطرح می شود شامل دو بخش است؛ اول اینکه هنگام ورود یک مشتری اگر در بیش از یک بلوک، سرویس‌دهندگانی وجود دارند که می‌توانند به او سرویس دهند مشتری بایستی به کدام بلوک واگذار شود. و دوم این که، یک سرویس‌دهنده که تازه بی‌کار شده است از بین کلاس‌های مشتری که توان سرویس آن‌ها را دارد، بایستی از کدام کلاس برای سرویس بعد مشتری انتخاب کند. به قانونی که به سوال اول پاسخ می‌دهد قاعده واگذاری [۴] و به پاسخ دومین سوال، قاعده برنامه‌ریزی [۵] می‌گویند.

۵) ظرفیت سامانه — در بعضی از فرآیندهای صف یک محدودیت فیزیکی برای مکان انتظار وجود دارد. این امر موجب می‌شود موقعی که صف به طول معینی می‌رسد، دیگر مشتریان نتوانند وارد سامانه شوند تا زمانی که با تکمیل یک سرویس فضا مهیا شود.

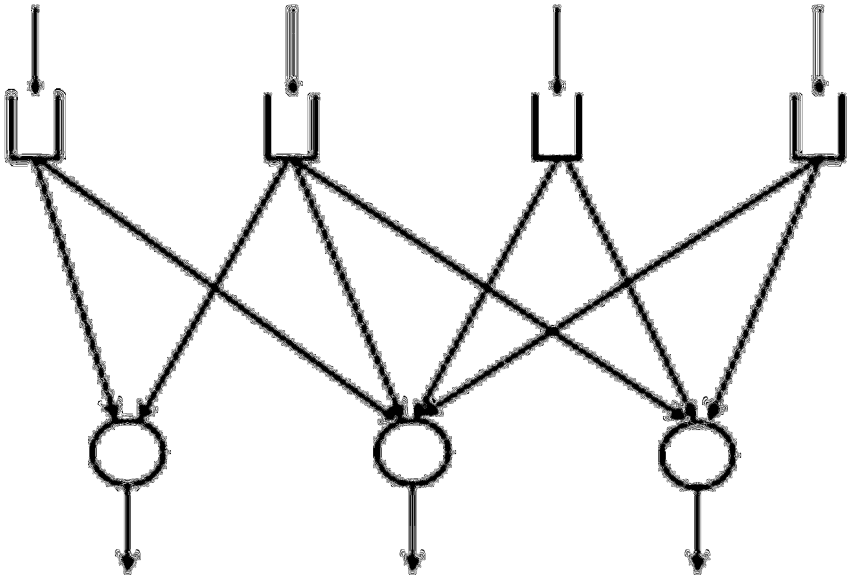
۶) مراحل سرویس — یک سامانه ممکن است تنها یک مرحله سرویس داشته باشد و یا ممکن است دارای چند مرحله سرویس باشد. یعنی ممکن است مشتری بعد از تکمیل اولین سرویس، سامانه را ترک کند و یا وارد سرویس دیگری شود.

این شش ویژگی به مشخصه‌های اصلی یک سامانه صف اشاره می‌کند. در زیر نوع خاصی از سامانه‌های صف که در این پایان‌نامه به عنوان مدل صف یک مرکز تلفن استفاده شده است، توضیح داده می‌شود.

سامانه سرویس‌دهی موازی (PSS[۶]) — یک سامانه سرویس‌دهی موازی، یک سامانه صف با چند کلاس مشتری و چند کلاس سرویس‌دهنده است. سرویس‌دهندگان هر بلوک می‌توانند مشتریان تعدادی از کلاس‌ها (و نه لزوماً همه مشتریان) را سرویس دهند. مشتریان یکی یکی وارد صف می‌شوند و

[۳] First Come, First Serve
[۴] Routing Rule
[۵] Scheduling Rule
[۶] Parallel-Server System

هر مشتری پس از دریافت یک سرویس سامانه را ترک می‌کند . شکل زیر نمایشی از یک سامانه سرویس‌دهی صوازی با چهار کلاس مشتری و سه بلوک سرویس‌دهنده است ، هر مستطیل در بالا یک کلاس مشتری و هر دایره در پایین یک بلوک سرویس را نشان می‌دهند . کلاس‌های مشتری با فلش به بلوک‌هایی که می‌توانند مشتریان آن کلاس را سرویس دهند ، متصل شده‌اند .



