مدل سازی و ارزیابی سیستم های کامپیوتری

تألیف:
دکتر محمد جعفر تارخ

موسسه علمی فرهنگی

تارخ، محمدجعفر
مدلسازی وارزیابی سیستمهای کامپیوتری [مؤلف] محمد جعفر تارخ .
تهران: موسسه علمی و فرهنگی نص، / مرکز تحقیقات مخابرات ایران / ۱۳۸۱ .
۲۰۸ ص، مصور،
ISBN: 964-5801-50-8 : ۱۲۰۰۰ ریال
فهرستنویسی بر اساس اطلاعات فیپا.
۱. سیستمهای کامپیوتری الف. عنوان
۴ م ۲ ت/ QA۷٦ ۰۰۴/۳
محل نگهداری:
کتابخانه ملی ایران ۱۳٦۳٦-۸۱م

مرکز تحقیقات مخابرات ایران

مدلسازی و ارزیابی سیستمهای کامپیوتری

مؤلف: دکتر محمد جعفر تارخ
چاپ اول: پاییز ۸۱
چاپ و صحافی: سازمان چاپ و انتشارات وزارت فرهنگ و ارشاد اسلامی
طراحی، آماده سازی: موسسه علمی فرهنگی «نص»
قیمت: ۱۲۰۰ تومان

تهران: انتهای کارگر شمالی - مرکز تحقیقات مخابرات ایران
تلفن: ۴-۶۳۸۰۰۶۱ فکس ۶۳۵۵۸۸ ص.پ. ۳۹۶۱-۱۴۱۵۵
وب سایت : www.itrc.ac.ir

تهران: میدان انقلاب، خیابان اردیبهشت، بن بست مبین، شماره ۲۳۷
تلفکس: ۶۴۱۲۳۸۵ ـ تلفن: ۶۴۶۵۶۷۴ / ص.پ. ۱۳۱۴۵/۸۶۳

شابک: ۸-۵۰-۵۸۰۱-۹۶۴ ISBN:964-5801-50-8

تقدیم به همسرم

که پژوهش را از او آموختم

دیباچه:

سیستم‌های نرم‌افزاری، سخت‌افزاری و شبکه‌های کامپیوتری در فعالیتهای مختلف بشری مؤثر و حائز اهمیت می‌باشند. با توجه به آنکه کارآیی و قابلیت اطمینان موضوع حساس و دقیق در همهٔ زمینه‌های صنعت، تجارت و خدمات می‌باشد، ارزیابی سیستم‌های فوق‌الذکر بیش از بیش بعنوان یکی از مباحث اساسی در علوم کامپیوتر، تکنولوژی ارتباطات و اطلاعات مورد بررسی و مطالعه قرار می‌گیرد. کتاب حاضر مروری در مدلسازی و ارزیابی سیستم‌های کامپیوتری جهت استفاده دانشجویان و پژوهشگران در زمینه‌های مخابرات، کامپیوتر، صنایع و ریاضیات کاربردی می‌باشد.

وظیفهٔ خود می‌دانم از جناب آقای مهندس مزینانی که با ارائه نظرات خود در تدوین این کتاب همکاری داشته‌اند تشکر و قدردانی نمایم.

محمد جعفر تارخ

استادیار دانشکدهٔ مهندسی صنایع

دانشگاه صنعتی خواجه نصیرالدین طوسی

تهران، تابستان ۸۱

فهرست مطالب

۱

مقدمه‌ای بر احتمالات

۱-۱ مقدمه

باتوجه به مطالعه نظریه احتمالات، آزمایش عبارتست از یک عمل فیزیکی که در دنیای واقعی می‌تواند به یک نتیجه ممکن منجر شود. مثلاً اگر یک جفت تاس را یکمرتبه بریزیم دو عددی که ممکن است مشاهده کنیم از جفت یک تا جفت شش می‌تواند باشد. اگر شخص معینی بخواهد صدمتر را با هر سرعتی که بدوَد، زمانیکه از لحظه شروع تا لحظه پایان می‌گذرد ممکن است هر عددی بین ۷ تا ۳۰ ثانیه باشد.فرد مشخصی که راه و رسم آداب و عادت خاصی در زندگی داشته باشد طول عمرش ممکن است عددی بین صفر تا صد سال باشد.

در یک سیستم چند برنامه‌ای (Multiprograming) با توجه به تعداد کارهای I/O و CPU، زمان دقیق اجرای برنامه به‌طور قطعی قابل پیش‌بینی نیست. همچنین در یک سیستم سخت‌افزاری زمان بین دوخرابی مدولها (مثلاً مدولهای حافظه) دقیقاً مشخص نمی‌باشد.

در هر یک از این حالات نتیجه ویژه‌ای راکه ممکن است مشاهده شود نمی‌توان پیش بینی کرد، اما مجموعه تمام نتایج را می‌توان در نظر گرفت. وقتی برای یک آزمایشی مدل احتمال درست می‌کنیم منظور ما این است که:

۱- مجموعه تمام نتایج چه می‌توان باشد؟

۲- براساس تجزیه و تحلیل آزمایش، فراوانی نسبی وقوع این نتایج چه می‌تواند باشد؟

۱-۲ فضای نمونه

فضای نمونه یک آزمایش عبارتست از مجموعه تمام نتایجی که ممکن است براین آزمایش مترتب شوند. اینک چند مثال می‌آوریم که درآنها این تعریف مورد استفاده قرارمی گیرد.

مثال ۱-۱

تاسی را یک بار می‌ریزیم فضای نمونه این آزمایش عبارتست از مجموعه {۱،۲،۳،۴،۵،۶}=S

مثال ۱-۲

در یک سیستم سادهٔ کامپیوتری که دارای حافظهٔ اصلی ،ثبات ، حافظه سریع (cache) و دیسک می‌باشد. نتایج ممکنِ دسترسی به اطلاعات مورد نظر توسط CPU را فضای نمونهٔ {دیسک و حافظه اصلی، ثبات، حافظهٔ سریع }=S خواهد داشت.

مثال ۱-۳

در یک سیستم جستجوی کتابخانه‌ای هوشمند، اگر حداقل زمان پاسخ گارانتی شده ۱۰ثانیه باشدبنابراین فضای نمونه زمان دسترسی به اطلاعات درخواستی عبارتست از $S=\{t\in R | t\geq ۱۰\}$

۱-۳ پیشامد

پیشامد زیر مجموعه‌ای از فضای نمونه است. یک پیشامد زمانی روی می‌دهد که یکی از عضوهایش نتیجه آزمایش باشد.

مثال ۱-۴

در مثال ریختن تاس هر یک از مجموعه ذیل پیشامد می باشند.

$C= \{۲\}$ $B= \{۴،۵،۶\}$ $A=\{۱\}$

مثال ۱-۵

در مثال سیستم جستجوی کتابخانه هوشمند مجموعه $A= \{t: ۱۰ \leq t \leq ۲۰\}$ یک پیشامد زمان پاسخ سیستم می‌باشد.

مثال ۱-۶

در یک سیستم مالتی پروسسور اشتراک زمانی، برای دسترسی به اطلاعات حافظهٔ مشترک که از M مدولهای تشکیل یافته، فضای نمونه $S= \{M_۰,M_۱,....M_{m-۱}\}$ می‌باشد و دسترسی به اطلاعات مدولهای $A=\{M_۲\}$ و $B= \{M_{m-۲}\}$ پیشامدهایی از فضای مورد نظر می‌باشند.

۱-۴ اصول احتمال

همانطور که در این بخش خواهیم دید، نظریهٔ احتمال با اعداد مربوط به پیشامدها (زیر مجموعه‌های فضای نمونه) سروکار دارد. این اعداد احتمال وقوع این پیشامدها نامیده می‌شوند.

تابع احتمال، یک تابع مجموعه‌ای حقیقی است که روی کلاس تمام زیر مجموعه‌های فضای نمونه S

تعریف می‌شود. مقداری که به زیر مجموعه معین A اطلاق می‌گردد با نماد P(A) نشان داده می‌شود. تابع مجموعه‌ای را، تابع احتمال می‌نامیم به طوریکه دارای سه دستور ذیل باشد:

۱- $P(A) \geq 0, \forall\ A \subset S$

۲- $P(S)=1$

۳- $P(A_1 \cup A_2 \cup \ldots)=P(A_1)+P(A_2)+\ldots$ هرگاه $A_i \cap A_j=\phi, \forall\ i \neq j$

مثال ۱-۷

یک مجموعه ۱۰۰ تایی از برنامه‌های کامپیوتری، برای یافتن خطاهای (Errors) گوناگون، مورد بررسی قرار می‌گیرد. ۲۰ برنامه دارای خطای دستوری (Syntax)، ۱۰ برنامه خطای ورودی / خروجی (I/O)، ۵ برنامه خطاهای مختلف، ۶ برنامه شامل هر دو خطای دستوری و I/O می‌باشند. همچنین ۳ برنامه شامل خطای دستوری و دیگر خطاهای مختلف و ۲ برنامه شامل خطای I/O و خطاهای مختلف دیگر می‌باشد. یک برنامه نیز شامل هر نوع خطا می‌باشد.

یک برنامه به طور تصادفی از این مجموعه انتخاب می‌گردد. فرض کنیم S پیشامد مربوط به برنامه‌های دارای خطای دستوری و I پیشامد مربوط به برنامه‌های دارای خطای I/O و O پیشامد مربوط به حل برنامه‌های دارای خطاهای مختلف،باشد. در این صورت احتمالات زیر را محاسبه می‌نماییم.

حل:

الف- احتمال اینکه برنامه شامل خطای دستوری یا خطای I/O یا هر دو خطا شود:

$$P(S \cup I)=P(S)+P(I)-P(I \cap S)=\frac{20}{100}+\frac{10}{100}-\frac{6}{100}=\frac{24}{100}=\frac{6}{25}$$

ب -احتمال اینکه برنامه خطا داشته باشد (هر خطایی):

$$P(S \cup I \cup O)=P(S)+P(I)+P(O)-P(S \cap O)-P(I \cap O)-P(S \cap I)$$

$$+P(I \cap S \cap O)=\frac{20}{100}+\frac{10}{100}+\frac{5}{100}-\frac{6}{100}-\frac{3}{100}-\frac{2}{100}+\frac{1}{100}=\frac{25}{100}=\frac{1}{4}$$

ج- احتمال آنکه برنامه خطا نداشته باشد

$$P(N)=1-\frac{1}{4}=\frac{3}{4}$$

احتمالات پیشامدهای مختلف مثال فوق را در جدول زیر در نظر می‌گیریم:

پیشامد	S	I	O	$S \cap I$	$S \cap O$	$I \cap O$	$S \cap I \cap O$
احتمال	$\frac{20}{100}$	$\frac{10}{100}$	$\frac{5}{100}$	$\frac{6}{100}$	$\frac{3}{100}$	$\frac{2}{100}$	$\frac{1}{100}$

۱-۵ فنون شمارش

جوابهای بسیاری از مسائل احتمال بستگی به امکان شمردن تعداد عضوهای متعلق به مجموعه‌های معین دارند. تعدادی از فنون شمردن، کمک‌های با ارزشی به شمارش تعداد عضوهای مجموعه‌های

معین می‌کنند. در نتیجه در حل مسائلی از احتمالات که شامل این مجموعه‌ها باشند قابل استفاده‌اند. اولین فن، که تا اندازه‌ای کلی می‌باشد روش بسیار ساده‌ای است که اغلب به عنوان قانون حاصلضرب به آن اشاره می‌شود. تعریف این روش چنین است:

اگر عملی به n_1 روش و عمل دیگری به n_2 روش انجام پذیر باشد، هر دو عمل (دوّمی بلافاصله بعد از اولی) به $n_2 \times n_1$ روش انجام پذیر خواهد بود.

دومین فن که به آن ترتیب می‌گویند به صورت زیر بیان می‌گردد که تعداد جایگشت‌های k تا از n تا عنصر بدون تکرار برابر است. $P(n,k)= \frac{n!}{(n-k)!}$

سومین فن که به آن ترکیب می‌گویند، انتخاب k عنصر از n عنصر با هر ترکیب عبارتست از:

$$C(n,k)= \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

مثال ۱-۸

فرض می‌کنیم ۵ ترمینال توسط یک خط ارتباطی به یک سیستم کامپیوتر مرکزی متصل شده‌اند. خط ارتباطی مرتب به ۵ ترمینال سرکشی می‌کند تا ترمینالی که آمادگی ارتباط با سیستم مرکزی را دارد، مرتبط نماید. فضای نمونه‌ای که وضعیت سیستم را مشخص می‌کند عبارتست از: $S=\{x_1,x_2,x_3,x_4,x_5\}$

که x_i ۰ یا ۱ می‌باشد. $x_i = 1$ معرف آمادگی ترمینال iام جهت برقراری ارتباط و $x_i = 0$ معرف عدم آمادگی ترمینال iام جهت ارتباط با سیستم مرکزی می‌باشد.

مثلاً مجموعه {۰،۱،۱،۰،۰} گویای آن است که ترمینالهای ۲ و ۳ آماده ارتباط و ترمینالهای ۱، ۴ و ۵ آماده ارتباط نیستند. بر اساس قانون حاصلضرب تعداد نمونه‌ها برابر $۳۲ = ۲^{۵}$ می‌باشد. حال فرض می‌کنیم دقیقاً ۳ ترمینال از ۵ ترمینال آمادگی ارتباط داشته باشند در این صورت ترکیب زیر را خواهیم داشت:

حل

$$\binom{۵}{۳} = \frac{۵!}{۳!\ (۵-۳)\ !} = ۱۰$$

۱-۶ احتمال شرطی

در بعضی از موارد عملی، به ما اعلام می‌شود که پیشامد A رخ داده ولی مطلوبست احتمال رخ دادن پیشامد B. فرض می‌کنیم پس از انجام آزمایش طبی شخصی که به بیماری دچار شده، در شرح حال خانواده شخص مرض قند سابقه داشته است. بنابراین احتمال اینکه این شخص نیز مرض قند داشته باشد چقدر است؟ در این مثال وقوع یک پیشامد A اعلام می‌گردد و سئوال می‌شود احتمال وقوع پیشامد B چقدر است.

در این راستا به تعریف احتمال مشروط وقوع B به فرض آنکه A رخ داده باشد می‌پردازیم.

$$P[A \mid B]= \begin{cases} \dfrac{P(A\cap B)}{P(B)} & \text{اگر } P(B)>0 \\ 0 & \text{اگر } P(B)=0 \end{cases}$$

مثال ۱-۹

با توجه به مثال ۱ بخش ۴-۱، اگر پیشامدهای A و B را به صورت ذیل تعریف نمائیم در این صورت احتمال شرطی $P(A \mid B)$ برابر خواهد بود:

حل ۱:

A : پیشامدی که برنامه حداقل یک خطای I/O دارد.

B : پیشامدی که برنامه حداقل یک خطای دستوری دارد.

$$P(A\cap B)=\frac{۶}{۱۰۰}$$

$$P(B)=\frac{۲۰}{۱۰۰} \qquad P(A\mid B)=\frac{\frac{۶}{۱۰۰}}{\frac{۲۰}{۱۰۰}}=\frac{۳}{۱۰}$$

مثال ۱-۱۰

فرض کنید در بررسی ۱۰۰ کامپیوتر در یک مرکز کامپیوتری، حداقل ۷۵ کامپیوتر داری مدل ۴۸۶ می‌باشند. اگر سه کامپیوتر بطور تصادفی یکی پس از دیگری انتخاب گردد (بدون جایگزینی) احتمال آنکه هر سه دارای مدل ۴۸۶ باشد چقدر است؟

حل ۲: فرض کنید A_1، A_2 و A_3 پیشامدهای اولین، دومین و سومین انتخابی که دارای مدل ۴۸۶ است، تعریف می‌گردد. احتمال مورد نظر برابر است با:

$$P[A_1\cap A_2\cap A_3]=P(A_1)\times P(A_2 \mid A_1)\times P(A_3 \mid A_1\cap A_2)$$

با توجه به قانون حاصلضرب این مقدار برابر است با:

$$\frac{۷۵}{۱۰۰}\times\frac{۷۴}{۹۹}\times\frac{۷۳}{۹۸}=۰/۴۱۸$$

۲باتوجه به قانون احتمال کل، فرض کنیم $A_1,A_2,\dots,A_n$ پیشامدهایی باشند به طوریکه:

۱- $A_i\cap A_j=\phi$ برای $i\neq j$ (پیشامدهای دو به دو ناسازگار)

۲- $P(A_i)>0$ برای $i=1,2,\dots,n$

۳- $A_1\cup A_2\cup\dots A_n=S$

بنابراین برای هر پیشامد A داریم [1]:

$$P(A)=P(A_1)\times P(A\mid A_1)+P(A_2)\times P(A\mid A_2)+\dots\dots+P(A_n)\times P(A\mid A_n)$$

A: عبارتست از مجموعه پیشامدهایی که شرایط ۳،۲،۱ را دارا و افرازی (Partition) از S می‌باشد (Partition).

مثال ۱-۱۱

در خواستهایی به یک سیستم کامپیوتری درون خطی (On-line) از طریق ۵ خط ارتباطی می‌رسد. درصد پیغامهای دریافتی از خطوط ۱، ۲، ۳، ۴ و ۵ به ترتیب برابر است با ۲۰، ۳۰، ۱۰، ۱۵ و ۲۵. احتمال آنکه طول یک پیغام بیشتر از ۱۰۰ کاراکتر باشد به ترتیب خطوط برابر با ۰/۴، ۰/۶، ۰/۲، ۰/۸ و ۰/۹ می‌باشد. احتمال آنکه یک پیغام دریافتی بیشتر از ۱۰۰ کاراکتر باشد به طریق زیر محاسبه می‌گردد:

حل ۳:

پیشامدی که پیغام انتخاب شده بیشتر از ۱۰۰ کاراکتر دارد =A

پیشامدی که پیغام در خط i دریافت شده است =A_i

بنابراین با توجه به قانون احتمال کل داریم:

$$P(A)= P(A_1)\times P(A|A_1)+....+P(A_5)\times P(A|A_5)$$

$$=0/2\times0/4+0/3\times0/6+0/1\times0/2+0/15\times0/8+0/25\times0/9=0/625$$

۱-۷ متغیر تصادفی

نظریه احتمال، خود شاخه جالب توجهی از ریاضیات است. این نظریه را می‌توان در زمینه‌های مختلف به کاربرد. یکی از مهمترین موارد کاربرد آن در توصیف متغیرهای تصادفی می‌باشد.

تعریف: متغیر تصادفی X، یک تابع حقیقی از عضوهای فضای نمونه S است. متغیر تصادفی که مقادیر قابل شمارش به خود اختصاص دهد، متغیر تصادفی گسسته نامیده می‌شود. همچنین متغیر تصادفی که مقادیر پیوسته را در برداشته باشد، متغیر تصادفی پیوسته نامیده می‌شود.

مثال ۱-۱۲

فرض کنید متغیر تصادفی X تعداد خطوط ارتباط فعال در یک سیستم کامپیوتر درون خطی تعریف می‌گردد. فضای نمونه عبارتست از مجموعه‌های nتائی $(x_1, x_2, ..., x_n)$ بطوریکه:

$$x_i = \begin{cases} 1 & \text{اگر خط i فعال باشد} \\ 0 & \text{اگر خط i فعال نباشد} \end{cases}$$

مثال ۱-۱۳

متغیر تصادفی که طول عمر دیسک‌های سخت یک سیستم کامپیوتری را بیان می‌کند یک متغیر تصادفی پیوسته می‌باشد.

۱-۸ تابع توزیع و تابع چگالی

تعریف: تابع توزیع برای یک متغیر تصادفی X (که با $F_X(t)$ نشان داده می‌شود) تابعی است از یک متغیر حقیقی t چنانچه:

۱- قلمرو تعریف F_X تمام محور حقیقی است

۲- برای هر عدد حقیقی $F_X(t)= P(X\leq t)$,t

لازم به ذکر است که تابع توزیع $F_X(t)$ تجمع کلی احتمالات را برای هر مقدار ارائه می‌دهد. تابع توزیع چنان که خواهیم دید ابزار نیرومندی، در بسیاری از موارد عملی می‌باشد.

خواص تابع توزیع [۲]:

$$۱-\lim_{x \to +\infty} F(x)=۱$$

$$۲-\lim_{x \to -\infty} F(X)=۰$$

۳- برای تابع غیر نزولی F داریم:

$$F(x) \le F(Y),\ \forall x \le y$$

برای $x < y$ می‌توان نوشت:

$$P(X \le Y)=F(Y)-F(X)$$

تعریف: تابع چگالی احتمال برای یک متغیر تصادفی پیوسته X (که با $f_X(x)$ نشان داده می‌شود) تابعی از یک متغیر حقیقی x می‌باشد به طوریکه:

۱- قلمرو $f_X(۰)$ تمام محور حقیقی است

۲- برای هر عدد حقیقی t

$$F_x(t)=\int_{-\infty}^{t} f_x(x)\,dx$$

بنابراین تابع چگالی را می‌توان به صورت زیر نوشت:

$$f_x(t)=\frac{d}{dt}\left[F_x(t)\right]$$

برای متغیر تصادفی گسسته X تابع $P(a)=P(X=a)$ تعریف می‌گردد که a عددی از یک مجموعه قابل شمارش می‌باشد.

۱-۹ امید ریاضی و واریانس

اگر X یک متغیر تصادفی گسسته باشد می‌توانیم با استفاده از تابع توزیع و با بکاربردن تابع احتمال (P_x)، پیشامدهای مربوط به X را محاسبه کنیم.

هرگاه X متغیر تصادفی پیوسته‌ای باشد، می‌توان تابع توزیع $F(x)$ یا تابع چگالی $f(x)$ را برای محاسبه احتمال پیشامدهای مربوط به X کار برد.

اغلب مسائلی مطرح می‌شوند که تنها به بیان احتمال وقوع X در فاصله معینی احتیاج نیست بلکه نیاز به تعیین مقدار متوسط X دارند.

مثال ۱-۱۴

میانگین زمان بین دو خرابی در یک سیستم کامپیوتری متأثر از خطوط ارتباطی، سخت افزار، نرم افزار و... قابل توجه می‌باشد. همچنین میانگین زمان تعمیر خرابی‌های مذکور در طراحی سیستمهای برنامه ریزی تعمیر و نگهداری از اهمیت خاصی برخوردار است.

تعریف: اگر X متغیر تصادفی گسسته با تابع احتمالی $P(x)$ باشد امید ریاضی $E(x)$ به صورت زیر خواهد بود.

$$E[X]=\sum_{x=i} x_i\ P(x_i)=x_۱ P(x_۱)+x_۲ P(x_۲)+\ldots\ldots\ldots$$

حال اگر x متغیر تصادفی پیوسته با تابع چگالی $f(x)$ باشد، امید ریاضی به صورت زیر تعریف می‌گردد:

$$E(X)=\int_{-\infty}^{+\infty} x\, f(x)\,dx$$

میانگین یک متغیر تصادفی اندازه‌ای برای نقطه میانی توزیع احتمال متغیر تصادفی را ارائه می‌کند. همانطور که خواهیم دید، در تمام کاربردهای اساسی علاقمندیم مقداری نیز برای تغییر پذیری توزیع احتمال (متغیر تصادفی) داشته باشیم. متداول‌ترین مقدار تغییر پذیری یک متغیر تصادفی، مقدار متوسط مربع، فاصله بین متغیر تصادفی و میانگینش می‌باشد.

اگر تغییر X نسبتاً کم باشد، X همواره نزدیک میانگین است و این مقدار متوسط فاصله، وقتی که مجذور شود، نسبتاً کوچک خواهد بود.

اگر X زیاد تغییرکند، گاهی با میانگینش به مقداری نسبتاً بزرگ، اختلاف خواهد داشت و مقدار متوسط فاصله وقتی مربع شود زیاد خواهد بود. این مقدار متوسط "انحراف مربع شده" را "واریانس" متغیر تصادفی می‌نامند. جذر واریانس "انحراف معیار" متغیر تصادفی می‌باشد. برای متغیرهای تصادفی گسسته و پیوسته، واریانس به ترتیب زیر تعریف می‌گردد:

$$\sigma^{۲}=\mathrm{Var}\,(x) = \sum_{x_i} \left(x_i-E\,(X)\right)^{۲} P\,(x_i)$$

$$\sigma^{۲}=\mathrm{Var}\,(x) = \int_{-\infty}^{\infty} \left(x-E\,(X)\right)^{۲} f\,(x)\,dx$$

که σ_X انحراف معیار متغیر تصادفی X نامیده می‌شود.

مثال ۱-۱۵

در مثال ۱ بخش ۵-۱، متغیر تصادفی گسسته X را برای تعداد سرکشی‌های (Polling) سیستم برای یافتن ترمینال آماده ارتباط، تعریف می‌کنیم. فرض می‌کنیم که X فقط مقادیر ۱، ۲ و ۳ را اختیار می‌کند. تابع احتمال به صورت: $P(۱)=۰/۶$ $P(۲)=۰/۳$ $P(۳)=۰/۱$

برای مقادیر مختلف متغیر تصادفی تعریف می‌شود. احتمال ۲ بار سرکشی و کمتر عبارتست از:

$$F(۲)=P(۱)+P(۲)=۰/۹$$

به راحتی دیده می‌شود:

$$\mu=E\,(x) = \sum_{i=۱}^{۳} x_i P\,(x_i) = ۱\times۰/۶+۲\times۰/۳+۳\times۰/۱=۱/۵$$

$$\sigma^{۲}=\mathrm{Var}\,(x) = \sum_{i=۱}^{۳} \left(x_i-۱/۵\right)^{۲} P\,(x_i)$$

$$= (۱-۱/۵)^{۲}\times۰/۶+(۲-۱/۵)^{۲}\times۰/۳+(۳-۱/۵)^{۲}\times۰/۱=۰/۴۵$$

۱-۱۰ متغیرهای تصادفی چند بعدی

بسیاری از مسائل مهم ایجاب می‌کند که چند متغیر تصادفی همزمان با هم در نظر گرفته شوند. در این صورت می‌خواهیم عمل مشترک چنین متغیرهای تصادفی را مطالعه کنیم. اگر چه این موضوع خارج از بحث کتاب می‌باشد ولی امیدواریم با ساده‌ترین نوع، یعنی متغیرهای تصادفی دوبعدی با ذکر یک مثال

این موضوع را به پایان برسانیم.

تعریف: هرگاه آزمایشی را در نظر بگیریم، زوج (X,Y) را یک متغیر تصادفی دوبعدی گویند،اگر هر یک از (X,Y) به هر عضو S (فضای نمونه) یک عدد حقیقی منحصر بفرد نسبت دهند.

برای متغیرهای تصادفی گسسته و پیوسته به ترتیب ذیل داریم:

$$F(x,y) = \sum_{x_i \le x} \sum_{y_j \le y} P(x_i, y_j)$$

$$F(u,v) = \int_{-\infty}^{v} \int_{-\infty}^{u} f(x,y)\, dx dy$$

مثال ۱-۱۷

با توجه به مثال ۲ بخش قبلی (۹-۱)، فرض می‌کنیم ۲ ترمینال از ۴ ترمینال موجود آمادگی ارتباط (جهت انتقال) دارند. ترکیب‌های مختلف حالت آمادگی انتقال سیستم به صورت زیر می‌باشد.

(۱,۱,۰,۰),(۱,۰,۱,۰),(۱,۰,۰,۱),(۰,۱,۱,۰),(۰,۱,۰,۱),(۰,۰,۱,۱)

که ۱ معرف آمادگی جهت انتقال و ۰ معرف عدم آمادگی می‌باشد.

متغیرهای تصادفی Y,X را به صورت زیر تعریف می‌نمائیم.

تعداد سرکشی جهت یافتن اولین ترمینال آماده جهت انتقال $=X$

تعداد سرکشی جهت یافتن دومین ترمینال آماده جهت انتقال $=Y$

بنابراین داریم:

$$P(۱,۲)=P(X=۱,Y=۲)=\frac{۱}{۶}$$
$$P(۲,۲)=P(۳,۳)=P(۳,۲)=۰$$

همچنین داریم:

$$P_X(۱)=P(۱,۲)+P(۱,۳)+P(۱,۴)=\frac{۱}{۲}$$
$$P_X(۲)=P(۲,۳)+P(۲,۴)=\frac{۱}{۳}$$

واضح است که Y مقادیر ۲، ۳و ۴ را فقط می‌تواند در برگیرد. بنابراین جدول زیر را خواهیم داشت:

جدول ۱-۱ احتمال سرکشی جهت یافتن ترمینال آماده

X \ Y	۲	۳	۴	تابع کناری P_X
۱	$\frac{۱}{۶}$	$\frac{۱}{۶}$	$\frac{۱}{۶}$	$\frac{۱}{۲}$
۲	.	$\frac{۱}{۶}$	$\frac{۱}{۶}$	$\frac{۱}{۳}$
۳	.	.	$\frac{۱}{۶}$	$\frac{۱}{۶}$
تابع کناری P_Y	$\frac{۱}{۶}$	$\frac{۱}{۳}$	$\frac{۱}{۲}$	

۲

توزیع های احتمال

توزیعهای احتمال در حل بسیاری از مسائل مختلف کاربرد فراوان دارند. ما در این فصل چندین توزیع احتمال را همراه مثالهای مختلف بررسی خواهیم نمود.

۲-۱ توزیع برنولی

اصولاً آزمایش برنولی دارای ۲ نتیجه پیروزی با احتمال p و شکست با احتمال ۱-p می‌باشد. اگر آزمایش برنولی n بار تکرار گردد در این صورت به بررسی توزیع دو جمله‌ای با پارامترهای n و p منجر می‌شود.

۲-۲ توزیع دو جمله‌ای

هر گاه در n تکرار آزمایش k بار موفقیت با احتمال p صورت گیرد، خواهیم داشت:

$$P\ (x=k)\ =\ \binom{n}{k}\ P^{k}q^{n-k}=b\ (k,n,p)$$

که مقدار q برابر ۱-p می‌باشد

مثال ۲-۱

یک کمپانی تجاری دارای ۲۰ خط ارتباطی مستقل می‌باشد. فرض می‌کنیم احتمال مشغول بودن هر خط ۰/۶ باشد، در این صورت احتمال اینکه حداقل ۱۰ خط مشغول باشد چقدر است؟

حل:

متغیر تصادفی دو جمله‌ای x را به عنوان تعداد خطوط مشغول با پارامترهای n=۲۰ و p=۰/۶ در نظر

می‌گیریم. بنابر این احتمال مورد نظر عبارتست از:

$$P(x \geq ۱۰) = \sum_{x=۱۰}^{۲۰} \begin{bmatrix} ۲۰ \\ x \end{bmatrix} (۰/۶)^{x} * (۰/۴)^{۲۰-x} = ۰/۸۷$$

مثال ۲-۲

در یک فایل اصلی (Master) ۱۲۰،۰۰۰ رکورد را به صورت ترتیبی (Sequential) ذخیره کرده‌ایم. فایل از بلاکهای ۶ رکوردی تشکیل شده است. یک فایل تراکنشی (transaction) هر روز جهت بهنگام (Update) شدن فایل اصلی اجرا می‌شود. این امر باعث می‌شود که حدود ۵٪ از کل رکوردهای فایل اصلی بهنگام شوند. اگر یکی از رکوردهای بلاک نیاز به بهنگام شدن داشته باشد کل بلاک باید بهنگام شود. در ضمن تمامی رکوردها به صورت یکسان در فایل اصلی پراکنده شده‌اند. میانگین و انحراف معیار تعداد بلاکهایی که باید بهنگام شوند چقدر است؟

حل ۲:

X را متغیر تصادفی در نظر می‌گیریم که تعداد رکوردهای بهنگام شده یک بلاک را معرفی می‌کند. می‌توان توزیع X را دو جمله‌ای با پارامترهای p=۰/۰۵ و n=۶ در نظر گرفت (چون یک رکورد یا باید بهنگام شود یا نشود). بنابراین اگر X≥۱ باشد یک بلاک باید بهنگام شود که داریم:

$$P(X \geq ۱) = ۱ - P(X=۰)$$

برای بهنگام شدن یک بلاک

$$۱ - b(۰،۶،۰/۰۵) = ۱ - (۰/۹۵)^{۶} = ۱ - ۰/۷۳۵ = ۰/۲۶۵$$

اکنون متغیر تصادفی Y را تعداد بلاک‌هایی که نیاز به بهنگام دارند در نظر می‌گیریم. Y نیز دارای توزیع دو جمله‌ای با پارامترهای n=۲۰۰۰۰ بلاک و p=۰/۲۶۵ می‌باشد. بنابراین میانگین بلاکهای که باید بهنگام شوند عبارتست از:

$$E(Y) = ۲۰/۰۰۰ \times ۰/۲۶۵ = ۵۳۰۰$$

انحراف معیاری برابر:

$$\sigma = (۲۰،۰۰۰ \times ۰/۲۶۵ \times ۰/۷۳۵)^{\frac{۱}{۲}} = (۳۸۹۵ - ۵)^{\frac{۱}{۲}} = ۶۲/۴۱۴$$

خواهیم داشت.

۳-۲ توزیع هندسی

فرض کنید که آزمایشهای برنولی را انجام دهیم و بدانیم که احتمال موفقیت در هر آزمایش p می‌باشد. اگر X تعداد آزمایشهای لازم برای بدست آوردن اولین موفقیت باشد X را متغیر تصادفی هندسی با پارامتر p (احتمال موفقیت) می‌نامیم و تابع احتمال آن به صورت زیر می‌باشد:

$$p_x(K) = p.q^{k-۱}, \quad K = ۱،۲،۳،۴،...$$

میانگین و واریانس این نوع متغیر تصادفی برابر است با:

$$E(x) = \frac{q}{p}, \quad Var(X) = \frac{q}{p^{۲}}$$

متغیر تصادفی هندسی در تئوری صف و آمار کاربردی مورد استفاده قرار می‌گیرد[۱].

مثال ۲-۳

با توجه به مثال ۲ بخش قبلی X را متغیر تصادفی در نظر می‌گیریم که معرف تعداد بلاک‌هایی که قبل از رسیدن به اولین بلاکی که نیاز به بهنگام دارد، از فایل اصلی خوانده می‌شوند. X دارای توزیع هندسی با پارامترهای P=۰/۲۶۵ ،میانگین $\frac{q}{p}=$ ۲/۷۷۴ و انحراف معیار $\frac{\sqrt{q}}{p}=$ ۲/۲۳۵ بلاک، می‌باشد.

۲-۴ توزیع پواسن

بطوریکه خواهیم دید متغیر تصادفی پواسن کاربردهای عملی فراوانی دارد. قبل از اینکه درباره این توزیع بحث نمائیم، به بررسی آنچه "فرآیند پواسن" نامیده می‌شود می‌پردازیم. فرآیند قابل شمارش $\{N(t), t \geq ۰\}$ یک فرآیند پواسن بانرخ $۰ < \lambda$ خواهد بود اگر موارد ذیل رعایت گردند.

۱- در فرآیند، رخداد پیشامدها مستقل از یکدیگر باشند.

۲- توزیع تعداد پیشامدها در هر فاصله زمانی فقط به طول فاصله بستگی دارد.

۳- احتمال آنکه دقیقاً یک حادثه در هر فاصله زمانی "h" اتفاق بیفتد برابر است با:

$$P(N(h)=۱)=\lambda h+O(h)$$

۴- احتمال آنکه بیش از یک حادثه در هر فاصله زمانی h اتفاق بیفتد عبارتست از:

$$P(N(h)\geq ۲)=O(h)$$

تابع f برابر o(h) اگر رابطه زیر برقرار باشد:

$$\lim_{h\to 0}\frac{f(h)}{h}=O(h)$$

فرض کنید متغیر تصادفی X تعداد اتفاقات در فاصله t، دارای توزیع پواسن با پارامتر λ باشد. تابع توزیع این متغیر را چنین می نویسیم:

$$P(X=K)=e^{-\lambda t}\frac{(\lambda t)^k}{k!}=P_k(t),\ k=۰,۱,۲,...$$

میانگین و واریانس متغیر تصادفی که دارای توزیع پواسن باشد عبارتست از λ

مثال ۲-۴

فرض کنید تعداد دستورهای ویرایش (Edit) که در ترمینال‌های یک سیستم کامپیوتر مرکزی تایپ می‌شود (ارسال) دارای فرآیند پواسن با پارامتر λ برابر ۱۲۰ دستور در ساعت باشد.

حل:

فرض کنید X تعداد دستورهای ویرایش که در یک دوره یک دقیقه‌ای وارد می‌شوند. در این صورت X یک متغیر تصادفی پواسن با پارامتر ۲ =λو:

$$P_x(k)=\frac{۲^k}{k!}e^{-۲},\ k=۰,۱,۲,...$$

احتمال اینکه در این فاصله یک دقیقه‌ای دستوری نیاید برابر است با:

$$P_x(۰)=e^{-۲}=۰/۱۳۵۳$$

احتمال اینکه بین ۱ و ۵ (شامل ۱ و ۵) دستور ویرایشی بیاید، برابر است با:

$$P(1 \le X \le 5) = F_x(5,2) - F_x(0,2) = 0/9834 - 0/1353 = 0/8481$$

هرگاه سه دوره یک دقیقه‌ای مجزای متوالی رادرنظر بگیریم، احتمال بین ۱و۵ورود دستور (شامل ۱و۵) در هر فاصله یک دقیقه‌ای برابر $0/61 = (0/8481)^3$ است. زیرا تعداد دستورات ویرایشی که در این فواصل یک دقیقه‌ای وارد می‌شوند، مستقل هستند. احتمال بین ۱و۵ (شامل ۱و۵) در میان ۲ فاصله از ۳ فاصله برابر است با:

$$\binom{3}{2} (0/8481)^2 (0/1519) = 0/3278$$

۲-۵ توزیع نمائی

متغیر تصادفی پیوسته X دارای توزیع‌نمائی با پارامتر $\lambda > 0$ می‌باشد اگر تابع چگالی آن به صورت زیر تعریف می‌گردد:

$$f(x) = \begin{cases} \lambda . e^{-\lambda x} & x > 0 \\ 0 & x \le 0 \end{cases}$$

تابع توزیع احتمال F(x) به صورت زیر بیان می‌گردد:

$$F(x) = \begin{cases} 1 - e^{-\lambda x} = 1 - \exp(-x / E(X)) & x > 0 \\ 0 & x \le 0 \end{cases}$$

خواص توزیع نمائی به شرح ذیل می‌باشند[1]:

۱- $E(x) = \frac{1}{\lambda}$ و $Var(x) = \frac{1}{\lambda^2}$

۲- یکی از دلایل اهمیت توزیع نمائی در سیستم‌های صف خاصیت مارکفی آن می‌باشد. گاهی اوقات این خاصیت را بدون حافظه می‌نامند.

$$P(X > t + h \mid x > t) = P(X > h), t > 0, h > 0$$

۳- گشتاور kام عبارتست از:

$$E(X^k) = \frac{K!}{\lambda!} = k!\, E(X)^k \qquad ,K = 1,2,3...$$

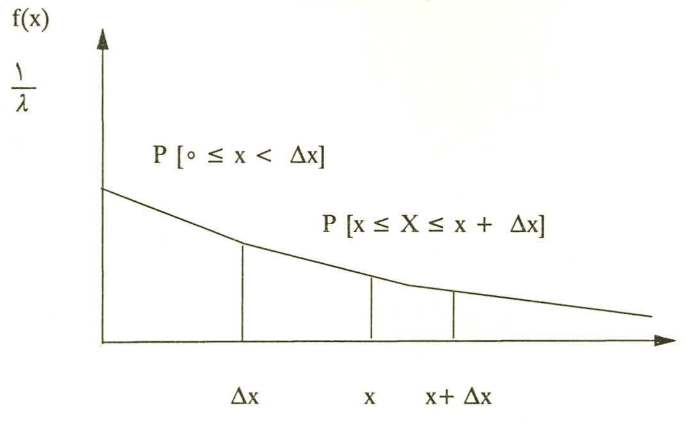
۴- مقدار r-امین درصد $\prod(r)$ که به صورت $P(X \le \prod(r)) = \frac{r}{100}$ تعریف می‌گردد، برابر است با:

$$\prod(r) = E(X)\, L_n\left(\frac{100}{100 - r}\right)$$

۵- F(x) یک تابع نزولی می‌باشد. (برای $x \ge 0$) نتیجه این مشخصه آن است که احتمال اینکه X در فواصل کوتاهتر واقع شود، بیشتر است یعنی

$$P\{0 \le X \le \Delta x\} > P\{x \le X \le x + \Delta x\} \qquad \Delta x \text{ و } x > 0$$

این مشخصه در شکل ۲-۱ کاملاً مشاهده میگردد. توزیع نمائی در مدلهای صف به دو دلیل دارای



شکل ۲-۱ توزیع نمائی

اهمیت فراوان می‌باشد. نخست آنکه، فرض مربوط به این توزیع از نقطه نظر ریاضی و محاسباتی ساده می‌باشد.دوم آنکه فرض این توزیع به اندازه کافی واقعی و بدین ترتیب مدل در حد قابل قبولی سیستم را پیش‌بینی می‌نماید.

۶- اگرفاصله زمانی بین دو واقعه (۲ ورود یا ۲ اتمام سرویس پیاپی) دارای توزیع نمائی با پارامتر λ باشد، تعداد وقوع واقعه در یک فاصله زمانی دارای توزیع پواسن خواهدبود. اگر تعداد وقوع پیشامد را در فاصله t بامتغیر تصادفی $X(t)$ نشان دهیم، خواهیم داشت:

$$P(X(t)=n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \quad \text{و} \quad n = ۰,۱,۲,...$$

به عبارت دیگر $X(t)$ دارای توزیع پواسن با پارامتر λt می‌باشد.

مثال ۲-۵

کارکنان یک شرکت کامپیوتری جهت محاسبات مهندسی خود از یک ترمینال استفاده می‌کنند. اگر زمانی که هر کارمند از ترمینال استفاده کند، دارای توزیع نمائی با مقدار میانگین ۳۶ دقیقه باشد، موارد ذیل رامحاسبه می‌نمائیم:

۱- احتمال اینکه یک کارمند کمتر از ۳۰ دقیقه ترمینال را اشغال نماید؟

حل:

فرض کنید T زمانی که یک کارمند صرف استفاده از ترمینال می‌کند باشد. احتمال اینکه T حداکثر t دقیقه افزایش یابد برابر است با:

$$F(t) = \begin{cases} 1 - e^{-\frac{t}{E(T)}} & , t>0 \\ 0 & t\leq 0 \end{cases} \quad , \quad F(t) = 1 - e^{-\frac{t}{36}}$$

$$P(T\leq 30) = 1 - e^{\frac{-30}{36}} = 1 - 0/4346 = 0/5654$$

۲- احتمال اینکه یک کارمند بیش از یک ساعت ترمینال را اشغال نماید؟

حل:

$$P(t>60) = e^{\frac{-60}{36}} = 0/1889$$

۳- اگر یک کارمند قبلاً ۳۰ دقیقه ترمینال را مشغول کرده باشد. احتمال آنکه یک ساعت دیگر ترمینال را مشغول کند چقدر است؟

حل: باتوجه به خاصیت مارکف (بدون حافظه) در حقیقت زمانی که شخص برای ۳۰ دقیقه از ترمینال استفاده می‌کند هیچ اثری بر احتمال اینکه حداقل یک ساعت دیگر از آن استفاده کند، ندارد. بنابراین احتمال این حالت برابر است با ۰/۱۸۸۹ .

۴- اگر ۹۰٪ زمانی که از ترمینال استفاده می‌کند کمتر از R دقیقه باشد، مقدار R چقدراست؟

حل: بنابر خاصیت توزیع نمائی داریم [۱]:

$$\pi(90) = E(X)\ Ln\left(\frac{100}{100-90}\right) = 2/3\,E(X) = 2/3\times 36 = 82/8 \text{ دقیقه}$$

۲-۶ توزیع یکنواخت

در این قسمت راجع به متغیر تصادفی پیوسته یکنواخت که اغلب در عمل با آن سر و کار داریم بحث می‌کنیم. متغیر تصادفی X روی فاصله [a,b] بطور یکنواخت توزیع می‌گردد اگر تابع چگالی و توزیع آن به ترتیب ذیل باشند:

$$f(x) = \begin{cases} \frac{1}{b-a} & a<x<b \\ 0 & \text{سایر موارد} \end{cases}$$

$$F(x) = \begin{cases} 0 & x<a \\ \frac{x-a}{b-a} & a\leq x<b \\ 1 & x\geq b \end{cases}$$

میانگین و واریانس این متغیر تصادفی به ترتیب برابر است با:

$$E(x) = \int_b^a x f(x)dx = \frac{a+b}{2}$$

$$Var(x) = \int_b^a (x-E(x))^2 f(x)dx = \frac{(b-a)^2}{12}$$

مثال ۲-۶

فرض کنیم در یک سیستم کامپیوتری، مدت زمان یک دوره چرخش دیسک ۰/۰۲۵ ثانیه طول می‌کشد. برای خواندن و نوشتن در یک رکورد خاص واقع در قسمتی از دیسک، لازم است دیسک مدتی چرخش کند تا رکورد مورد نظر مقابل خواننده دستگاه قرار بگیرد. زمان تأخیر چرخش مذکور دارای توزیع یکنواخت بین ۰ تا ۰/۰۲۵ ثانیه می‌باشد. مطلوبست میانگین، واریانس و احتمال آنکه تأخیر چرخش مذکور بین ۰/۰۰۵ تا ۰/۰۱۵ ثانیه باشد

حل:

$$E(T)=\frac{a+b}{۲} = \frac{۰+۲۵}{۲}=۱۲/۵ \quad \text{میکرو ثانیه}$$

$$Var(T)=\frac{(b-a)^{۲}}{۱۲}=\frac{۲۵^{۲}}{۱۲}=۵۲/۰۸۲۳ \quad \text{میکرو ثانیه}$$

$$P(۵\leq T\leq ۱۵) = \frac{۱۵-۵}{۲۵-۰}=\frac{۱۰}{۲۵}=۰/۴$$

۲-۷ توزیع نرمال

از میان تمام توابع احتمال تابع چگالی نرمال بیشتر کاربرد دارد. بدلیل آنکه متغیر تصادفی نرمال در مسائل عملی دیده می‌شود. در ضمن وسیله‌ای برای تقریب دقیق تعداد زیادی از قانونهای احتمال، فراهم می‌سازد. تابع چگالی برای یک متغیر تصادفی نرمال متقارن، دارای منحنی نمایشی به شکل ناقوس است. در نظر افراد این منحنی محک مناسبی برای بکار بردن در امتحاناتی که در آنها از روی منحنی نمره داده می‌شود، می‌باشد. چون منحنی نمایش تابع چگالی به شکل ناقوس است متغیر تصادفی به طور نرمال توزیع می‌شود. مقادیر نزدیک به μ را با بیشترین احتمال و، متناظراً مقادیر دور از μ(در دو طرف) را با کمترین احتمال قبول می‌کند.