شکل ۱ : یک سامانه PSS با گراف متناظرش .

هر مستطیل یک کلاس مشتری و هر دایره یک بلوک سرویس‌دهنده را نشان می‌دهد .

۲—۲ فرآیندهای تصادفی

تعریف ۱.۲ (فرآیند تصادفی)

برای یک مجموعه T ، خانواده متغیرهای تصادفی $\{X(t), t \in T\}$ را یک فرآیند تصادفی با فضای پارامتر T و فضای حالت Z می گویند ؛ وقتی که Z ، مجموعه تمام مقادیری است که متغیرهای تصادفی $X(t)$ ، برای هر $t \in T$ ، می‌توانند اختیار کنند .

اگر T یک مجموعه پیوسته (معمولا $[0, \infty)$) باشد ، آن گاه فرآیند $\{X(t), t \in T\}$ را زمان پیوسته

می گویند . فرآیند $\{X(t), t \in [0, \infty)\}$ را با $\{X_n\}_{n \geq 0}$ نیز نشان می‌دهند .

برای هر $\omega \in \Omega$ ، تابع $t \mapsto X_n(\omega)$ از T به Z را با x نمایش داده و آن را یک مسیر نمونه ای از فر آیند $\{X(t), t \in T\}$ می‌گویند .(صفحات ۹۳و۹۴ بیشلت [۴])

تعریف ۲.۲ (فرآیند پواسون)

فرآیند شمارشی $\{N_n\}_{n \geq 0}$ را یک فر آیند پواسون با نرخ λ $(\lambda > 0)$ گویند هرگاه

i) $N_0 = 0$

ii) $\{N_n\}_{n \geq 0}$ دارای نموهای مستقل باشد . (یعنی برای هر $t_1 \leq t_2 \leq \ldots \leq t_n$ ، متغیرهای $N_{t_2} - N_{t_1}, N_{t_3} - N_{t_2}, \ldots, N_{t_n} - N_{t_{n-1}}$ از یکدیگر مستقل باشند)

iii) نموهای $\{N_n\}_{n \geq 0}$ ، $0 \leq s < t$, $N(t, s) = N_t - N_s$ ، دارای توزیع پواسون با پارامتر $\lambda(t-s)$ باشند،

$$P(N(t, s) = i) = \frac{(\lambda(t-s))^i}{i!} e^{-\lambda(t-s)}$$

اگر $\{N_n^\lambda\}_{n \geq 0}$ فرآیند پواسون با نرخ λ باشد ، آن گاه برای هر $t \in R^+$ (با استفاده از قانون قوی اعداد بزرگ) می توان نوشت :

$$\lim_{\lambda \to \infty} \frac{N_t^\lambda}{\lambda} = t \qquad (1)$$

(بیشلت [۴])

تعریف ۳.۲ دنباله متغیرهای تصادفی حقیقی مقدار ، نا منفی ، مستقل و همتوزیع $\{Y_i;\ i = 1, 2, \ldots\}$ را فرآیند تجدید معمولی[۲] می گویند .

Y_i زمان بین $(i-1)$امین و iامین تجدید است ؛ $Y_0 = 0, i = 1, 2, \ldots$. زمان تصادفی که n امین تجدید اتفاق می‌افتد با $T_n = \sum_{i=1}^{n} Y_i$ نمایش داده می‌شود . فرآیند تصادفی $\{T_1, T_2, \ldots\}$ فرآیند زمان‌های تجدید نامیده می‌شود .

فرآیند شمارشی متناظر $\{A(t); t > 0\}$ که با $A(t) = \max\{n : T_n \leq t\}$ تعریف می‌شود را فرآیند شمارش تجدید می‌نامند (۳.۳§ بیشلت [۴]).

[۲] ordinary renewal process