تعریف: یک متغیر تصادفی X دارای توزیع نرمال است اگر و فقط اگر تابع چگالی احتمال آن به صورت زیر باشد :

$$f_X(x) = = \frac{۱}{\sigma\sqrt{۲\pi}} e^{\frac{-(x-\mu)^{۲}}{۲\sigma^{۲}}} \qquad \forall x \in R$$

μ می‌تواند هر عدد حقیقی باشد اما σ باید مثبت باشد.

تعریف: هر گاه Z یک متغیر تصادفی نرمال با $\mu = ۰$ و $\sigma^{۲} = ۱$ باشد آنرا متغیر تصادفی نرمال استاندارد می‌نامند. لازم به ذکر است که $Var(x)= \sigma^{۲}$، $E(x)=\mu$ می‌باشد.

فرض کنیم متغیرهای تصادفی $X_1,X_2,...,X_n$ مستقل از یکدیگر باشند. به طوریکه هر یک از آنها دارای توزیع نرمال بامیانگین‌های به ترتیب $\mu_۱,\mu_۲,....,\mu_n$ و همچنین واریانسهای $\sigma_۱^{۲}\sigma^{۲}_۲...\sigma^{۲}_n$ می‌باشند. در این صورت متغیر تصادفی Y که برابر است با $Y = y_۱+y_۲+....y_n$ دارای توزیع نرمال با میانگین $\mu = \mu_۱ + \mu_۲ + ... + \mu_n$ و واریانس $\sigma^{۲}= \sigma^{۲}_۱ +\sigma^{۲}_۲+.....+\sigma^{۲}_n$ می‌باشد.

مثال ۲-۷

فرض کنید تعداد پیغامهای بافر شده در یک سیستم کامپیوتری که با متغیر تصادفی X بیان می‌گردد، دارای

توزیع نرمال با میانگین ۱۰۰ و انحراف معیار ۱۰ باشد. احتمال تعداد بافرهای استفاده شده در دو حالت ذیل محاسبه کنید.

۱- بیشتر از ۱۲۰ بافر نباشد

۲- بیشتر از ۱۳۰ بافر باشد

حل: اگر متغیر Z را به صورت $Z = \frac{x-\mu}{\sigma}$ تعریف نمائیم بنابراین :

۱- $$Z = \frac{۱۲۰-۱۰۰}{۱۰} = ۲$$

مقدار $P(Z \leq 2)$ از طریق جدول تابع توزیع نرمال برابر ۰/۹۷۷۲۵ می‌باشد.

۲- با توجه به جدول مذکور

$$P(X > ۱۳۰) = P(Z > ۳) = ۱-۰/۹۹۸۶۵ = ۰/۰۰۱۳۵$$

۸-۲ توزیع گاما

یکی از توزیعهایی که در مهندسی کاربرد دارد، توزیع گاما می‌باشد. متغیر تصادفی پیوسته X دارای توزیع گاما با پارامترهای $\alpha > ۰$ و $\lambda > ۰$ می‌باشد اگر تابع چگالی آن به صورت زیر تعریف گردد:

$$f(x)=\begin{cases} \dfrac{\lambda(\lambda x)^{\alpha-۱} e^{-\lambda x}}{\Gamma(\alpha)} & , \quad x>۰ \\ ۰ & , \quad x\leq ۰ \end{cases}$$

تابع گاما $\Gamma(۰)$ به صورت زیر تعریف می‌گردد:

$$\Gamma(t) = \int_{۰}^{\infty} x^{t-۱} e^{-x} dx \quad \text{و} \quad t>۰$$

همچنین در رابطه این تابع موارد ذیل برقرار است [۳]:

۱- $\Gamma(n + ۱) = n!, n = ۰,۱,۲,...$

۲- $\Gamma(t + ۱) = t\Gamma(t), t>۰$

۳- اگر $x_۱,x_۲,...x_n$ متغیرهای تصادفی مستقل و دارای توزیع نمائی باشند(با پارامتر λ)، مجموع آنها (Y) دارای توزیع گاما با پارامترهای n و λ خواهد بود همچنین تابع توزیع Y برابراست با [۱]:

$$F_y(x)=G_n(x)=۱-e^{\lambda X}\left\{ ۱+\lambda x+\frac{(\lambda x)^۲}{۲!}+........+\frac{(\lambda x)^{n-۱}}{(n-۱)!}\right\} \quad , \quad x\geq ۰$$

مثال ۲-۸

در یک سیستم کامپیوتری ، زمان بین درخواستهای مشتریان (X) دارای توزیع نمائی با میانگین ۱۰ ثانیه باشد. اگر متغیر تصادفی T را به صورت زمان سپری شده ۵ در خواست اول، تعریف کنیم در این صورت موارد ذیل را محاسبه می‌نمائیم:

۱- مقدار میانگین و واریانس T

۲- احتمال اینکه T از ۶۰ ثانیه بیشتر نشود

۳- احتمال اینکه T بیش از ۹۰ ثانیه باشد

حل:

در نظر می گیریم: $T = X_1 + X_2 + X_3 + X_4 + X_5$

که در آن X_i , $i = 1, ..., 5$ مستقل و دارای توزیع نمائی با میانگین ۱۰ می‌باشند بنابراین متغیر T دارای توزیع گاما با پارامترهای $\alpha = 5$, $\lambda = \frac{1}{10}$ می‌باشد. پس بر اساس تابع گاما[۱]:

حل ۱: $E(t) = \frac{\alpha}{\lambda} = 50$ ثانیه و $Var(T) = \frac{\alpha}{\lambda^2} = 500$

حل ۲: $P(T \le 60) = G_5(60) = 1 - e^{-6}\left(1 + 6 + \frac{6^2}{2!} + \frac{6^3}{3!} + \frac{6^4}{4!}\right) = 0/7149$

حل ۳: $P(T > 90) = 1 - P(T \le 90) = e^{-9}\left(1 + 9 + \frac{9^2}{2!} + \frac{9^3}{3!} + \frac{9^4}{4!}\right) = 0/055$

۲-۹ توزیع ارلانگ Erlang

متغیر تصادفی T دارای توزیع ارلانگ با پارامترهای k و λ می‌باشد اگر T دارای توزیع گاما با تابع چگالی زیر باشد. توزیع ارلانگ یک فرم خاصی از توزیع گاما می‌باشد[۱].

$$f(t) = \begin{cases} \frac{\lambda k (\lambda k t)^{k-1}}{(k-1)!} e^{-\lambda k t} & t > 0 \\ 0 & t \le 0 \end{cases}$$

توزیع ارلانگ در حقیقت یک سرویس k مرحله‌ای می‌باشد که تمامی مراحل آن مستقل و یکسان هستند. هر یک از مراحل مذکور دارای توزیع نمائی باپارامتر λ می‌باشد. اگر کل زمان سرویس را به صورت T در نظر بگیریم در این صورت T دارای توزیع ارلانگ باپارامتر λk خواهدبود. بنابراین امید ریاضی و واریانس و تابع توزیع ارلانگ به صورت زیر می‌باشد:

$$E(T) = \frac{k}{k\lambda} = \frac{1}{\lambda} \quad , Var(T) = \frac{k}{(k\lambda)^2} = \frac{1}{(k\lambda)^2} = \frac{E(T)^2}{k^2}$$

$$F(t) = P(T \le t) = 1 - e^{-\lambda k t}\left[1 + \frac{\lambda k t}{1!} + \frac{(\lambda k t)^2}{2!} + \dots\dots + \frac{(\lambda k t)^{k-1}}{(k-1)!}\right]$$

مثال ۲-۹

فرآیند تعمیر قطعات یک مرکز کامپیوتری، به ۵ مرحله مستقل تقسیم می‌گردد.زمان تعمیر برای هر یک از این مراحل دارای توزیع نمائی با میانگین ۱۰ دقیقه می‌باشد. موارد ذیل را محاسبه می‌نمائیم.

۱- احتمال اینکه تعمیر کار در حداکثر یک ساعت فعالیت خود را به انجام برساند.

۲- احتمال اینکه تعمیر کار بیش از ۹۰ دقیقه صرف زمان تعمیر نکند

حل ۱: زمان تعمیر T دارای توزیع ارلانگ با k = ۵ مرحله و میانگین ۵۰ دقیقه می‌باشد.

$$E(T) = 10 + 10 + 10 + 10 + 10 = 50 \quad \text{دقیقه}$$

$$Var(T) = \frac{E(T)^2}{k} = \frac{2500}{5} = 500 \quad \text{دقیقه}$$

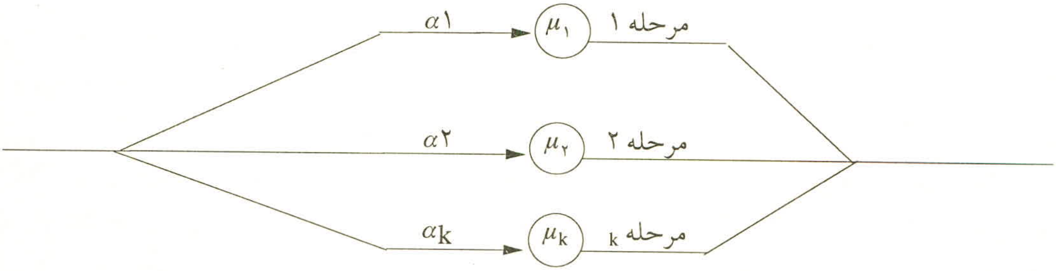
$$F(60) = 1 - e^{-\frac{60}{10}}\left[1 + \frac{60}{10} + \frac{1}{2}\left(\frac{60}{10}\right)^2 + \frac{1}{6}\left(\frac{60}{10}\right)^3 + \frac{1}{24}\left(\frac{60}{10}\right)^4\right] = 0/7149$$

حل ۲:

$$F(90) = 1 - e^{\frac{-90}{10}}\left[1 + 9 + \frac{9^2}{2} + \frac{9^3}{6} + \frac{9^4}{25}\right] = 0/945$$

۲-۱۰ توزیع hyperexponential [1]

اگر زمان سرویس در یک سیستم صفی دارای انحراف معیار زیادی نسبت به مقدار میانگین آن باشد، در این صورت توزیع زمان سرویس به توزیع .hyperexp تقریب زده می‌شود. همچنین اگر انحراف معیار کمتر از میانگین باشد، توزیع زمان سرویس به توزیع .hypoexp تقریب زده می‌شود. مدل زیر، شکل ۲-۲، توزیع hyperexp با k مرحله را معرفی می‌نماید.



شکل ۲-۲ مدل k مرحله‌ای توزیع .hyperexp

هر یک از مراحل دارای توزیع نمائی با پارامتر $\mu_i, i = 1, \ldots, k$ می‌باشد. یک متقاضی جهت دریافت سرویس با احتمال $\alpha_i, i = 1, \ldots k$ وارد یکی از مراحل می‌شود.

لازم به ذکر است که

$$\sum_{i=1}^{k} \alpha_i = 1$$

تابع چگالی، میانگین و واریانس این توزیع به ترتیب زیر می‌باشد:

$$f_S(t) = \sum_{i=1}^{k} \alpha_i \mu_i e^{-\mu_i t}$$

$$E(S) = \sum_{i=1}^{k} \left(\frac{\alpha_i}{\mu_i}\right), \quad E\left(S^2\right) = 2\sum_{i=1}^{k} \frac{\alpha_i}{\mu_i^2}$$

$$Var(S) = E\left(S^2\right) - E(S)^2$$

S: متغیر زمان سرویس

مثال ۲-۱۰

در یک سیستم کامپیوتری اطلاعات به صورت بلاکهای ۱۰ کلمه‌ای در حافظه اصلی ذخیره شده‌اند. در این سیستم فرض می‌کنیم کنترل کننده دسترسی مستقیم به حافظه (Direet Mercry Accen control) (DMAC) و CPU همزمان به اطلاعات هر بلاک حافظه جهت سرویس (پردازش) دسترسی داشته باشند. بر اساس اطلاعات موجود، ۳ کلمه از هر بلاک توسط DMAC و ۷ کلمه توسط CPU مورد دسترسی قرار می‌گیرند. زمان سرویس مذکور توسط CPU و DMAC برای اطلاعات هر بلاک دارای توزیع نمائی با میانگین‌های ۱ و ۳ ثانیه به ترتیب می‌باشد.

۱- احتمال اینکه یک بلاک از اطلاعات ، حداکثر ۵ ثانیه نیاز به سرویس مذکور داشته باشد چقدر است؟

۲- میانگین زمان سرویس (پردازش) برای یک بلاک چقدر می‌باشد؟

حل ۱:

$\alpha_۲ =$ احتمال دسترسی DMAC به اطلاعات هر بلاک $= ۰/۳$

$\alpha_۲ =$ احتمال دسترسی CPU به اطلاعات هر بلاک $= ۰/۷$

$$P(S\leq t) = ۱-\alpha_۱ e^{-\mu_۱ t}-\alpha_۲ e^{-\mu_۲ t} \quad , \mu_۱=\frac{۱}{۳}$$

$$P(S\leq ۵) = ۱-۰/۳\, e^{\frac{-۵}{۳}}-۰/۷ e^{-۵} \quad , \quad \mu_۲=۱$$

$$= ۱-(۲/۰۲\times ۱۰^{-۳}+۴/۷۱\times ۱۰^{-۳})$$

$$= ۱ - ۰/۰۰۶۷ = ۰/۹۹۳۳$$

حل ۲:

$$E(S) = \left[\frac{۰/۳}{\frac{۱}{۳}}\right] + \left[\frac{۰/۷}{۱}\right] = ۱/۶ \text{ ثانیه}$$

۳

زنجیره های مارکف و دامنه کاربرد آنها

۳-۱ فرآیندهای تصادفی

فرآیند تصادفی به فرآیندی اطلاق می‌شود که قابل پیش بینی نبوده و در طول زمان تغییر می‌کند. مثلاً تعداد برنامه‌هایی که در یک سیستم کامپیوتری، در صف CPU منتظر دریافت سرویس می‌باشند، یک فرآیند تصادفی است. در طول زمان تعداد برنامه‌ها افزایش یافته و با توجه به سرعت CPU تعداد آنها دائم تغییر می‌کند. فرآیند تصادفی به صورت ریاضی نیز تعریف می‌شود.

تعریف: یک مجموعه از متغیرهای تصادفی $\{X(t), t \in T\}$ را فرآیند تصادفی می‌گویند. اگر مجموعه T قابل شمارش باشد، فرآیند را فرآیند تصادفی گسسته و اگر عناصر T متعلق به مجموعه پیوسته‌ای از اعداد حقیقی باشند، آن را فرآیند تصادفی پیوسته می‌نامند.

تعریف: فضای حالت (State Space) فرآیند عبارتست از مجموعه مقادیری که متغیر تصادفی $X(t)$ به خود اختصاص می‌دهد به عبارت دیگر $\{X(t), t \in T\}$.

مثال ۳-۱

فرض کنید مجموعه $\{X_n, n = ۱,...,۳۶۵\}$ تعداد برنامه‌های اجرا شده در یک مرکز کامپیوتری در روز nاُم سال باشد. این فرآیند بصورت فضای حالت گسسته با پارامتر گسسته می‌باشد.

مثال ۳-۲

زمان انتظار یک درخواست در یک سیستم کامپیوتر مرکزی، تا شروع اجرای آن یک فرآیند تصادفی می‌باشد $\{W(t), t \geq o\}$. در این صورت زمان ورود درخواست به سیستم `t` یک پارامتر پیوسته و همچنین فضای

حالت نیز پیوسته می‌باشد.

۳-۲ فرآیند تولد و مرگ

این فرآیند علاوه بر آنکه یکی از ابتدائی‌ترین شکل مدلهای صف می‌باشد، فرموله کردن وبررسی تحلیلی آن، امکان اثبات فرمولهای مدلهای دیگر صف را فراهم می‌سازد.

این فرآیند در تئوری احتمالات و فرآیندهای تصادفی کاربردهای گوناگونی دارد. ولی در این کتاب فقط از لحاظ مدلهای صف مورد بررسی قرار می‌گیرد.

در این فرآیند منظور از تولد، ورود برنامه (مشتری و...)به سیستم و مرگ به معنی خروج برنامه (مشتریان و...) از سیستم می‌باشد.

وضعیت سیستم با تعداد برنامه‌های (مشتری و...) سیستم یعنی N در زمان T به صورت $N(T)$ نشان داده می‌شود. در فرآیند تولد و مرگ، بررسی می‌شود که متغیر تصادفی $N(T)$ چگونه با زمان تغییر پیدا می‌کنند. در حالی که می‌دانیم تولد (ورود) و مرگ (خروج) به طور تصادفی، بوقوع می‌پیوندند. در مطالعه فرآیند ورود به، و خروج از سیستم موارد ذیل را در نظر می‌گیریم:

الف -اگر وضعیت سیستم $N(t)=n$ باشد، مدت زمان باقی مانده برای آنکه ورود بعدی اتفاق بیافتد، دارای توزیع نمائی با پارامتر λ_n , $n = ۰,۱,۲,...$ می‌باشد.

ب-اگر وضعیت سیستم $N(t) = n$ باشد، مدت زمان باقی مانده برای آنکه خروج بعدی اتفاق بیافتد، دارای توزیع نمائی با پارامتر μ_n , $n = ۱,۲,...$ می‌باشد.

ج- در هر زمان تنها یک ورود یا یک خروج می‌تواند رخ دهد.

طبق رابطه توزیع نمائی و توزیع پواسن که قبلاً اشاره کردیم (فصل ۲)، λ_nو μ_nبه ترتیب میانگین نرخ ورود و خروج، در فرآیند می‌باشند. طبیعی است که اگر مشتری (برنامه و...) در سیستم باشد با یک ورود (تولد) وضعیت سیستم از n به $n+1$ انتقال می‌یابد و اگر وضعیت سیستم n باشد، یک خروج (مرگ) وضعیت سیستم را به $n-1$ انتقال می‌دهد. شکل زیر فرآیند ورود و خروج (تولد و مرگ) رابیان می‌کند. بطوریکه P_n عبارتست از احتمال این که سیستم در وضعیت n باشد.

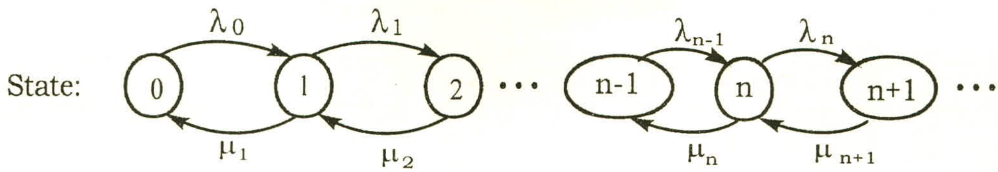
بدیهی است که تعیین مشخصات سیستم هنگامی که در حالت انتقالی یا گذرا (transient)، باشد غالباً مشکل است. در حالی که بعد از آن که سیستم به یک وضعیت یکنواخت (steady state) رسید، می‌توان آن را به ترتیب زیر مورد بررسی قرار داد.

فرض کنید که سیستم در وضعیت n باشد. در این صورت تعداد دفعاتی که فرآیند وارد وضعیت n و از وضعیت n خارج می‌شود را بررسی می‌نمائیم . چون ورود / خروج در وضعیت n به صورت یکی در میان اتفاق می‌افتد، پس تعداد دفعات ورود و خروج یا باهم مساوی و یا آنکه ۱ واحد باهم اختلاف دارند. این اختلاف در مواردی که تعداد دفعات ورود / خروج سیستم زیاد باشد ناچیز است.

بر اساس این بررسی، به یک اصل در فرآیند تولد و مرگ می‌رسیم که عبارتست از اصل تساوی نرخ ورود و خروج.

Rate in = Rate out

بر اساس تساوی مذکور برای تمام وضعیتهای ممکن، احتمال اینکه دقیقاً n مشتری در سیستم باشد را



شکل ۳-۱ فرآیند ورود و خروج

میتوان محاسبه نمود.

برای توضیح بیشتر فرض کنید سیستم در وضعیت صفر است. یعنی هیچ مشتری در سیستم وجود ندارد. طبق شکل ۳-۱ سیستم فقط از وضعیت یک می‌تواند وارد وضعیت صفر بشود (یعنی یک مشتری در سیستم باشد و این مشتری از سیستم خارج شود).

بنابراین اگر احتمال اینکه سیسستم در وضعیت یک باشد P_1 بنامیم و میانگین نرخ خروج از آن و ورود به وضعیت صفر دارای نرخ μ_1 باشد داریم:

$\mu_1 P_1 =$ میانگین نرخ ورود به وضعیت صفر

همینطور برای خروج از وضعیت صفر، می‌بایست یک ورود (تولد) اتفاق بیافتد. احتمال اینکه سیستم در وضعیت صفر باشد P_0 و نرخ ورود در وضعیت صفر برابر λ_0 است. پس میانگین نرخ خروج از وضعیت صفر عبارتست از:

$\lambda_0 P_0 =$ میانگین نرخ خروج از وضعیت صفر

بنابراین معادله تعادل (Balance Equatioon) برای وضعیت صفر عبارتست از: $\mu_1 P_1 = \lambda_0 P_0$

به همین ترتیب برای وضعیت یک داریم:

$\lambda_0 P_0 + \mu_2 P_2 =$ میانگین نرخ ورود به وضعیت یک

$\mu_1 P_1 + \lambda_1 P_1 = (\mu_1 + \lambda_1) P_1 =$ میانگین نرخ خروج از وضعیت یک

بنابراین معادله تعادل وضعیت یک به صورت زیر بیان می‌شود:

$$\lambda_0 P_0 + \mu_2 P_2 = (\mu_1 + \lambda_1) P_1$$

با همین استدلال می‌تواند جدول ۳-۱ را برای وضعیت‌های مختلف تشکیل داد:

جدول ۳-۱ معادلات انتقال وضعیتها

شماره‌وضعیت	معادله‌وضعیت نرخ خروج = نرخ ورود
۰	$\mu_1 P_1 = \lambda_0 P_0$
۱	$\lambda_0 P_0 + \mu_2 P_2 = (\mu_1 + \lambda_1) P_1$
۲	$\lambda_1 P_1 + \mu_3 P_3 = (\mu_2 + \lambda_2) P_2$
⋮	⋮
$n-1$	$\lambda_{n-2} P_{n-2} + \mu_n P_n = (\lambda_{n-1} + \mu_{n-1}) P_{n-1}$

جدول ۳-۱ (ادامه)

n	$\lambda_{n-1}P_{n-1} + \mu_{n+1}P_{n+1} = (\lambda_n + \mu_n)P_n$
⋮	⋮

ملاحظه می‌شود که معادله اول (معادله وضعیت صفر) دارای دو احتمال P_0 ، P_1 و معادله دوم شامل سه احتمال P_0، P_1 و P_2 می‌باشد. بنابراین تمام معادلات را می‌توان برحسب احتمال P_0 بدست آورد. توضیح فوق را به زبان ریاضی زیر میتوان فرموله نمود:

احتمال مورد نظر	وضعیت
$P_1 = \frac{\lambda_0}{\mu_1} P_0$	۰
$P_2 = \frac{\lambda_1}{\mu_2} P_1 + \frac{1}{\mu_2}(\mu_1 P_1 - \lambda_0 P_0) = \frac{\lambda_1}{\mu_2} P_1 = \frac{\lambda_1 \lambda_0}{\mu_2 \mu_1} P_0$	۱
	⋮
$P_n = \frac{\lambda_{n-1}}{\mu_n} P_{n-1} + \frac{1}{\mu_n}(\mu_{n-1} P_{n-1} - \lambda_{n-2} P_{n-2}) = \frac{\lambda_{n-1}}{\mu_n} P_{n-1} = \frac{\lambda_{n-1} \cdots\cdots \lambda_0}{\mu_n \cdots\cdots \mu_1} P_0$	n−۱
$P_{n+1} = \frac{\lambda_n}{\mu_{n+1}} P_n + \frac{1}{\mu_{n+1}}(\mu_n P_n - \lambda_{n-1} P_{n-1}) = \frac{\lambda_n}{\mu_{n+1}} P_n = \frac{\lambda_n \cdot \lambda_{n-1} \cdots\cdots \lambda_0}{\mu_{n+1}\mu_n \cdots\cdots \mu_1} P_0$	n

برای تسهیل بیان ریاضی فرض کنید:

$$R_n = \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0}{\mu_n \mu_{n-1}\cdots\mu_1} \qquad n = 0,1,2,\ldots$$

از روابط فوق نتیجه گیری می شودکه:

$$P_n = R_n P_0$$

چون میدانیم که مجموع احتمال حالات ممکن باید برابر یک شود یعنی:

$$\sum_{n=0}^{\infty} P_n = 1$$

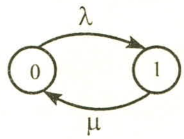
بنابراین رابطه زیر نتیجه گیری می شود:

$$P_0 = 1\left(1 + \sum_{n=1}^{\infty} R_n\right) \quad \text{و یا} \quad P_0 = \frac{1}{1 + \sum_{n=1}^{\infty} R_n}$$

مثال ۳-۳

در یک سیستم کامپیوتری وضعیت سرویس CPU به دو صورت تعریف می‌گردد: مشغول (BUSY) و بیکار (idel). فرآیند ورود برنامه‌ها به سیستم دارای توزیع پواسن با پارامتر λ و مدت زمان دریافت سرویس CPU، دارای توزیع نمائی با پارامتر μ می باشد.

دیاگرام وضعیت انتقال سیستم برای این مدل تولد و مرگ به صورت زیر است:



لازم به ذکر است که وضعیت صفر به معنی بیکار بودن CPU و وضعیت یک به معنی مشغول بودن آن می‌باشد.

بنابر قانون احتمال داریم: (۱) $P_۰ + P_۱ = ۱$

معادله مربوط به وضعیت‌های سیستم به صورت زیر می‌باشد:

$$\mu P_۱ = \lambda_۰ P_۰$$

بنابراین:

$$P_۱ = (\frac{\lambda}{\mu}).P_۰$$

پس معادله (۱) را می‌توان به صورت زیر بیان نمود:

$$P_۰ + P_۱ = P_۰ (۱ + \frac{\lambda}{\mu})$$

بنابراین:

$$P_۰ = \frac{\mu}{(\lambda + \mu)}$$

و

$$P_۱ = \frac{\lambda}{\mu} P_۰ = \frac{\lambda}{(\lambda + \mu)}$$

۳-۳ زنجیره‌های مارکف

فرآیند تصادفی $\{X(t_i), t_i \in T\}$ را یک فرآیند مارکف گوئیم، اگر برای زمانهای مختلف $t_۱ < t_۲ < ... < t_n < t_{n+۱}$ و $n+۱$ حالت مختلف (وضعیت) رابطه ذیل برقرار باشد:

$$P[X(t_{n+۱})=x_{n+۱} | X(t_۱)=x_۱, X(t_۲)=x_۲, ..., X(t_n)=x_n] = P[X(t_{n+۱})=x_{n+۱} | X(t_n)=x_n]$$

معادله فوق نشان می‌دهد که آینده فرآیند فقط به حالت فعلی وابسته می‌باشد و ارتباطی به گذشته فرآیند ندارد. این مشخصه که قبلاً نیز بحث شد، به خاصیت بدون حافظه معروف است. فرآیندهای مارکف در جدول ذیل دسته بندی می‌گردند:

با توجه به جدول ۳-۲، فرآیند مارکف را زنجیره مارکف گوییم که فضای حالت (وضعیت) به صورت گسسته باشد.

فرض کنید زنجیره مارکف با زمان گسسته $\{X_n \mid X_n = X(t_n)\}$ باشد در این صورت داریم:

جدول ۲-۲ فرآیندهای مارکف

فضای حالت / زمان	گسسته	پیوسته
زمان‌گسسته	زنجیره‌مارکف زمان‌گسسته	فرآیندمارکف زمان گسسته
زمان پیوسته	زنجیره مارکف زمان پیوسته	فرآیند مارکف زمان پیوسته

$$P[X_{n+1} = X_m | X_n = X_n \ldots, X_0 = X_0] = P[X_{n+1} = X_m | X_n = X_n]$$

اگر زنجیر در حالت i در زمان n باشد، بنابراین احتمال انتقال یک مرحله‌ای از وضعیت i به وضعیت j عبارتست از : $i,j,n = 0$ و ۱ و ۲ و...

$$P_{ij}(n) = P[X_{n+1} = j \mid X_n = i]$$

اگر زنجیر مارکف برای احتمالات انتقال یک مرحله‌ای، مستقل از زمان گسسته n باشد در این صورت احتمالات انتقال به صورت P_{ij} نوشته می‌شود. این نوع زنجیر مارکف را همگن در زمان می‌نامند.

احتمالات انتقال یک مرحله‌ای P_{ij} به صورت ماتریس زیر بیان می‌شود. ماتریس انتقال یک ماتریس تصادفی است که مجموع عناصر هر سطر برابر یک می‌باشد:

$$\begin{matrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ P_{i0} & P_{i1} & P_{i2} & \cdots \\ \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdots \end{matrix}$$

به طوریکه :

$$\sum_{j=0}^{\infty} P_{ij} = 1 \quad P_{ij} \geq 0 \quad i, j = 0, 1, 2, \ldots$$

۳-۴ معادلات چپمن -کولموگرف

در بخش قبلی احتمالات انتقال یک مرحله‌ای P_{ij} را بررسی کردیم. در این بخش احتمالات انتقال چند مرحله‌ای P_{ij}^{n} که عبارتست از احتمال اینکه فرآیند در وضعیت i باشد و پس از n انتقال به وضعیت j برسد، را به صورت ذیل تعریف می‌کنیم :

$$P_{ij}^{n} = P\{X_{n+m} = j | X_m = i\} \quad m,n \geq 0 \quad i,j \geq 0$$

همچنین داریم:

$$P_{ij}^{0} = \begin{cases} 1 & i=j \quad \text{اگر} \\ 0 & i \neq j \quad \text{اگر} \end{cases}$$

بنابراین احتمالات انتقال n مرحله‌ای توسط معادله چپمن کولموگرف محاسبه می‌شود:

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty} P_{ik}^{n} P_{kj}^{m} \qquad \text{برای همه } n,m,i,j \geq 0$$

بویژه $n = ۲,۳,\ldots\ldots\ldots$

$$P_{ij}^{n} = \sum_{k=۰}^{\infty} P_{ik}^{(n-۱)} P_{kj}$$

مثال ۳-۴

در یک شبکه انتقال پیغام (به صورت Paketهایی از ۰ و ۱) چندین مرحله جهت انتقال Paket داریم. احتمال آنکه یک رقم ۰ یا ۱ به درستی (همان رقم) به مرحله بعدی انتقال یابد ۰/۷۵ می‌باشد. در این صورت احتمال آنکه یک رقم ۰ در ۴ مرحله بعد نیز ۰ در یافت شود چقدراست؟

حل: در این مسئله یافتن $P_{۰۰}^{۴}$ مطلوب می‌باشد بنابراین ماتریس احتمالات انتقال عبارتست از:

$$P = \begin{bmatrix} ۰/۷۵ & ۰/۲۵ \\ ۰/۲۵ & ۰/۷۵ \end{bmatrix}$$

بنابراین:

$$P^{۲} = \begin{bmatrix} ۰/۶۲۵ & ۰/۳۷۵ \\ ۰/۳۷۵ & ۰/۶۲۵ \end{bmatrix}, \; P^{۴} = P^{۲} \cdot P^{۲} = \begin{bmatrix} ۰/۵۳۱۲۵ & ۰/۴۶۸۷۵ \\ ۰/۴۶۸۷۵ & ۰/۵۳۱۲۵ \end{bmatrix}$$

بنابراین احتمال آنکه یک رقم ۰ در ۴ مرحله بعد نیز ۰ دریافت شود: $P_{۰۰}^{۴} = ۰/۵۳۱۲۵$ در مورد زنجیر مارکف تعاریف ذیل مطرح می‌باشند [۴].

تعریف ۱: هر گاه هر یک از وضعیتهای یک زنجیر مارکف، از تمامی وضعیتهای دیگر قابل دستیابی باشد، زنجیر را تحویل ناپذیر گوییم.

تعریف ۲: اگر وضعیت jاز وضعیت i وهمچنین i از وضعیت jقابل دستیابی باشد این دو وضعیت را متبادل می‌نامیم. رابطه تبادل بین دو وضعیت i,jرابا $j \leftrightarrow i$ نشان می‌دهیم. بر اساس تعریف فوق خواص رابطه تبادل عبارتست از:

۱- انعکاسی : $i \leftrightarrow i$، چون

$$P_{ij}^{0} = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$$

۲- تقارن :اگر $j \leftrightarrow i$ آنگاه $i \leftrightarrow j$

۳- شرکت پذیری :اگر $j \leftrightarrow i$ و $k \leftrightarrow j$ آنگاه $i \leftrightarrow k$

تعریف ۳: وضعیت iرا متناوب و دارای دوره تناوب $d>۱$ گویند اگر:

۱- $P_{ii}^{n} = ۰$ مگر آنکه $n = md$ برای یک مقدار صحیح و مثبت d.

۲- dبزرگترین عددی می‌باشد که شرط (۱) را دارا باشد $(d>۱)$

اگر $d=۱$ باشد وضعیت iرا غیر متناوب می‌نامیم.

تعریف ۴: برای وضعیت i، احتمال اولین بازگشت به وضعیت iبعد از nانتقال خروج از وضعیت i رابه صورت زیر تعریف می‌کنیم:

$$f_i^{(n)} = P[\, X_n = i, X_k \neq i, k = ۱,۲,\ldots,n-۱ \;|\; X_۰ = i]$$

و:

$$f_i^{(n)} = ۱$$

تعریف ۵: احتمال آنکه زنجیر مارکف $\{X_n\}$، در n – امین مرحله در وضعیت j باشد، را به صورت زیر تعریف می‌کنیم:

$$\Pi_j = P[X_n = j]$$

دراین صورت $\Pi = (\Pi_1, \Pi_2, \ldots)$ را توزیع احتمال ایستائی می‌گوییم اگر رابطه $\Pi = \Pi P$ برقرار باشد.

مثال ۳-۵

با توجه به مثال ۱ که ماتریس احتمالهای انتقال آن به صورت زیر می‌باشد:

$$P = \begin{bmatrix} ۰/۷۵ & ۰/۲۵ \\ ۰/۲۵ & ۰/۷۵ \end{bmatrix}$$

زنجیر مارکف در این حالت تحویل ناپذیر وغیر متناوب می‌باشد. بنابراین برای محاسبه توزیع احتمال ایستائی، $\Pi = (\Pi_۰, \Pi_۱)$، داریم:

$$(_۰, \ \Pi_۱) \begin{bmatrix} ۰/۷۵ & ۰/۲۵ \\ ۰/۲۵ & ۰/۷۵ \end{bmatrix} = (\Pi_۰, \ \Pi_۱)$$

پس:

$$\begin{cases} ۰/۷۵\Pi_۰ + ۰/۲۵\Pi_۱ = \Pi_۰ \\ ۰/۲۵\Pi_۰ + ۰/۷۵\Pi_۱ = \Pi_۱ \end{cases}$$

باتوجه به این که $\Pi_۰ + \Pi_۱ = ۱$ می‌باشد بنابراین :

$$\Pi_۰ = ۰/۵ \quad , \quad \Pi_۱ = ۰/۵$$

۴ سیستم های صف

۴-۱ مقدمه

آیا می‌دانید برنامه شما در یک سیستم کامپیوتر مرکزی، که تعداد زیادی ترمینال به آن متصل است، چه مدت زمانی جهت دریافت سرویس از CPU یا دیسک یا ... انتظار می‌کشد؟ یا آنکه در یک شبکه کامپیوتری پیغام جهت ارسال به یک کامپیوتر مقصد چه مدت در صف انتظار (بافر) باقی می‌ماند؟

در دنیای تکنولوژی اطلاعات، هنگامی که یک پیغام از طریق شبکه اینترنت، ارسال می‌کنیم، با توجه به محدودیتهای گیتهای شبکه چه مدت زمانی طول می‌کشد تا پیغام به دریافت کننده برسد.

در تمام مثالهای فوق مسئله صف موجود بوده و مطالعه آن در مباحث مدرن امروز ضروری به نظر می‌رسد. همچنین کاربرد مسائل صف در سیستمهای اقتصادی، اجتماعی، صنعتی و بهداشتی اهمیت فوق‌العاده‌ای دارد که در ذیل به چند نمونه اشاره می‌کنیم:

۱ـ صف جلو گیشه بانک

۲ـ صف اتوبوس

۳ـ صف پرداخت پول در فروشگاه

۴ـ صف تحویل کالا به انبار

۵ـ صف باراندازی و بارگیری در بنادر

۶ـ صف بیماران جهت دریافت خدمات بهداشتی

٤-۲ عناصر تشکیل دهنده سیستم صف

۱- **صف (Queue):** عبارتست از یک گروه مشتری (برنامه و ..) که انتظار دریافت سرویس (CPU, I/O,....) را می‌کشند.

۲- **مشتری (Job):** عبارتست از شخص، برنامه یا... که انتظار سرویس (CPU، دیسک و...) را می‌کشد.

۳- **سرویس دهنده (Server):** عبارتست از عاملی که سرویس را فراهم می‌کند. مثلاً در یک سیستم کامپیوتری، CPU که برنامه جهت اجراء در پشت آن قرار می‌گیرد، را سرویس دهنده می‌نامیم.

٤-۳ مشخصات صف

بهترین روش در جهت بررسی صف، تجزیه و تحلیل مشخصات عناصر تشکیل دهنده صف می‌باشد.

۱- **مشتری:** فرآیند ورودی، زمان ورود مشتری به صف را نشان می‌دهد. این فرآیند به صورت گروهی و یا فردی می‌تواند صورت بگیرد. برای بهتر فهمیدن ورود مشتری به سیستم چهار حالت را در نظر می‌گیریم:

الف- مشتری صف را ترک می‌کند قبل از آنکه سرویس دریافت کند (RENEGING).

ب- مشتری به محض ورود، صف را ترک می‌کند (BALKING).

پ- مشتری حضور خود را از صفی به صف دیگر تغییر می‌دهد. (JOCKING).

ت- مشتری جهت دریافت سرویس انتظار می‌کشد. (WAITING).

مثلاً در یک سیستم کامپیوتری که کاربران از طریق ترمینالهای مختلف به یک کامپیوتر مرکزی متصل می‌باشند، کاربر پس از اتمام تایپ یک دستور سیستمی، قبل از قرار گرفتن در نوبت CPU، دستور مذکور را حذف نماید (BALKING).

۲- **سرویس دهنده:** فرآیند سرویس عبارتست از زمانیکه جهت سرویس مشتری منظور می‌گردد. زمان سرویس ممکن است ثابت و یا متغیر باشد. بحثی که در سرویس دهنده‌ها قابل اهمیت است مربوط به نوع آنها می‌باشد.

۳- **صف:** یکی از مشخصات مهم صف نظم آن می‌باشد که عبارتست از ترتیبی که سرویس در اختیار مشتری قرار می‌گیرد. به طور مثال می‌توان موارد زیر را در نظر گرفت.

الف- FCFS: تحت عنوان اولین ورود، اولین سرویس گیرنده.

ب- LCFS: تحت عنوان آخرین ورود، اولین سرویس گیرنده.

البتّه انواع تقدّم‌ها (priority) در سیستم‌های صف موجود می‌باشند که در این فصل در مورد آنها بحث و بررسی صورت خواهد گرفت.

متغییرهای اساسی در سیستم‌های صف عبارتند از:

الف- مدت انجام سرویس (service time): مدت زمانی که ارائه سرویس به یک مشتری طول می‌کشد.

پ- زمان ورود (arrival time) : زمانی که مشتری وارد سیستم می‌شود.

ت- زمان خروج (departure time) : زمانی که مشتری سیستم را ترک می‌کند.

ث- مدت انتظار (waiting time) : مدت زمانی که مشتری در صف منتظر می‌ماند تا سرویس دریافت کند.(این زمان شامل زمان انجام سرویس نمی‌باشد).

ج- مدت بین دو ورود (interarrival time) : فاصله زمانی بین ورود پیاپی دو مشتری

چ- مدت بین دو خروج (interdeparture time): فاصله زمانی بین خروج پیاپی دو مشتری

ح- طول صف (queueing lenght) : تعداد افرادی که در صف جهت دریافت سرویس منتظر می‌مانند.

خ- نرخ ورود (arrival rate) λ:تعداد ورود مشتری به سیستم در واحد زمان (ثانیه و ...)

د- نرخ سرویس (service rate) μ: نسبت سرویس دهی مشتری در واحد زمان (که در واقع همان تعداد خروج مشتری در واحد زمان می‌باشد).

ذ- شدت ترافیک (traffic intencity) ρ: نسبت نرخ ورود به نرخ سرویس دهی را شدت ترافیک گویند $(\rho=\frac{\lambda}{\mu})$.

اگر در سیستم صف، همه متغیرها معین باشند و عامل تصادفی در آنها وجود نداشته باشد، مسئله بدین صورت مطرح می‌شود که یک ورودی رأس موعد معینی وارد صف می‌شود و در زمان مشخصی مثلاً سه دقیقه سرویس می‌گیرد.اگر فرض کنیم فاصله بین ورود مشتری‌ها ۵ دقیقه باشد بنابراین سرویس دهنده بیکار می‌باشد.

مدت زمان بین دو ورود (در این مثال ۵ دقیقه) را a و مدت انجام سرویس را s می‌نامیم. بدیهی است تا زمانی که a>s باشد، صفی تشکیل نمی‌شود و سرویس دهنده هر بار به مدت a-s بیکار می‌ماند. بنابراین نسبت استفاده از سرویس دهنده $\frac{s}{a}$ و نسبت زمان بیکاری آن $\frac{a-s}{a}$ می‌باشد. در این مثال سیستم $\frac{3}{5}$ اوقات مشغول و $\frac{2}{5}$ اوقات بیکاری می‌باشد.

حالت دیگر سیستم آن است که فاصله زمانی بین دو ورود دقیقاً برابر زمان انجام سرویس باشد (به عبارت دیگر s=a). در این مورد نیز صفی تشکیل نمی‌شود و سرویس دهنده دائم مشغول می‌باشد. این وضعیت ایده‌آل می‌باشد ولی اگر s>a باشد، یعنی فاصله زمانی بین دو ورود بیشتر از مدت زمان سرویس دهی باشد، در این صورت صف تشکیل می‌شود و مرتباً طول آن افزایش می‌یابد.

بنابراین اگر عوامل تصادفی وجود نداشته باشند، مسائل صف به آسانی قابل حل می‌باشد.ولی در دنیای واقعی هر پیشامدی دارای عوامل تصادفی می‌باشد. مثلاً ورود یک مشتری به سیستم ممکن است یک میانگین ۵ دقیقه‌ای داشته باشد، ولی یک مشتری امکان دارد بفاصله ۳ دقیقه از مشتری قبلی یا به فاصله ۷ دقیقه وارد شود. همچنین زمان انجام سرویس ممکن است یک میانگین ۳ دقیقه‌ای داشته باشد ولی زمان انجام سرویس برای یک مشتری ۲ دقیقه و برای یک مشتری دیگری ۵ دقیقه طول بکشد. بنابراین فاصله زمانی ورود مشتری به سیستم یک متغیر تصادفی است و مدت انجام سرویس نیز یک متغیر تصادفی می‌باشد. در نتیجه زمان انتظار مشتری و زمان بیکاری سرویس دهنده تصادفی خواهد

شد. در این صورت اگر زمان بین ورود دو مشتری به سیستم کمتر از زمان لازم برای اتمام سرویس‌دهی به مشتری اول باشد، مشتری دوم باید انتظار بکشد و در نتیجه صف تشکیل خواهد شد. بنابراین حتی در موقعی که میانگین مدت انجام سرویس کمتر از میانگین بین ورود دو مشتری می‌باشد، به علت وجود عوامل تصادفی صف تشکیل می‌شود. پس هدف بررسی در سیستم‌های صف آن است که هنگامی که ورود و مدت سرویس دهی تصادفی باشد، مدت انتظار مشتری، نرخ سرویس دهی و... مورد تحلیل و بررسی قرار گیرد.

٤-٤ نمادگذاری سیستم صف

برای آنکه نوع صف و توزیعهای احتمالی زمان ورود و مدت انجام سرویس با یک علامت ریاضی مشخص گردد، ترتیب زیر را در نظر می‌گیریم که در کنفرانس استاندارد نمودن علامتگذاری در تئوری صف در سال ۱۹۷۱ پیشنهاد شده است.

(نوع ترتیب صف/ جمعیت مشتری/ ظرفیت سیستم/ تعداد سرویس دهنده/ فرایند انجام سرویس/ فرایند ورود) سه مورد سمت راست ترکیب فوق در مواردی که صف با ظرفیت نامحدود و ترتیب (دیسیپلین) FCFS مشخص شده حذف می‌شوند و سیستم صف با سه پارامتر اول بیان می‌گردد. معمولاً جهت معرفی توزیع بین دو ورود (فرایند ورود) و توزیع فرایند انجام سرویس علائم زیر بکار می‌رود:

M: توزیع نمائی

E_k: توزیع ارلانگ k مرحله‌ای K-Erlang

D: توزیع ثابت (معین)

GI: توزیع احتمال عمومی برای فاصله زمانی بین دو ورود پیاپی

G: توزیع احتمال عمومی برای مدت انجام سرویس

طبق تعریف فوق در سیستم صفی، فاصله بین دو ورود پیاپی آن دارای توزیع نمائی، مدت زمان انجام سرویس توزیع ارلانگ ۲-Erlang و دارای سه سرویس دهنده باشد، سیستم را با علامت $M/E_۲/۳$ نشان می‌دهیم.

با توجه به بررسی متغیرهای اساسی در سیستم صف که در بخش قبلی بیان شد، مشخصه‌های زیر در سیستم‌های صف مطرح می‌باشند.

$N(t)$: متغیر تصادفی که تعداد مشتری در سیستم را در زمان t، بیان می‌کند.

$P_n(t)$: احتمال وجود n مشتری در سیستم در زمان t (به شرط آنکه تعداد مشتری در سیستم در زمان صفر مشخص باشد).

$\frac{۱}{\lambda}$: میانگین فاصله ورود پیاپی دو مشتری

$\frac{۱}{\mu}$: میانگین مدت انجام سرویس برای یک مشتری

لازم به ذکر است در حالتی که c تعداد سرویس دهنده باشد، شدت ترافیک به صورت زیر بیان می‌شود:

$$\rho = \frac{\lambda}{C\mu}$$

$N_q(t)$: متغیر تصادفی که تعداد مشتری در صف در زمان t را بیان می‌کند.

$N_s(t)$: متغیر تصادفی که تعداد مشتریانی که در زمان t سرویس دریافت نموده‌اند، را بیان می‌کند.

معمولاً در بررسی سیستم صف موقعیکه سیستم در یک وضعیت یکنواخت قرار گیرد آن را مورد تجزیه و تحلیل قرار می‌دهیم. زیرا در وضعیتهای انتقالی (گذرا) تحلیل سیستم پیچیده خواهد بود. رایج‌ترین پارامترهایی که برای وضعیت یکنواخت سیستم حاسبه می‌شوند به قرار ذیل می‌باشد.

P_n: احتمال وجود n مشتری در سیستم.

L: متوسط تعداد مشتری در سیستم (شامل مشتری داخل سرویس)، $L=E(N)$.

Lq: متوسط تعداد مشتری در صف، $Lq=E(N_q)$.

W: متوسط زمان انتظار در سیستم، $E(W)=E(q)+E(s)$. لازم به ذکر است که متغیرهای s,q,w متغیرهای تصادفی مربوط به زمان انتظار در سیستم، زمان سپری شده در صف و زمان سرویس می‌باشند به طوریکه $w=q+s$.

Wq: متوسط زمان انتظار در صف $E(q)=W-E(s)$

ثابت شده است که در وضعیت یکنواخت داریم [۵].

$$L=\lambda.W$$

$$Lq=\lambda.Wq$$

بنابراین:

$$W=Wq+\frac{1}{\mu}$$

۴-۵ معرفی سیستم‌های صف مختلف

ساده‌ترین مدل صفی که بر اساس فرآیند تولد و مرگ باشد مدل M/M/1 است. بطوریکه فواصل ورود و انجام سرویس دارای توزیع نمائی می‌باشند. (هنگامی که فواصل ورود نمائی باشد، فرآیند ورود به صورت توزیع پواسن خواهد بود).

شکل ۴-۱ فرآیند این مدل را تجسم می‌نماید.

شرط اینکه این مدل دارای وضعیت یکنواخت باشد آن است که $\rho=\frac{\lambda}{\mu}<1$ باشد. یعنی میانگین نرخ ورود از میانگین نرخ خروج کمتر باشد.

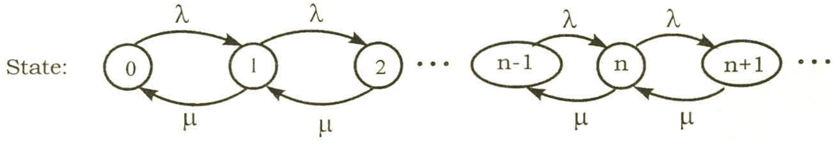
با استفاده از روابط موجود در فرآیند تولد و مرگ (که در فصل سوم بیان کردیم) و همچنین معادلات بررسی شده بخش ۴-۳، مشخصه‌های W,L_q,L,P_n و Wq قابل محاسبه می‌باشند.

عامل R_n که در فصل قبل مورد بررسی قرار گرفت به صورت زیر خلاصه می‌شود:

۱) $$R_n=\left(\frac{\lambda}{\mu}\right)^n=\rho^n \quad , \quad n=1,2,\dots\dots\dots$$

پس خواهیم داشت:

۲) $$P_n=\rho^n P_0$$



شکل ۴-۱ دیاگرام انتقال وضعیت مدل M/M/1

همچنین برای p رابطه زیر صادق است:

$$P_{\circ}=\frac{۱}{۱+\sum_{n=۱}^{\infty}\rho^{n}} \qquad (۳)$$

با استفاده از رابطه تصاعد هندسی داریم:

$$p_{\circ}=\left(\sum_{n=\circ}^{\infty}\rho^{n}\right)^{-۱}=\left(\frac{۱}{۱-\rho}\right)^{-۱}=۱-\rho$$

با استفاده از روابط (۲) و (۳) نتیجه گیری می‌شود:

$$p_n=(۱-\rho)\rho^{n} \qquad (۴)$$

متوسط تعداد مشتریان در سیستم عبارتست از:

$$L=E(n)=\sum_{n=\circ}^{\infty} np_n$$

با استفاده از رابطه (۴) داریم:

$$L=\sum_{n=\circ}^{\infty} n\,(۱-\rho)\,\rho^{n}=(۱-\rho)\,\rho\sum_{n=\circ}^{\infty}\frac{d}{d\rho}(\rho^{n})$$

$$=(۱\ \ \rho)\rho\frac{d}{d\rho}\left[\sum_{n=\circ}^{\infty}\rho^{n}\right]$$

و

$$=(۱-\rho)\,\rho\frac{d}{d\rho}\left[\frac{۱}{۱-\rho}\right]=(۱-\rho)\,\rho\left[\frac{۱}{(۱-\rho)^{۲}}\right]$$

$$=\frac{\rho}{۱-\rho}=\frac{\lambda}{\mu-\lambda}$$

همچنین می‌توان L_q را با استفاده از رابطه زیر محاسبه نمود:

$$Lq=\sum_{n=۱}^{\infty}(n-۱)P_n$$

$$=L-(۱-P_{\circ})=\frac{\rho^{۲}}{۱-\rho}$$

$$=\frac{\lambda^{۲}}{\mu(\mu-\lambda)}$$ (لازم به ذکر است $L-Lq=\rho$)

همچنین میانگین مدت انتظار در سیستم وصف برای یک مشتری عبارتست از:

$$W=\frac{L}{\lambda}=\frac{۱}{\mu-\lambda}$$

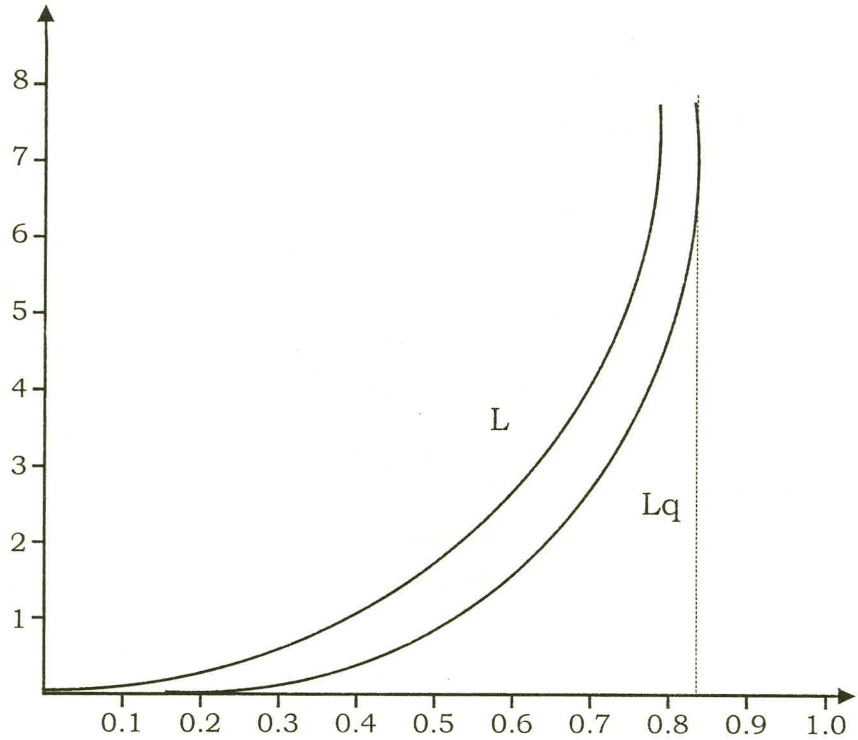
و

$$W_q=\frac{Lq}{\lambda}=\frac{\lambda}{\mu(\mu-\lambda)}$$

رابطه فوق همچنین به صورت ذیل نیر محاسبه می‌شوند:

$$W=\frac{L}{\lambda}=\frac{E(s)}{۱-\rho}\quad,\quad Wq=W-E(s)=\frac{\rho E(s)}{۱-\rho}$$

در شکل ۴-۲، میانگین تعداد مشتری در سیستم (L) و میانگین تعداد مشتری در صف (Lq) بر حسب



شکل ۴-۲ میانگین تعداد مشتری در سیستم و صف بر حسب ρ

ρ(شدت ترافیک) نشان داده شده است. همانطور که ملاحظه می‌شود اگر میزان استفاده یا شدت ترافیک‌افزایش یابد میانگین تعداد مشتری در صف به سرعت افزایش می‌یابد. از طرف دیگر اگر بخواهیم میانگین تعداد مشتری در صف ۵ نفر باشد، عملاً شدت ترافیک باید ۰/۸۵ باشد. یعنی باید ۱۵٪ اوقات، سرویس دهنده بیکار باشد تا میانگین تعداد مشتری در صف از ۵ نفر تجاوز نکند[۶].

حال اگر بخواهیم مدت زمان انتظار در سیستم و صف را در زمان t مورد بررسی قرار دهیم محاسبات زیر را در نظر می‌گیریم [۱]:

$$W(t)=P[w\le t]=\sum_{n=0}^{\infty} P[w\le t \mid N=n]\,P\,[N=n]$$

$$=\sum_{n=0}^{\infty}\int_0^t \frac{\mu^{n+1}x^n e^{-\mu x}}{n!}\left(\frac{\lambda}{\mu}\right)^n\left(1-\frac{\lambda}{\mu}\right)dx$$

$$=\int_0^t \mu e^{-\mu x}\left(1-\frac{\lambda}{\mu}\right)\sum_{n=0}^{\infty}\frac{x^n\lambda^n}{n!}\,dx$$

$$=\int_0^t \mu e^{-\mu x}\left(1-\frac{\lambda}{\mu}\right)e^{\lambda x}\,dx=\int_0^t \mu\left(1-\frac{\lambda}{\mu}\right)e^{-\mu\left(1-\frac{\lambda}{\mu}\right)x}\,dx$$

$$=1-e^{-\mu(1-\rho)t}=1-e^{-t/w}$$

لازم به توضیح است که تابع چگالی $f(x) = \frac{\mu^{n+1}X^n e^{-\mu x}}{n!}$ برای $X \geq 0$، دارای توزیع گاما می‌باشد.
همچنین $W_q(t)$ برابر است با:

$$W_q(t)=P(q\leq t)=P(q=0)+P(0<q\leq t)=1-\rho+\rho\left(1-e^{-\mu(1-\rho)t}\right)$$

$$=1-\rho e^{-\mu(1-\rho)t}=1-\rho e^{-t/w}$$

مثال ۴-۱

در سیستم کامپیوتری زمان پردازش برنامه‌ها دارای توزیع نمائی با میانگین سه دقیقه می‌باشد. برنامه‌ها به صورت تصادفی با میانگین ۴ دقیقه وارد سیستم می‌شوند و دیسیپلین (نظم) صف به صورت FCFS میباشد. مدیر مرکز کامپیوتر با مسائل زیر مواجه می‌باشد:

(a): احتمال اینکه برنامه برای پردازش بیش از ۲۰ دقیقه لازم داشته باشد چقدر است؟

(b): متوسط تعداد برنامه‌های در صف انتظار چقدر می‌باشد؟

(c): فرض کنید متوسط زمان انتظار در سیستم به ۳۰ دقیقه افزایش یابد. در این صورت:

۱- نرخ ورود به سیستم

۲- متوسط تعداد برنامه‌ها در سیستم

۳- درصد افزایش نرخ ورود، را بیابید.

(d): با فرض اینکه زمان انتظار ۹۰٪ برنامه‌هاحداکثر ۴۰ دقیقه باشد، در این صورت نرخ ورود و متوسط تعداد برنامه‌ها در صف انتظار را محاسبه نمائید.

حل (a): متغیر تصادفی بین دو ورود را t و متغیر تصادفی زمان سرویس را s تعریف می‌کنیم.

داریم:

$$E(t) = ۴ \text{ دقیقه}$$

$$\lambda = \frac{1}{E(t)} = ۰/۲۵ \frac{\text{برنامه}}{\text{دقیقه}} \text{(ورود)}$$

$$\mu = \frac{1}{E(S)} = \frac{1}{3} \frac{\text{برنامه}}{\text{دقیقه}} \text{(پردازش)}$$

$$\rho = \frac{\lambda}{\mu} = ۰/۷۵$$

متوسط زمان انتظار در سیستم برابر است با:

$$W = \frac{E(s)}{1-\rho} = ۱۲ \text{ دقیقه}$$

بنابراین طبق قانون احتمال کل داریم:

$$W(t)=P(w\leq t)=1-e^{-t/w}=1-e^{-t/12}$$

یا

$$P(w>t)=e^{-t/12}$$

بنابراین احتمال اینکه زمان انتظار (W) بیش از ۲۰ دقیقه شود عبارتست از:

$$e^{\frac{-20}{12}}=e^{\frac{-5}{3}}=0.1889$$

حل (b): میانگین طول صف به صورت زیر محاسبه می‌شود.

$$Lq=E(N_q) = \frac{\rho^{۲}}{۰/۲۵} = \frac{(۰/۷۵)^{۲}}{۰/۲۵} = ۲/۲۵ \text{ برنامه}$$

حل (c): هنگامی که زمان انتظار به ۳۰ دقیقه می‌رسد (W=۳۰)، فرض می‌کنیم متوسط زمان پردازش همان ۳ دقیقه باشد. در این صورت:

$$W=\frac{E(s)}{۱-\rho}$$

$$=\frac{E(s)}{۱-\frac{\lambda}{\mu}}=\frac{E(s)}{۱-\lambda E(s)}$$

بنابراین داریم:

$$۳۰= \frac{۳}{۱-۳\lambda}$$

یا

$$\lambda = \frac{۲۷}{۹۰} = ۰/۳ \frac{\text{برنامه}}{\text{دقیقه}} \quad \text{یا} \quad \lambda = ۱۸ \frac{\text{برنامه}}{\text{ساعت}}$$

متوسط تعداد برنامه‌ها در سیستم برابر است با:

$$L = \frac{\rho}{۱-\rho} = \frac{۰/۹}{۱-۰/۹} = ۹ \quad \text{برنامه}$$

درصد افزایش نرخ ورود عبارتست از:

$$\lambda_۱=۰/۲۵ \frac{\text{برنامه}}{\text{دقیقه}} \text{ یا } \lambda_۱ = ۱۵ \frac{\text{برنامه}}{\text{ساعت}}$$

$$\lambda_۲ = ۱۸ \frac{\text{برنامه}}{\text{ساعت}}$$

بنابراین:

$$\frac{۱۰۰(۱۸ - ۱۵)}{۱۵} = \%۲۰$$

حل (d): با توجه به اینکه ۹۰٪ تمام برنامه‌ها کمتر از ۴۰ دقیقه در سیستم انتظار می‌کشند، طبق آنچه در خواص توزیع نمائی (بخش ۵-۲) بیان شد می‌توان نوشت:

$$\pi_w(r) = E(w)Ln(\frac{۱۰۰}{۱۰۰ - r})$$

$$\pi_w(۹۰) = ۲/۳ \text{ w}$$

بنابر این:

$$۴۰=\pi_w(۹۰) = ۲/۳ \text{ w} = \frac{۲/۳ E(s)}{۱-\lambda E(s)} = \frac{۲/۳}{۱-۳\lambda}$$

که نتیجه می‌دهد:

$$\lambda = ۱۶/۵۵ \frac{\text{برنامه}}{\text{ساعت}} \quad \text{و یا} \quad \lambda = \frac{۳۳/۱}{۱۲۰} \frac{\text{برنامه}}{\text{دقیقه}}$$

افزایش نرخ ورودی برابر است با:

$$۱۰۰\times\frac{(۱۶/۵۵-۱۵)}{۱۵} = \%۱۰/۳$$

همچنین متوسط تعداد برنامه در صف جهت پردازش برابر است با:

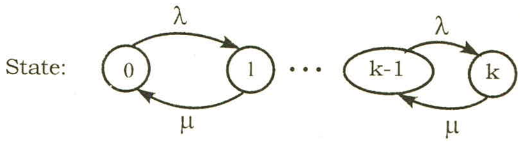
$$L_q=E(N_q)=\frac{\rho^2}{1-\rho}=3/97 \text{ برنامه}$$

لازم به ذکر است که:

$$\rho=\lambda E(s)=0/8275$$

۴-۶ مدل صف با ظرفیت محدود M/M/1/K

در مواردیکه ظرفیت سیستم محدود باشد با تغییرات جزئی در مدل M/M/۱ می‌توان پارامترهای اساسی سیستم را بدست آورد. در مدل مورد نظر K، نشاندهنده حداکثر تعداد مشتری، برنامه و...در سیستم می‌باشد. در نتیجه وقتی مشتری وارد شود و K مشتری قبلاً در سیستم باشند، این نوع مشتریان ورودشان مسدود شده و به سیستم راه داده نمی‌شوند (blocked) یا آنکه خودشان تصمیم می‌گیرندبه جای دیگر مراجعه نمایند (balking). در شکل ۴-۳ انتقال سیستم از یک وضعیت به وضعیت دیگر با استفاده از فرآیند تولد و مرگ نشان داده شده است.



شکل ۴-۳ وضعیت انتقال سیستم M/M/۱/K

در این مدل داریم:

$$\lambda_n=\begin{cases}\lambda & n=0,1,\dots,k-1\\ 0 & n\geq k\end{cases}$$

و

$$\mu_n=\begin{cases}\mu & n=1,2,\dots,k\\ k & n<0\end{cases}$$

همینطور برای R_n داریم:

$$R_n=\begin{cases}(\frac{\lambda}{\mu})^n=\rho^n & \\ n=1,2,\dots,N & n>N\end{cases}$$

برای P_0 رابطه حاصله از فرآیند تولد و مرگ به صورت زیر می‌باشد:

$$P_0=\frac{1}{\sum_{n=0}^{k}\rho^n}$$

$$=\frac{1}{\left(\frac{1-\rho^{k+1}}{1-\rho}\right)}=\frac{1-\rho}{1-\rho^{k+1}} \qquad (\text{اگر } \lambda\neq\mu)$$

با توجه به اینکه حداکثر K مشتری در سیستم می‌باشند، سیستم به یک حالت یکنواخت خواهد رسید. بنابراین نیازی به شرط $\mu>\lambda$ برای حالت یکنواخت سیستم نمی‌باشد. اگر $\lambda=\mu$ پس $\rho=1$ و

داریم:

$$P_0=\frac{1}{k+1}=P_n \qquad , \qquad n=1,2,\dots,k$$

بنابراین احتمالات در وضعیت یکنواخت عبارتست از:

$$P_n=\begin{cases}\dfrac{[1-\rho]\,\rho^n}{1-\rho^{k+1}}, & \lambda\neq\mu \;,\; n=0,1,\dots,k\\[2ex] \dfrac{1}{k+1}, & \lambda=\mu \;,\; n=0,1,\dots,k\end{cases}$$

در اینجا اگر $\lambda<\mu$ و $k\to\infty$ در این صورت داریم:

$$P_n=(1-\rho)\rho^n$$

متوسط تعداد مشتری در سیستم به صورت زیر (در دو حالت)محاسبه می‌شود:

$\lambda\neq\mu$ -(۱)

$$L=E(N)=\sum_{n=0}^{k} n\rho_n=\sum_{n=1}^{k}\left(\frac{1-\rho}{1-\rho^{k+1}}\right)(n\rho^n)$$

$$=\left(\frac{1-\rho}{1-\rho^{k+1}}\right)\rho\sum_{n=1}^{k} n\rho^{n-1}$$

$$=\left(\frac{1-\rho}{1-\rho^{k+1}}\right)\rho\sum_{n=1}^{k}\frac{d\rho^n}{d\rho}=\left(\frac{1-\rho}{1-\rho^{k+1}}\right)\rho\frac{d}{d\rho}\sum_{n=0}^{k}\rho^n$$

$$=\left(\frac{1-\rho}{1-\rho^{k+1}}\right)\rho\frac{d}{d\rho}\left(\frac{1-\rho^{k+1}}{1-\rho}\right)$$

$$=\left(\frac{1-\rho}{1-\rho^{k+1}}\right)\rho\left[\frac{-(k+1)\rho^k}{1-\rho}+\left(1-\rho^{k+1}\right)(1-\rho)^{-2}\right]$$

$$=\frac{\rho}{1-\rho}-\frac{(k+1)\rho^{k+1}}{1-\rho^{k+1}}$$

$\lambda=\mu$ -(۲)

$$L=\sum_{n=0}^{k} n\rho_n=\frac{1}{k+1}(1+2+\dots+k)=\frac{k(k+1)}{2(k+1)}=\frac{k}{2}$$

روابط فوق را به صورت زیر خلاصه می‌کنیم:

$$L=\begin{cases}\dfrac{\rho}{1-\rho}-\dfrac{[k+1]\rho^{k+1}}{1-\rho^{k+1}} & ، \quad \lambda\neq\mu\\[2ex] \dfrac{k}{2} & ، \quad \lambda=\mu\end{cases}$$

برای Lq، میانگین تعداد مشتری در صف، می‌توان نوشت.

$$Lq=L-(1-P_0)$$

با توجه به اینکه ظرفیت سیستم K می‌باشد و تعدادی از مشتریان که با نرخ λ وارد سیستم می‌شوند، موفق به

ورورد به صف (که دارای نرخ ورودی $\lambda\alpha$است) نمی‌شوند، خواهیم داشت:

$$\lambda\alpha=\lambda(1-P_k)$$

با استفاده از فرمول Little Law داریم[۵]:

$$W=E(w)=\frac{L}{\lambda\alpha}$$

و

$$Wq=Eq(w)=\frac{Lq}{\lambda\alpha}$$

همچنین احتمال اینکه سرویس دهنده مشغول باشد برابر است با:

$$\rho=\lambda\alpha E(s)=\lambda(1-P_k)E(s)$$

مثال ۴-۲

در یک مرکز سوئیچینگ پیغام، پیغامها به طور تصادفی با نرخ میانگین ۲۴۰ پیغام در دقیقه، به مرکز می‌رسند. سرعت انتقال داده خط ارتباطی ۸۰۰کاراکتر در ثانیه می‌باشد. توزیع طول پیغام به صورت نمائی با میانگین طول ۱۷۶ کاراکتر برای هر پیغام، می‌باشد. محاسبه کنید چه تعداد بافر لازم است تا هنگامی که پیغامی وارد می‌شود به احتمال کمتر از($P<0/0005$) همه بافرها پر شده باشند. برای تعداد بافر محاسبه شده، مقادیر W,Lq,L و Wq را حساب کنید.

همچنین‌احتمال آنکه یک پیغام ورودی در سیستم حداکثر ۲/۵ ثانیه سپری کند، چقدر است؟

احتمال آنکه یک پیغام قبل از ارسال (در صف) حداکثر ۲/۵ ثانیه سپری کند، چقدراست؟

حل: برای این سیستم مدل M/M/1/K مناسب می‌باشد. چون ظرفیت سیستم برابر Kپیغام‌می‌باشد. بنابراین هنگامی که یک پیغام در حال ارسال می‌باشد K-1 بافر برای ذخیره پیغامهای سیستم نیاز است. پس:

$$\lambda=240\ \frac{\text{پیغام}}{\text{دقیقه}} \quad \text{و} \quad \lambda=4\ \frac{\text{پیغام}}{\text{ثانیه}}$$

متوسط زمان ارسال یک پیغام برابر است با:

$$\frac{176\ \text{کاراکتر}}{800\ \text{کاراکتر}}=0/22\ \text{ثانیه}$$

بنابراین شدت ترافیک (بکارگیری سرویس دهنده) عبارتست از:

$$\rho=\lambda E(s)=4\times0/22=0/88$$

احتمال آنکه همه بافرها پرشده باشند (k پیغام در سیستم باشد)برابر است با:

$$P_k=\frac{(1-\rho)\ \rho^k}{1-\rho^{k+1}}$$

برای حل معادله فوق روش زیر را دنبال می‌نمائیم.

فرض کنید K=۲۰ ظرفیت سیستم باشد در این صورت:

$$P_{25}=\frac{(1-0/88)(0/88)^{20}}{1-(0/88)^{21}}=0/00998936$$

این مقدار تقریباً برابر با ۱٪ می‌باشد. حال اگر ظرفیت سیستم را برابر ۲۵ پیغام در نظر بگیریم خواهیم داشت:

$$P_{25}=\frac{(1-0/88)(0/88)^{25}}{1-(0/88)^{26}}=0/005095$$

این مقدار تعداد بافر مورد نظر را نشان نمی‌دهد ($P\geq0/005$) بنابراین:

$$P_{26}=0/004464<0/005$$

پس ظرفیت سیستم ۲۶ می‌باشد که به ۲۵ بافر نیاز دارد.
با توجه به فرمولهای این بخش مقادیر زیر محاسبه می‌گردد.

۱- متوسط تعداد پیغام در سیستم:	پیغام	$L=E(N)=۶/۴۴۹$
۲- متوسط تعداد پیغام در صف	پیغام	$Lq=Eq(N)=۵/۵۷۳$
۳- میانگین زمان سیری شده در سیستم	ثانیه	$W=E(w)=۱/۶۲$
۴- میانگین زمان سیری شده در صف	ثانیه	$Wq=E_q(w)=۱/۴$

جهت محاسبه $W(t)$ و $Wq(t)$ در مدل M/M/1/K داریم[۱]:

$$W(t)=P(W\leq t)=1-\sum_{n=0}^{k-1} q_n P[\mu t,n]$$

$$P[\mu t,n]=\sum_{k=0}^{n} e^{-\mu t}\frac{(\mu t)^k}{k!}$$

$$q_n=\frac{p_n}{1-p_k} \qquad n=0,1,\ldots k-1$$

همچنین برای $Wq(t)$ خواهیم داشت:

$$Wq(t)=P(q\leq t)=1-\sum_{n=0}^{k-2} q_{n+1}P(\mu t,n)$$

برای این مثال، مقادیر $W(t)$, $Wq(t)$ عبارتند از:

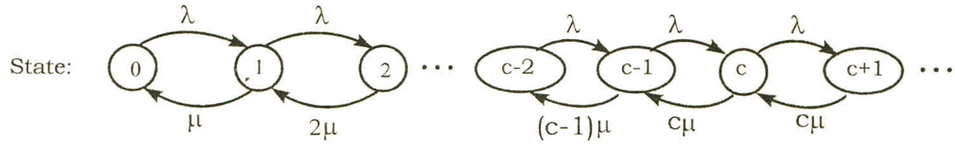
$$Wq(t\leq ۲/۵)=P(q\leq ۲/۵)=۰/۸۰۳۹ \quad \text{و} \quad W(t\leq ۲/۵)=۰/۷۷۲۰۸$$

۴-۷ مدل صف با سرویس دهنده های متعدد M/M/C

یکی از مواردی که در سیستم‌های واقعی معمول است، وجود سرویس دهنده‌های متعدد (Multiple servers) در سیستم می باشد. فرض می‌کنیم توزیعهای احتمالی فواصل بین ورود و فواصل انجام سرویس (زمان سرویس)، دارای توزیع نمایی و تعداد سرویس دهنده‌ها C می‌باشد در این صورت مدل صف به صورت M/M/C است.

اگر در چنین مدلی ظرفیت سیستم محدود باشد (N مشتری) مدل M/M/C/N را خواهیم داشت. حالت خاص این مدل وقتی است که امکان صف کشیدن نباشد، و ظرفیت سیستم برابر تعداد سرویس دهنده‌ها باشد (M/M/C/C). به طور کلی در تمام این مدلها تعداد مشتریان بالقوه بی‌نهایت و دیسیپلین صف بر اساس FCFS می‌باشد. فرض کنید که تمام سرویس دهنده‌ها دارای نرخ میانگین، مدت زمان سرویس، برابر μ باشند، در این صورت لازم نیست بدانیم در هر لحظه زمانی، کدام سرویس دهنده مشغول است. بلکه تعداد سرویس دهنده‌های مشغول حائز اهمیت می‌باشد. همچنین با وجود آنکه ممکن است تعداد C مشتری در یک زمان در حال سرویس باشند، در هر لحظه فقط یک مشتری سیستم را ترک می‌کند. زیرا احتمال آنکه دو واقعه در یک لحظه زمانی به وقوع پیوندد صفر است.

بنابراین مدل را به صورت فرآیند تولد و مرگ می‌توان فرموله نمود. با این تفاوت که انتقال سیستم از یک وضعیت به وضعیت دیگر فقط به وضعیت سیستم در آن لحظه بستگی دارد. یعنی اگر فقط یک مشتری در



شکل ۴-۴ دیاگرام وضعیت انتقال مدل M/M/C

حال سرویس باشد و بقیه سرویس دهندگان بیکار باشند، انتقال سیستم از وضعیت ۱ به ۰ با نرخ μ خواهد بود. ولی اگر دو مشتری در حال سرویس باشند، انتقال سیستم از وضعیت ۲ به ۱ وقتی اتفاق می‌افتد که یکی از دو مشتری سیستم را ترک کند. در نتیجه نرخ انتقال سیستم از وضعیت ۲ به ۱ دو برابر می‌شود یعنی 2μ.

همینطور اگر سه مشتری در سیستم در حال سرویس باشند، نرخ انتقال از وضعیت سه به دو، سه برابر حالت اول یعنی 3μ می‌باشد. طبیعی است که این استدلال تا C سرویس دهنده تکرار می‌شود. بنابراین نرخ انتقال وضعیت سیستم $C\mu$ باقی می‌ماند. شکل ذیل وضعیت انتقال سیستم در این مدل صف رانشان می‌دهد.

بنابر استدلال فوق و فرمولهای فرآیند تولد و مرگ، برای بدست آوردن مشخصه‌های مدل M/M/C به ترتیب زیر عمل می‌کنیم:

$$\lambda_n = \lambda \qquad , \qquad n = 0, 1, 2, \ldots$$

$$\mu_n = \begin{cases} n\mu & , \quad n = 1, 2, \ldots, C \\ C\mu & , \quad n \geq C \end{cases}$$

و

$$R_n = \frac{\lambda_{n-1} \; \lambda_{n-2} \, , \cdots \; \lambda_0}{\mu_n \; \mu_{n-1} \cdots\cdots \; \mu_1}$$

که در این مدل تبدیل می‌شود به:

$$R_n = \begin{cases} \dfrac{\lambda.\lambda.\ldots.\lambda}{n\mu.\,(n-1)\,\mu.\ldots.\mu} = \dfrac{\left(\frac{\lambda}{\mu}\right)^n}{n!} & , \quad n \leq C \\[2ex] \dfrac{\left(\frac{\lambda}{\mu}\right)^n}{C!} \left(\dfrac{\lambda}{C\mu}\right)^{n-C} = \dfrac{\left(\frac{\lambda}{\mu}\right)^n}{C!C^{n-C}} & , \quad n \geq C \end{cases}$$

می‌دانیم که وقتی ظرفیت سیستم نامحدود است، باید $\lambda < C\mu$ شود تا سیستم دارای وضعیت یکنواخت باشد. با توجه به این فرض P_0 را طبق فرآیند تولد و مرگ به ترتیب زیر محاسبه می نماییم:

$$P_0 = \frac{1}{1 + \sum_{n=1}^{\infty} R_n}$$

$$= \frac{1}{\left[\sum_{n=0}^{C-1} \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} + \frac{\left(\frac{\lambda}{\mu}\right)^C}{C!} . \sum_{n=C}^{\infty} \left(\frac{\lambda}{C\mu}\right)^{n-C} \right]}$$

$$=\frac{1}{\left[\sum_{n=0}^{c-1}\frac{\left(\frac{\lambda}{\mu}\right)^n}{n!}+\frac{\left(\frac{\lambda}{\mu}\right)^C}{C!}\cdot\frac{1}{1-\frac{\lambda}{C\mu}}\right]}$$

و برای P_n خواهیم داشت:

$$P_n=R_n\,.\,P_0=\begin{cases}\dfrac{\left(\frac{\lambda}{\mu}\right)^n}{n!}P_0 & 0\leq n\leq c\\[2ex] \dfrac{\left(\frac{\lambda}{\mu}\right)^n}{c!c^{n-c}}P_0 & n>c\end{cases}$$

همچنین با قرار دادن $\rho=\frac{\lambda}{C\mu}$ و استفاده از رابطه Lq در فرآیند تولد و مرگ داریم:

$$Lq=E\left(N_q\right)=\sum_{n=c}^{\infty}\left(n-c\right)P_n=\sum_{k=0}^{\infty}kP_{c+k}=\sum_{k=0}^{\infty}k\frac{\left(\frac{\lambda}{\mu}\right)^c}{c!}\rho^k p_0$$

$$=P_0\frac{\left(\frac{\lambda}{\mu}\right)^c}{c!}\cdot\sum_{k=0}^{\infty}k\rho^k=P_0\frac{\left(\frac{\lambda}{\mu}\right)^c}{c!}\{0+\rho+2\rho^2+3\rho^3+\ldots\}$$

$$=P_0\frac{\left(\frac{\lambda}{\mu}\right)^c}{c!}\rho\frac{d}{d\rho}\{1+\rho+\rho^2+\ldots\ldots\}$$

$$=P_0\frac{\left(\frac{\lambda}{\mu}\right)^c}{c!}\rho\frac{d}{d\rho}\left(\frac{1}{1-\rho}\right)=\frac{P_0\,\rho^{c+1}}{c!\ (1-\rho)^2}$$

متوسط زمان انتظار در صف و سیستم به ترتیب زیر محاسبه می‌گردند.

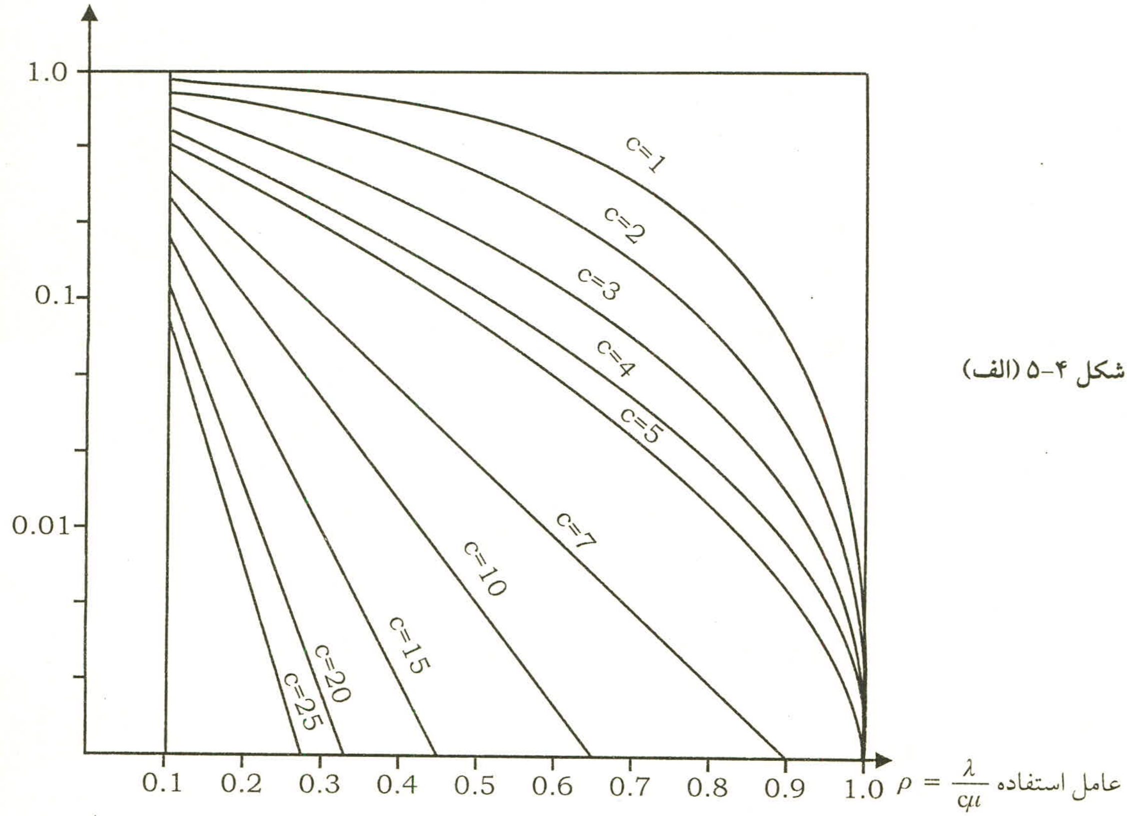
$$Wq=\frac{Lq}{\lambda}$$

$$W=Wq+E(s)=Wq+\frac{1}{\mu}$$

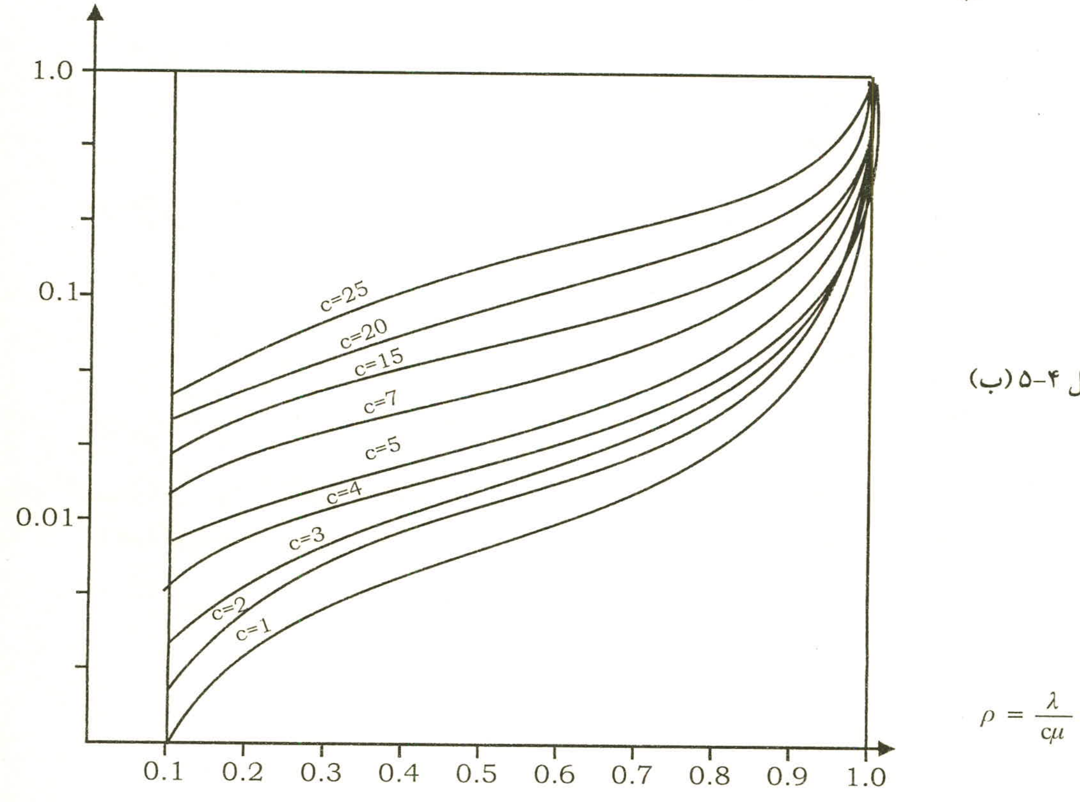
بنابراین:

$$L=\lambda W$$

نمودارهای زیر (شکل‌های ۵-۴ الف و ۵-۴ ب) تغییرات P_0 (احتمال اینکه هیچ مشتری در سیستم نباشد) و L_q (میانگین تعداد مشتری در صف) را، بر حسب عامل شدت ترافیک $\rho=\frac{\lambda}{c\mu}$ نشان می‌دهد.



شکل ۴-۵ (الف)



شکل ۴-۵ (ب)

مثال ۴-۳

در یک بندر تجاری کشتی‌ها بعد از ورود، باراندازی و بارگیری کرده و سپس از بندر خارج می‌شوند. در هر روز به طور متوسط دو کشتی وارد بندر شده و هر اسکله قادر است به طور متوسط سه کشتی را در روز سرویس (بارگیری) نماید. توزیع فواصل ورود و سرویس نمائی فرض می‌شود. پارامترهای ذیل را برای مواردیکه یک اسکله و دو اسکله در بندر باشد، محاسبه و مقایسه می‌نمائیم.

$$\lambda=2 \quad , \quad \mu=3$$

$$\rho=\begin{cases} \frac{۲}{۳} \quad , & C=1 \\ \frac{۲}{۶} \quad , & C=2 \end{cases}$$

با توجه به اینکه $\rho<۱$، سیستم در وضعیت یکنواخت قرار دارد. بنابراین روابط فوق‌الذکر را در جدول ۱-۴ داریم:

جدول ۴-۳ فرمولها و نتایج حاصله در حالتی که سه مرکز در یک مکان قرار گرفته‌اند

متغیر جواب	مورد ۱=c	مورد ۲=c
عامل استفاده ρ	$\frac{۲}{۳}$	$\frac{۱}{۳}$
احتمال نبودن کشتی در بندر یا احتمال بیکاری سرویس دهنده (P_0)	$\frac{۱}{۳}$	$\frac{۱}{۲}$
احتمال بودن یک کشتی در بندر (P_1)	$\frac{۲}{۶}$	$\frac{۱}{۲}$
احتمال بودن n کشتی در بندر ($n\geq۲$)	$\frac{۱}{۳}(\frac{۲}{۳})^n$	$(\frac{۱}{۳})^n$
میانگین تعداد کشتی در بندر L	۲	$\frac{۳}{۴}$
میانگین تعداد کشتی‌های منتظر سرویس L_q	$\frac{۴}{۳}$	$\frac{۱}{۱۲}$
میانگین تعداد روزهائی که کشتی در بندر است W	1	$\frac{۳}{۸}$
میانگین تعداد روزهائی که کشتی منتظر دریافت سرویس است	$\frac{۲}{۳}$	$\frac{۱}{۲۴}$

۴-۸ سیستم صف با ظرفیت محدود و سرویس دهندگان متعدد M/M/C/N

در این سیستم متغیرهای جواب به ترتیب زیر محاسبه می‌شوند [۶]:

$$R_n = \begin{cases} \dfrac{(\frac{\lambda}{\mu})^n}{n!} & , n \le c \\ \dfrac{(\frac{\lambda}{\mu})^c}{C!} - (\frac{\lambda}{c\mu})^{n-c} = \dfrac{(\frac{\lambda}{\mu})^n}{C!\ c^{n-c}} & , n \ge c \\ \circ & , n > N \end{cases}$$

احتمال وجود n مشتری در سیستم عبارتست از:

$$R_n = \begin{cases} \dfrac{(\frac{\lambda}{\mu})^n}{n!} P_\circ & , n \le c \\ \dfrac{(\frac{\lambda}{\mu})^n}{C!\ C^{n-c}} P_\circ & , N \ge n \le c \\ \circ & , n > N \end{cases}$$

که در آن احتمال اینکه هیچ مشتری در سیستم نباشد ($P_\circ$) برابر است با:

$$P_\circ = \frac{1}{\left[1 + \sum_{n=1}^{c} \frac{\left[\frac{\lambda}{\mu}\right]^n}{n!} + \frac{\left[\frac{\lambda}{\mu}\right]^n}{c!} \sum_{n=c+1}^{N} \left(\frac{\lambda}{c\mu}\right)^{n-c} \right]}$$

همچنین میانگین تعداد مشتری در صف از رابطه زیر بدست می‌آید.

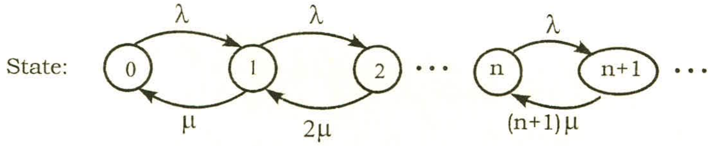
$$Lq = \frac{P_\circ \left(\frac{\lambda}{\mu}\right)^c \rho}{C!\ (1-\rho)^2} \left(1 - \rho^{N-C} - (N-C)\rho^{N-C}(1-\rho) \right)$$

با توجه به رابطهٔ فوق مقادیر W_q , W مورد ارزیابی قرار می‌گیرند.

۴-۹ سیستم صف با تعداد نامحدود سرویس دهنده $M/M/\infty$

در این سیستم بلافاصله پس از ورود هر مشتری یک سرویس دهنده برای آن فراهم می‌گردد. انتقال وضعیت سیستم در شکل ۴-۶ مشخص شده است:

همانطور که در شکل نیز دیده می‌شود.



شکل ۴-۶ دیاگرام وضعیت انتقال مدل M/M/∞

$$R_n = \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} \qquad , \qquad n = ۱, ۲, ...$$

همچنین داریم:

$$P_{۰} = \frac{۱}{\sum_{n=۰}^{\infty} \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!}} = \frac{۱}{e^{\frac{\lambda}{\mu}}}$$

احتمال وجود n مشتری در سیستم عبارتست از :

$$P_n = e^{-\left(\frac{\lambda}{\mu}\right)} \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} \qquad , \quad n = ۱, ۲, ...$$

مثال ۴-۴

در یک سیستم تلفن، پیامهای تلفنی به طور تصادفی با نرخ ۱۴۰ پیام در ساعت وارد می‌شوند. تعداد زیادی خط تلفن برای جوابگویی تهیه شده است به طوریکه یک میانگین ۳ دقیقه برای پیامهای ورودی فراهم می‌شود. میانگین تعداد خطوط مذکور چقدر است؟ ۹۰ امین و ۹۵ امین درصد تعداد خطوط مذکور را محاسبه نمائید.

حل مدل M/M/∞ برای این سیستم مناسب می‌باشد در این مورد داریم[۱]:

$$\rho = \lambda .\ E(s) = \frac{۱۴۰}{۶۰} \frac{\text{تلفن}}{\text{دقیقه}} \times ۳ \frac{\text{دقیقه}}{\text{تلفن}} = ۷$$

بنابراین میانگین تعداد خطوط مورد نیاز ۷ می‌باشد.

با استفاده از تقریب نرمال مقدار ۹۰ امین درصد عبارتست از مقدار میانگین بعلاوه ۱/۲۸ برابر انحراف معیار، به عبارت دیگر:

$$۷ + ۱/۲۸\sqrt{۷} = ۱۰/۳۸$$

که برابر ۱۱ خط منظور میگردد.

در مورد ۹۵ امین درصد تعداد خطوط مذکور، میانگین بعلاوه ۱/۶۴۵ برابر انحراف معیار، مقدار مطلوب را می‌دهد یعنی:

$$۷ + ۱/۶۴۵\sqrt{۷} = ۱۱/۳۵$$

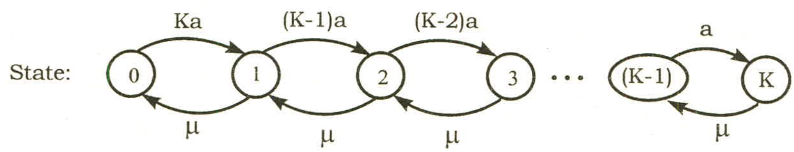
که برابر ۱۲ خط در نظر گرفته می‌شود.

۴-۱۰ سیستم صف با تعداد محدود منابع M/M/1/K/K

در این سیستم K اول ظرفیت سیستم را نشان می‌دهد و K دوم، نشان دهنده کل تعداد مشتری ممکن می‌باشد.

هرچند که ظرفیت سیستم بالقوه می‌تواند از K بیشتر باشد، ولی هنگامیکه تمام مشتریان وارد سیستم شوند تعداد آنها بیشتر از K نخواهد بود. زیرا دلیلی ندارد ظرفیتی در سیستم بوجود آوریم، که هرگز مورد استفاده قرار نگیرد. فرض می‌کنیم وقتی یک مشتری سرویس دریافت کرد، به جمعیتی برمی‌گردد که دوباره مراجعه می‌کنند. فرق این مدل، با مدلهای قبلی، در نرخ ورود می‌باشد. وقتی بیشتر مشتریان در سیستم باشند، نرخ ورود کمتر خواهد بود و در یک حالت خاص اگر تمام جمعیت در سیستم باشند، عملاً نرخ ورود صفر خواهد شد.

فرض کنید که λ نرخ ورود برای هر مشتری باشد. به عبارت دیگر از زمانی که مشتری به منبع باز می‌گردد، تا زمانی که تصمیم می‌گیرد دوباره از سرویس سیستم استفاده کند، دارای توزیع نمائی با میانگین $\frac{1}{\lambda}$ باشد. فرض می‌کنیم که λ برای همه مشتریان یکسان است (دقت شود که فاصله ورود بین دو مشتری در این مدل قابل اندازه گیری نیست، چون به تعداد افراد داخل سیستم بستگی دارد). اگر فقط یک مشتری در منبع باقی باشد، یعنی K-1 مشتری در سیستم باشد، بدیهی است که میانگین نرخ ورود، این مشتری باقیمانده، به سیستم λ است. اگر دو مشتری در منبع باشد، هریک تصمیم بگیرند جهت سرویس به سیستم مراجعه کنند، در واقع یک ورود اتفاق می‌افتد. بنابراین نرخ ورود به سیستم 2λ خواهد بود. با همین استدلال می‌توان نرخ انتقال سیستم از یک وضعیت به وضعیت دیگر را در شکل ۴-۷، با استفاده از فرآیند تولد و مرگ نشان داد:



شکل ۴-۷ وضعیت انتقال مدل M/M/1/K/K

بنابراین متغیرهای جواب بااستفاده‌ازفرآیند تولد و مرگ به صورت زیر محاسبه می‌شوند[۶]:

$$R_n = \begin{cases} \dfrac{\lambda_{n-1}\,\lambda_{n-2}\,\cdots\cdots\,\lambda_0}{\mu_n\cdot\mu_{n-1}\,\cdots\cdots\mu_1} = N(N-1)\,(N-2)\,\cdots\cdots\,(N-n+1)\left(\dfrac{\lambda}{\mu}\right)^n = \dfrac{N!}{(N-n)!}\left(\dfrac{\lambda}{\mu}\right)^n & ,\ n\le K \\ 0 & ,\ n>K \end{cases}$$

احتمال نبودن مشتری در سیستم، P_0 عبارتست از:

$$P_0 = \frac{1}{1+\sum\limits_{n=1}^{\infty} R_n} = \frac{1}{\left[1+\sum\limits_{n=0}^{N} \dfrac{N!}{(N-n)!}\left(\dfrac{\lambda}{\mu}\right)^n\right]}$$

احتمال بودن n مشتری در سیستم، P_n برابر است با:

$$P_n = \frac{N!}{(N-n)!}\left(\frac{\lambda}{\mu}\right)^n P_0$$

میانگین مشتریان در صف و در سیستم به ترتیب عبارتست از:

$$Lq = \sum_{n=1}^{\infty}(n-1)\,P_0 = \sum_{n=1}^{N}(n-1)\,P_n$$

$$= N - \frac{\lambda + \mu}{\lambda}(1 - P_0)$$

$$L = \sum_{n=0}^{\infty} nP_n = Lq + (1 - P_0) = N - \frac{\mu}{\lambda}(1 - P_0)$$

میانگین زمان انتظار در صف و در سیستم به ترتیب زیر می‌باشد.

$$W = \frac{L}{\lambda_\alpha} \quad , \quad Wq = \frac{Lq}{\lambda_\alpha}$$

که در آن داریم :

$$\lambda_\alpha = \sum_{n=0}^{\infty} \lambda_n P_n$$

$$= \sum_{n=0}^{N} (N-n)\ \lambda P_n = \lambda\ (N-L)$$

۴-۱۱ سیستم صف با تعداد محدود منابع و چندین سرویس دهنده M/M/C/K/K

این مدل را مدل تعمیر ماشین‌آلات نیز می‌گویند. در صورتی که K ماشین و C تعمیرکار وجود داشته باشند، نرخ خراب شدن ماشین α، (ورود) و تعداد ماشینی که در واحد زمان توسط یک تعمیرکار، تعمیر می‌شود μ، می‌باشد. شکل ۴-۸ ترتیب انتقال وضعیت سیستم را با توجه به فرایند تولد و مرگ نشان می‌دهد.

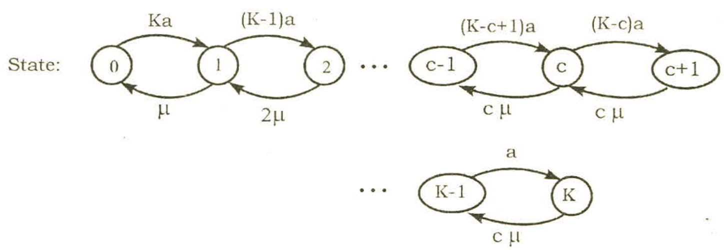
براساس فرآیند تولد و مرگ متغیرهای جواب برای این مدل به صورت زیر محاسبه می‌گردند [۶].

برای ضریب R_n داریم:

$$R_n = \begin{cases} \dfrac{N}{(N-n)!n!}\left(\dfrac{\lambda}{\mu}\right)^n & 0 \le n \le C \\ \dfrac{N}{(N-n)!C^{n-c}}\left(\dfrac{\lambda}{\mu}\right)^n & K \ge n > C \\ & n > N \end{cases}$$

احتمال نبودن مشتری در سیستم P_0 برابر است با:

$$P_0 = \frac{1}{\left[\sum_{n=0}^{c-1} \frac{N!}{(N-n)!n!}\left(\frac{\lambda}{\mu}\right)^n + \sum_{n=c}^{N} \frac{N!}{(N-C)!C^{n-c}\,C!}\left(\frac{\lambda}{\mu}\right)^n\right]}$$



شکل ۴-۸ دیاگرام وضعیت انتقال مدل M/M/C/K/K

همچنین احتمال بودن n مشتری در سیستم به صورت زیر می‌باشد:

$$P_n = R_n . P_o$$

میانگین تعداد مشتریان در صف و در سیستم به ترتیب عبارتند از:

$$Lq = \sum_{n=c}^{N} (n-c)\, P_n$$

$$L = \sum_{n=0}^{c-1} nP_n + Lq + c\left(1 - \sum_{n=0}^{c-1} P_n\right)$$

جهت محاسبه میانگین زمان انتظار در صف ودر سیستم از فرمولهای مورد C=۱ استفاده می‌کنیم یعنی :

$$W = \frac{L}{\lambda_\alpha}\ ,\ \ Wq = \frac{Lq}{\lambda_\alpha}$$

که در آن داریم:

$$\lambda_\alpha = \sum_{n=0}^{N-1} \lambda P_n = \lambda(1 - P_N)$$

مثال ۴-۵

یک شرکت کامپیوتری در هر یک از سه مرکز کامپیوتر خود ۶ عدد کامپیوتر دارد. مدت زمانی که هر کامپیوتر فعال باشد، دارای یک توزیع نمایی، با میانگین ۳۰ ساعت فاصله بین دو خرابی است. یک مهندس تعمیرکار در هر مرکز جهت تعمیر کامپیوترها حضور دارد. هریک از آنها به طور متوسط در ۳۰ دقیقه یک کامپیوتر را تعمیر می‌کنند.

زمان تعمیر کامپیوترها دارای توزیع نمایی می‌باشد. شرکت تصمیم می‌گیرد تمامی ۱۸ کامپیوتر را در یک مکان گردآوری نموده و یک گروه دو نفره از مهندسین را برای تعمیر این مجموعه درنظر بگیرد. برنامه‌ریزی تعمیرات این مرکز وقتی سیستم خراب می‌شود بصورت زیر تعریف میگردد. حداکثر احتمالیکه هیچیک از تعمیرکاران بیکار نباشند نباید بیشتر از ۰/۱ شود. همچنین هیچکدام از ماشینها نباید به طور متوسط بیشتر از ۳۵ دقیقه خراب باشند. میانگین زمان انتظار برای شروع تعمیر، نباید از ۲۷ دقیقه تجاوز نماید. آیا یک گروه ۲ نفره تعمیرکار کافی می‌باشد؟

حل: با محاسبه احتمالات n= ۰، ۱، ... ،۶ و P_n جدول ۲-۴ زیر بدست می‌آید. مقادیر در جدول ۳-۴ محاسبه شده است. به طور مثال یک مورد را در اینجا محاسبه می‌کنیم:

$$\frac{P_2}{P_0} = 2!\binom{6}{2}\left(\frac{30}{1800}\right)^2 = \frac{2\times 6!}{2!4!}\left(\frac{1}{60}\right)^2 = \frac{30}{3600} = \frac{1}{120} = 0/008333$$

در اینجا برای سه سیستم M/M//۱/۶/۶ و یک سیستم M/M/۲/۱۸/۱۸ که مناسب مثال فوق می‌باشد، نتایج را در جدول ۳-۴ نشان داده‌ایم. به عنوان مثال در سیستم مذکور با توجه به مقدار L داریم:

ساعت ۳×۰/۱۰۶۸۲×۲۴=۷/۶۹

(زمان خرابی در ۲۴ ساعت)

هنگامی که سه مرکز در کنار یکدیگر قرار می‌گیرند داریم:

ساعت ۰/۳۰۰۴۶×۲۴=۷/۲۱

(زمان خرابی در ۲۴ ساعت)

جدول ۴-۲ احتمال در حالتی که هر مرکز یک تعمیرکار دارد

P_n	$\frac{P_n}{P_0}$	n
۰/۹۰۱۷۸	۱	۰
۰/۰۹۰۱۷۸	۰/۱	۱
۰/۰۰۷۵۱۴۸	۰/۰۰۸۳۳۳۳	۲
۰/۰۰۰۵۰۰۹	۰/۰۰۰۵۵۵۵۶	۳
$۲/۵۰۴۹\times ۱۰^{-۵}$	$۲/۷۷۷۸\times ۱۰^{-۵}$	۴
$۸/۳۴۹۸\times ۱۰^{-۷}$	$۹/۲۵۹۳\times ۱۰^{-۷}$	۵
$۱/۳۹۱۶\times ۱۰^{-۸}$	$۱/۵۴۳۲\times ۱۰^{-۸}$	۶

جدول ۴-۳ فرمولها و نتایج حاصله در حالتی که سه مرکز در یک مکان قرار گرفته‌اند.

سیستم کل	سیستم موجود (در هر مرکز)	
۰/۰۰۵۴۶۵	۰/۰۰۸۵۹۵۴	Lq
۰/۳۰۰۴۶	۰/۱۰۶۸۲	L
۰/۵۵۷۸ دقیقه	۲/۶۲۵۳ دقیقه	Wq
۱۵/۳۱۱ دقیقه	۲۶/۷۲۹ دقیقه	E(q | q>0)
۳۰/۵۵۶ دقیقه	۳۲/۶۲۵ دقیقه	W
۰/۰۱۶۶۹۲	۰/۰۱۷۸۰۳	p(ماشین n خراب باشد)
۰/۰۳۶۳	۰/۰۹۸۲۲	P(انتظار برای تعمیر)

۴-۱۲ مدل صف با فواصل ورود نمائی و سرویس غیرنمائی M/G/1

مدل صفی را با سرویس دهنده واحد، فواصل ورودنمائی و فواصل سرویس دهی با توزیع عمومی G در نظر می‌گیریم (M/G/1). فرض کنید فواصل انجام سرویس مستقل از یکدیگر بوده و دارای توزیع احتمالی یکسان با میانگین $\frac{1}{\mu}$ و واریانس σ^2 باشد.

شرط اینکه این سیستم دارای وضعیت یکنواخت باشد آن است که $\rho<1$ باشد. ثابت می‌شود که [1]:

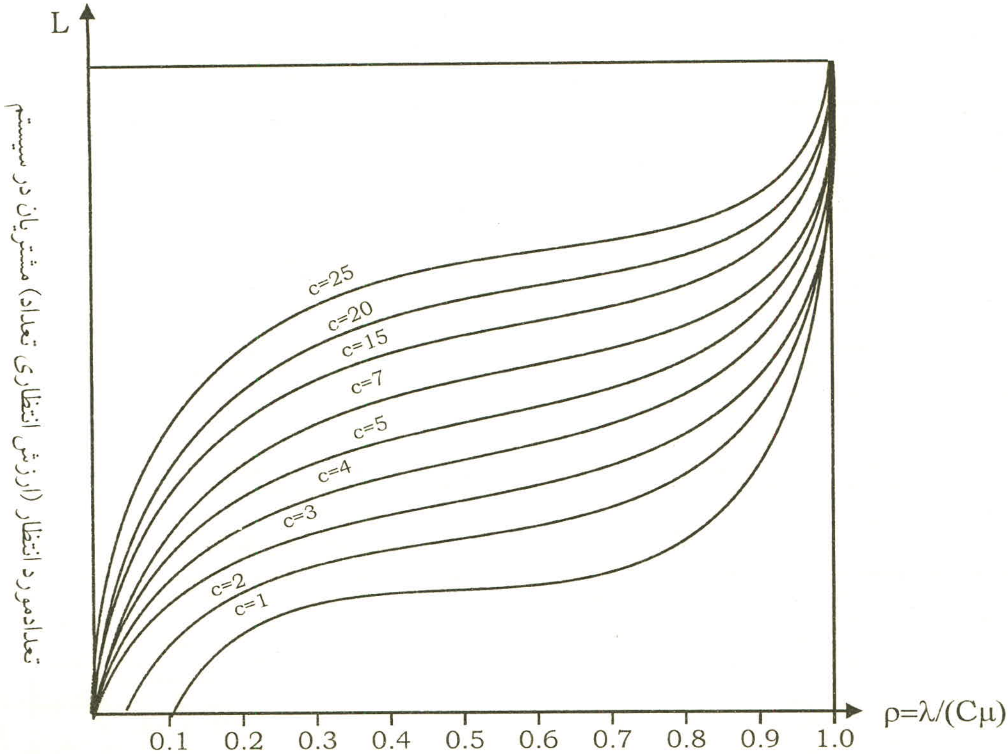
$$P_0 = 1-\rho$$

$$L = \frac{\lambda^2\sigma^2+\rho^2}{2(1-\rho)}+\rho$$

$$Wq = \frac{Lq}{\lambda}$$

$$W = Wq + \frac{1}{\mu}$$

اگر فواصل خدمت یک میزان ثابت باشد، یعنی مدل صف به صورت M/D/1، روابط فوق همچنان صادق است. با این تفاوت که در آن $\sigma^2=0$ و $Lq = \frac{\rho^2}{2(1-\rho)}$ برای مدل M/D/C روابط پیچیده‌ای ارائه می‌دهد. معمولاً از این روابط استفاده کرده و نتایج را به صورت گراف بر حسب $\rho=\frac{\lambda}{C\mu}<1$ نشان می‌دهند. نمونه‌ای از این گرافها در شکل ۴-۹ نشان داده شده است.



شکل ۴-۹ مقدار L برحسب عامل استفاده در مدل صف، فرآیند ورودی پواسن و مدت خدمت ثابت

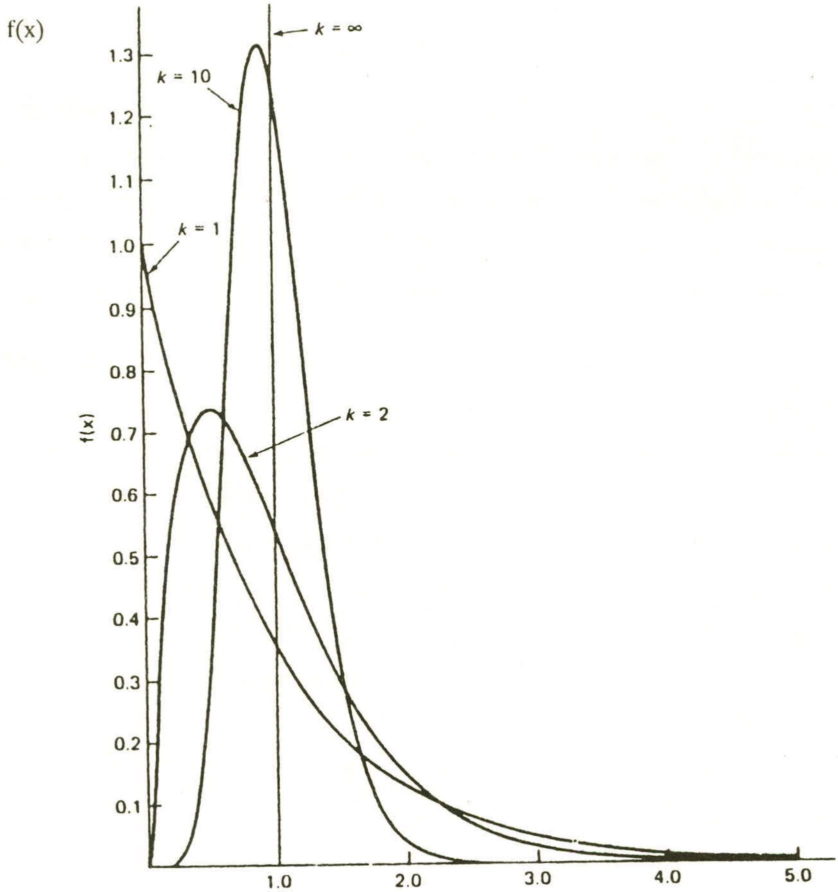
ملاحظه می‌شود که در مدل M/G/۱، اگر $\frac{1}{\mu}$ ثابت بماند با افزایش σ^2، واریانس فواصل انجام سرویس، بر میزان L, Lq, W و Wq اضافه می‌شود. در مواقعی که $\sigma^2 = \frac{1}{\mu^2}$ (واریانس توزیع‌نمائی) باشد نتایج فوق تبدیل به نتایج حاصله برابر مدل M/M/۱ می‌شود. تابحال دو مورد خاص از روابط فوق را مورد بررسی قرار دادیم.

مورد اول: وقتی توزیع فواصل انجام سرویس توزیع نمایی باشد ($\sigma^2 = \frac{1}{\mu^2}$) به عبارت دیگر مدل M/M/۱.

مورد دوم: وقتی فواصل انجام سرویس ثابت باشد ($\sigma^2 = 0$) یعنی مدل M/D/۱. یک مورد سوم آنستکه توزیع فواصل انجام سرویس توزیع ارلانگ (Erlang) باشد. به عبارت دیگر $M/E_k/1$، اگر K=۱ باشد عملاً مدل M/M/۱ را خواهیم داشت. توزیع ارلانگ در مدلهای صف به دو دلیل دارای اهمیت زیادی می‌باشد [۶].

اول آنکه اگر K متغیر تصادفی مستقل $x_1, x_2, \ldots, x_k$ در نظر بگیریم که هر یک دارای توزیع نمائی با میانگین $\frac{1}{K\mu}$ باشد، متغیر تصادفی مجموع $X = x_1 + \ldots + x_k$، دارای توزیع ارلانگ با پارامتر $K\mu$ خواهد بود. بنابراین اگر سرویس دادن به مشتری دارای K مرحله متفاوت و هر مرحله سرویس دارای توزیع نمائی باشد، فاصله سرویس مشتری دارای توزیع ارلانگ خواهد بود.

دوم آنکه توزیع ارلانگ یکسری از توزیعها (با Kهای مختلف) را دربر می‌گیرد که در آنها متغیر تصادفی مقدار غیرمنفی دارد. بنابراین بسیاری از فواصل انجام خدمت را می‌توان با تقریب به صورت توزیع ارلانگ نشان داد. حتی توزیع نمائی یک توزیع ارلانگ با K=۱ و فواصل ثابت، یک توزیع ارلانگ با K=∞ می‌باشد. سایر توزیعهای ارلانگ بین این دو حد واقع شده‌اند. شکل ۴-۱۰ توزیعها را برای مقادیر مختلف K نشان



شکل ۴-۱۰ توزیع ارلانگ K مرحله‌ای

۴-۱۳ مدل صف با فواصل ورود غیرنمائی G/M/۱

برای فواصل ورود غیرنمائی نتایج حاصله به سادگی روابط مندرج در بخش M/G/۱ نمی‌باشد. مثلاً برای G/M/۱ , G/M/C روابط غالباً پیچیده خواهد بود. همچنین نتایج برای مدل‌های E_k/E_k/۱ , D/E_k/۱ و E_k/D/۱ پیچیده می‌باشند. بنابراین یکی از راههای محاسبه کمیّت‌های مورد نظر برای مدلهای صف فوق، استفاده از شبیه‌سازی می‌باشد[۶].

۴-۱۴ تقدم (Priority) در صف

در یک سیستم صف لزومی ندارد تمامی مشتریان به صورت یکسان سرویس دریافت نمایند. بلکه تعدادی از مشتریان برای دریافت سرویس می‌توانند اولویت داشته باشند. به‌طوریکه در سیستم، مشتریان به دسته‌های مختلف، class n , ... , ۱ class تقسیم می‌شوند. فرض می‌شود که class ۱ بالاترین اولویت را و class n کمترین اولویت را دارا باشد. البته تمامی مشتریان یک دسته، دارای اولویت یکسان هستند و در داخل دسته به صورت FCFS سرویس دریافت می‌کنند. مشتری با اولویت بالاتر سرویس را زودتر دریافت

می‌کند. در تقدم مشتریان دو سیاست مطرح می‌شود.

۱- روش اول :

در این سیستم سرویس فعلی قطع می‌گردد و مشتری جدید (که اولویت بالاتر دارد) سرویس خود را آغاز می‌کند. مشتری که سرویسش قطع شده به ابتدای صف مربوط به دسته خود برمی‌گردد. البته در این روش یک اصلاحی در تقدیم (Preemptive) درنظر گرفته می‌شود. بدین معنی که مشتری سرویسش قطع شده، در دسترسی بعدی به سرویس، ادامه سرویس را از نقطه قطع سرویس به بعد دریافت می‌کند.

۲- روش دوم :

در این سیستم مشتری جدید که وارد سیستم می‌شود منتظر می‌ماند تا مشتری در حال دریافت سرویس به طور کامل سرویس دریافت کند، سپس اقدام به دریافت سرویس نماید. این روش را (Head - of - Line) (HOL) نیز می‌گویند.

فرض می‌کنیم در یک سیستم صف با تقدم، که در آن N دسته مشتری داریم، ورودی به سیستم به صورت پواسن باشد. نظم (دیسیپلین) صف را به صورت HOL درنظر می‌گیریم. در این صورت نرخ ورودی λ_i و زمان سرویس با میانگین $E(S_i) = \frac{1}{\mu_i}$ را برای class i خواهیم داشت. بنابراین کل نرخ ورودی به سیستم به صورت توزیع پواسن و با مقدار زیر خواهد بود.

$$\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$$

عملیات ریاضی جهت محاسبات معیارهای مورد نظر بصورت ذیل قرار گرفته است [۱].

$$E(s) = \frac{\lambda_1}{\lambda} E(s_1) + \frac{\lambda_2}{\lambda} E(s_2) + \dots + \frac{\lambda_n}{\lambda} E(s_n)$$

$$E(s^2) = \frac{\lambda_1}{\lambda} E(s_1)^2 + \frac{\lambda_2}{\lambda} E(s_2)^2 + \dots + \frac{\lambda_n}{\lambda} E(s_n)^2$$

$$u_j = \lambda_1 E(s_1) + \dots + \lambda_j E(s_j) \quad , j = 1, 2, \dots, n$$

$$u_n = u = \lambda E(s)$$

$$Wq_j = E(q_j) = \frac{\lambda E(s^2)}{2(1 - u_{j-1})(1 - u_j)} \quad , \ j = 1, 2, \dots n \ , \ u_0 = 0$$

$$Wq = E(q) = \frac{\lambda_1}{\lambda} E(q_1) + \frac{\lambda_2}{\lambda} E(q_2) + \dots + \frac{\lambda_n}{\lambda} E(q_n)$$

$$W_j = E(w_j) = E(q_j) + E(s_i) \ , \ j = 1, 2, \dots, n$$

با توجه به این که داریم:

$$W = E(w) = E(q) + E(s)$$

$$Lq = E(N_q) = \lambda E(q)$$

$$L = E(N) = \lambda E(w)$$

لازم به ذکر است در این سیستم‌ها، فواصل ورود دارای توزیع نمائی با یک سرویس دهنده می‌باشند. ولی چون مشتری‌های مختلف سرویس‌های متفاوتی دریافت می‌کنند عملاً مدل به صورت M/G/1 تلقی می‌شود. در مورد سیستم از نوع روش اول برخی از روابط مشابه حالت قبلی است. تعدادی دیگر که

مخصوص این روش می‌باشند را در اینجا ذکر می‌کنیم [۱].

$$W_j = E\left(w_j\right) = \frac{1}{1-u_{j-1}}\left[E\left(s_j\right) + \frac{\sum_{i=1}^{j}\lambda_i E\left(s_i^2\right)}{2\left(1-u_j\right)}\right], u_0 = 0,\ j = 1,2,\ldots,n$$

$$W_{qj} = E(q_j) = E(w_j) - E(s_j)$$

$$L_{qj} = E(N_{qj}) = \lambda_j E(q_j) \quad , \quad j = 1,2, \ldots ,n$$

$$Lq = E(Nq) = \lambda E(q) = \lambda W_q$$

$$L = E(N) = \lambda E(w) = \lambda W$$

مثال ۴-۶

یک سیستم پایگاه داده‌ای دو نوع تقاضا دریافت می‌کند. نوع اول توزیع پواسن و نرخ ۰/۹ در هر ثانیه، تقاضاها وارد می‌شوند. زمان لازم برای پاسخگویی تقریباً ثابت با مقدار میانگین ۰/۴ ثانیه می‌باشد. نوع دوم تقاضای ورودی نیز دارای توزیع پواسن با نرخ متوسط یک مورد در ۱۰ ثانیه و زمان لازم جهت پاسخگویی دارای توزیع دو مرحله‌ای فوق نمائی (hyper exponential)

با مقادیر ۰/۴ = α_1 و ۱۰ = $\frac{1}{\mu_1}$, $\frac{5}{3}$ = $\frac{1}{\mu_2}$ ثانیه می‌باشند. همچنین زمان سرویس برای نوع دوم ۵ ثانیه می‌باشد، بطوریکه گشتاور دوم آن ۸۳/۳۳ ثانیه است. مطلوبست:

۱- محاسبه متغیرهای سیستم بدون هیچگونه تقدم

۲- HOL که تقدم را به نوع دوم می‌دهد.

۳- مطلوبست محاسبهٔ متغیرها با تقدم نوع اول

حل ۱: با توجه به روابط ذکر شده برای این قسمت داریم:

$$E(s) = 0/9 \times 0/4 + 0/1 \times 5 = 0/86$$ ثانیه

$$E(s^2) = 0/9 \times 0/4^2 + 0/1 \times 83/33 = 8/477$$ ثانیه

$$\rho = \lambda.E(s) = 0/86$$ ثانیه

$$Wq = E(q) = \frac{\lambda E(s^2)}{2(1-\rho)} = 30/275$$ ثانیه

$$W_1 = E(q) = E(s) = 30/275 + 0/4 = 30/675$$ ثانیه

$$W_2 = 30/275 + 5 = 35/275$$ ثانیه

$$W = Wq + E(s) = 30/275 + 0/86 = 31/135$$ ثانیه

حل ۲: HOL که نوع دوم دارای تقدم می‌باشد:

$$W_{q1} = \frac{\lambda E(s)^2}{2(1-u_1)} = 6/6227 \text{ ثانیه}$$

$$W_{q2} = \frac{\lambda E(s)^2}{2-(1-u_1)-(1-u_2)} = 47/3047 \text{ ثانیه} \;,$$

$$W_1 = W_{q_1} + E(s_1) = 7/0227 \text{ ثانیه}$$

$$W_2 = W_{q_2} + E(s_2) = 52/3047 \text{ ثانیه}$$

متوسط کل زمان در صف

$$W_q = 0/9 \times 6/622 + 0/1 \times 47/3047 = 10/6909 \text{ ثانیه}$$

متوسط زمان انتظار کل در سیستم

$$W = W_q + E(s) = 11/5509 \text{ ثانیه}$$

حل ۳: تقدم از نوع اول:

$$W_1 = E(s_1) + \frac{\lambda_1 E(s_1)^2}{2(1-u_1)} = 0/4 + \frac{0/9\times 0/16}{2(1-0/36)} = 0/5125 \text{ ثانیه}$$

$$W_2 = \frac{1}{(1-u_1)}\left(E(s_2) + \frac{\lambda_1 E(s_1)^2 + \lambda_2 E(s_2)^2}{2(1-u_2)}\right)$$

$$= \frac{1}{1-0/36}\left(5 + \frac{0/9\times 0/16 + 0/1\times 83/33}{2(1-0/86)}\right) = 55/1172 \text{ ثانیه}$$

متوسط زمان صف برای دو نوع تقاضا، عبارتند از:

$$Wq_1 = E(q_1) = E(w_1) - E(s_1) = W_1 - E(s_1) = 0/1125 \text{ ثانیه}$$

$$Wq_2 = E(q_2) = E(w_2) - E(s_1) = W_2 - E(s_2) = 50/1172 \text{ ثانیه}$$

همچنین برای کل زمان مصرف شده در صف داریم:

$$Wq = 0/9 \times 0/1125 + 0/1 \times 50/1172 = 5/1173 \text{ ثانیه}$$

و میانگین زمان انتظار عبارتست از:

$$W_1 = 0/5125 \text{ ثانیه}$$

$$W_2 = 55/1172 \text{ ثانیه}$$

$$W = Wq + E(s) = 5/1130 + 0/86 = 5/973 \text{ ثانیه}$$

جدول ۴-۴ نتایج هر سه قسمت را خلاصه نموده است:

جدول ۴-۴ مقایسه زمانهای انتظار در صف و سیستم در حالات مختلف

	NO PRIORITY	HOL	P.RESUME
Wq_1 (نوع اول)	۳۰/۲۷۵	۶/۶۲۲۷	۰/۱۱۲۵
Wq_2 (نوع دوم)	۳۰/۲۷۵	۴۷/۳۰۴۷	۵۰/۱۱۷۲
W_1	۳۰/۶۷۵	۷/۰۲۲۷	۰/۵۱۲۵
W_2	۳۵/۲۷۵	۵۲/۳۰۴۷	۵۵/۱۱۷۲
Wq	۳۰/۲۷۵	۱۰/۶۹۰۹	۵/۱۱۳۰
W	۳۱/۱۳۵	۱۱/۵۵۰۹	۵/۹۷۳

۵

شبکه های صفی

۵-۱ مقدّمه

در سیستم‌های صفی که در فصل قبلی بررسی کردیم مشتریان (برنامه و ....) معمولاً با توزیع پواسن به سیستم وارد شده و پس از دریافت سرویس از سیستم خارج می‌شوند. در سیستم‌های واقعی (سیستم‌های کامپیوتری، مخابراتی و ....) توزیع ورودی به یک صف، معمولاً خروجی حاصل از صف دیگری است.

همچنین ممکن است ورودی به یک صف با ورودی‌های جدید دیگری نیز ترکیب شود. به عنوان مثال در یک شبکهٔ کامپیوتری که انتقال پیغام، توسط بسته‌هایی تحت پروتکل خاصّ صورت می‌گیرد، بیشتر بسته‌ها (Packet) که در حقیقت مشتریان سیستم می‌باشند، پس از انتقال از یک سوئیچ (سرویس دهنده) به بقیه سوئیچ‌ها رفته و از آنجا نیز به بقیه نقاط ارسال می‌شوند، تا در نهایت به مقصد برسند.

در هر سوئیچ تعدادی بسته ممکن است اضافه شود، که این امر نشان دهنده ورود پیغام‌های (مشتری) جدید به شبکه می‌باشد. همچنین تعدادی بسته از هر سوئیچ خارج می‌شود که معرف پیغام‌هایی که به مقصد رسیده‌اند.

شبکه‌های صفی مانند مثال فوق در مواردی که تعداد مشتری‌ها زیاد بوده و مستقل از یکدیگر عمل می‌نمایند، مدل مناسبی می‌باشند. معمولاً ظرفیت سیستم را ثابت فرض نموده و فقط توزیع مشتریان در بین صفها تغییر می‌کند.

به عنوان مثال، برای این موارد می‌توان از سیستم‌های کامپیوتری دسته‌ای (Batch) با سطح ثابت چند برنامگی (Multiprogramming) نام برد. در این سیستم‌ها، تعداد ثابتی برنامه وجود دارد که بعضی از آنها منتظر دریافت سرویس CPU و برخی منتظر دریافت سرویس دیسک و تعداد نیز در حال دسترسی به

دیسک می‌باشند. تعدادی از برنامه‌ها نیز اعمال I\O انجام می‌دهند. هنگامی که یک برنامه خاتمه می‌یابد، زمانبند (scheduler) مرکزی آن را بایک برنامه دیگر جایگزین می‌نماید. نکته مهم آن است که کل تعداد برنامه‌های فعال، ثابت بوده ولی توزیع آنها روی دیسک، I\O و CPU می‌تواند تغییر کند. (مثلاً در مقطعی از زمان ممکن است کارهای I\O بیشتر از CPU باشد).

مدلهای دیگری از ترکیب این نوع مدلها می‌تواند تشکیل شود. به عنوان مثال، تعدادی ترمینال را در نظر می‌گیریم که هر کاربر می‌تواند یک پردازش به سیستم و یا پردازش دسته‌ای به صف پردازش مربوطه ارسال کند. تعداد کل برنامه‌های ترمینالها ثابت است ولی پردازش دسته‌ای در صف مربوطه متغیر می‌باشد. در تحلیل بسیاری از سیستمهای واقعی به جای یک صف نیاز به شبکه‌ای از صفهای مختلف، جهت مدل کردن سیستم خواهیم داشت. در بخشهای بعدی به بررسی شبکه‌های مختلف صف می‌پردازیم.

۵-۲ شبکه صفی Tandem

ساده ترین شبکهٔ صفی دو یا چند صف M/M/۱ است که در یک ردیف منظم، پشت سرهم قرار می‌گیرند. همانطور که در شکل ۵-۱ دیده می‌شود، مشتریان به اولین سرویس دهنده وارد شده و پس از دریافت سرویس به صف بعدی می‌روند.

در مورد تعداد مشتریانی که برای هر سرویس دهنده منتظر هستند، محدودیتّی وجود ندارد.

فرض می‌کنیم ورود به اولین صف دارای توزیع پواسن با نرخ λ می‌باشد. زمان سرویس دهی دارای توزیع نمائی با میانگین $\frac{1}{\mu_1}$ (در صف اول) و $\frac{1}{\mu_2}$ (در صف دوم) می‌باشد. برای یک مشتری، زمان سرویس دهی در صف‌های اول و دوم مستقل از یکدیگر می‌باشد. برای اینکه سیستم به حالت تعادل برسد باید نرخ سرویس دهی سرویس دهنده از نرخ ورودی بیشتر باشد.

با فرض آنکه مشتری از یک صف مستقیماً به صف بعدی می‌رود، بنابراین نرخ ورود هر صف فقط به نرخ خروج صف قبلی آن بستگی دارد و در نتیجه پردازشها به صورت مارکف می‌باشند. وضعیت سیستم به صورت زوج‌هایی که معرف تعداد مشتری‌ها در هر صف میباشند، بیان می‌گردد [۷].

احتمال آنکه i مشتری در صف اول و j مشتری در صف دوم باشد = $\Pi_{i,j}$

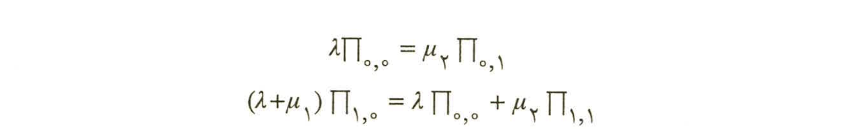
$$\lambda\Pi_{0,0} = \mu_2\Pi_{0,1}$$

$$(\lambda+\mu_1)\Pi_{1,0} = \lambda\Pi_{0,0} + \mu_2\Pi_{1,1}$$

$$(\lambda+\mu_2)\Pi_{0,1} = \mu_1\Pi_{1,0}$$

$$(\lambda+\mu_1+\mu_2)\Pi_{1,1} = \lambda\Pi_{0,1} + \mu_1\Pi_{2,0} + \mu_2\Pi_{1,2}$$

$$(\lambda+\delta(i)\mu_1+\delta(j)\mu_2)\Pi_{i,j} = \lambda\Pi_{i-1,j} + \delta(j)\mu_1\Pi_{i+1,j-1} + \delta(j+1)\mu_2\Pi_{i,j+1}$$



شکل ۵-۱ شبکهٔ صفی M/M/۱

$\delta(\circ)$را "تابع دلتای کرونکر" تعریف می‌کنند که عبارتست از:

$$\delta(i)=\begin{cases} ۱ & i>\circ \\ \circ & i\leq\circ \end{cases}$$

صف اول، متأثر از اتفاقات صف دوم نیست بنابر این احتمال وجود i مشتری در صف اول عبارتست از:

$$P_r(i,\circ)=\varphi_۱^{\,i}(1-\varphi_۱)\ ,\ \varphi_۱=\frac{\lambda}{\mu_۱}$$

در مورد صف دوم همانطور که فرآیند ورودی به سیستم به صورت تصادفی با نرخ λمی‌باشد، صف دوم نیز فرآیند تصادفی بوده و اگر این فرآیند پواسن باشد (در حالت کلی فرآیند تجدید (Renewal Process) می‌باشد) صف دوم از صف اول مستقل خواهد بود [۷]. پس داریم:

$$P_r(o,i)=\varphi^{i}_۲\ (۱-\varphi_۲)\ ,\ \varphi_۲=\frac{\lambda}{\mu_۲}$$

$$\pi_{i,j}=\varphi_۱^{\ i}(۱-\varphi_۱)\ \varphi_۲^{\ j}\ (۱-\varphi_۲)$$

$\pi_{i,j}$، احتمال وجود i مشتری در صف اول و j مشتری در صف دوم را تعیین می‌کند.

صف دوم دقیقاً یک M/M/۱ مجزا با نرخ ورودی λمی‌باشد. لازم به ذکر است که اگر خروجی صف اول پواسن باشد، دلیلی وجود ندارد که خروجی صف دوم (خروجی سیستم) پواسن باشد. هر چند که طول صف در دو سیستم می‌تواند یکسان باشد.

در یک صف M/M/۱، اگر خروجی قبلی صف را در حالی ترک کند که در سیستم مشتری وجود دارد، خروجی بعدی وقتی اتفاق می‌افتد که مشتری در حال سرویس، سروسیش را تمام کند. به عبارت دیگر توزیع فاصلهٔ بین دو خروجی همان توزیع زمان سرویس می‌باشد. حال اگر خروجی قبلی صف را در حالی ترک کند که هیچ مشتری در سیستم نیست، خروج بعدی وقتی اتفاق می‌افتد که مشتری بعدی وارد شود و سرویس دریافت نماید. به عبارت دیگر فاصلهٔ دو خروجی عبارتست از زمان بین دو ورودی بعلاوه زمان سرویس دهی. بنابراین توزیع زمانهای بین دو ورود همانند توزیع زمان بین دو خروج می‌باشد. پس سرویس دهنده دوم مستقل از اولی تحلیل می‌شود. این خاصیت علاوه بر صفهای M/M/۱ در صفهای M/M/C و M/G/∞ وجود دارد [۷].

به طور کلی در یک شبکهٔ صفی پشت سر هم (Tandem)، هر صف را مستقل در نظر گرفته و تحلیل و بررسی می‌کنیم. در این شبکه‌ها، نتیجه ادغام چندین فرآیند پواسن مستقل، خود فرآیند پواسن تشکیل می‌دهد. همچنین اگر توزیع پواسن به چندین قسمت تقسیم شود، هر یک از آنها یک فرآیند پواسن با نرخ متفاوت، خواهد بود. برای اثبات این مطلب یک فرآیند پواسن با نرخ λرا در نظر می‌گیریم. احتمال آنکه یک ورودی در زمان کوتاه $\delta(t)$اتفاق بیفتد عبارتست از:

$$P_r(\delta(t)\ \text{در}\ \text{یک ورودی})=\lambda\delta(t)+o(\delta(t))$$

$$P_r(\delta(t)\ \text{در}\ \text{صفر ورودی})=۱-(\lambda\delta(t)+o(\delta t))$$

$P_r(\delta(t)$ چندین ورودی در $) = o(\delta t)$

فرآیند خروجی از ادغام دو فرآیند پواسن با پارامترهای λ_1 , λ_2 تشکیل شده

$$=[\lambda_1\delta(t)+o(\delta t)].[1-\lambda_2\delta(t)+o(\delta t)]+[\lambda_2\delta(t)+o(\delta t)].[1-\lambda_2\delta(t)+o(\delta t)]$$

بنابراین داریم:

$P_r(\delta(t)$ یک خروجی در فاصله) = P_r (یک خروجی در فرآیند اول و صفر خروجی در فرآیند دوم) (صفر خروجی در فرآیند اول و یک خروجی در فرآیند دوم) $+P_r$

پس از عملیات ضرب روی عبارت فوق، به استثناء عبارت $(\lambda_1+\lambda_2)\delta(t)$، جمله‌ای بزرگتر از $o(\delta t)$ وجود ندارد. بنابراین احتمال وجود بیش از یک ورودی برابر $o(\delta t)$ و احتمال آنکه هیچ ورودی موجود نباشد برابر است با:

$P_r(\delta(t)$ هیچ ورودی در $)= 1-(\lambda_1+\lambda_2)\delta(t)+o(\delta t)$

که همان نتیجه‌ای است که در یک فرآیند پواسن با نرخ $\lambda_1+\lambda_2$ مشاهده می‌شود. با تحلیل مشابه می‌توان نشان داد که یک فرآیند پواسن پس از تقسیم شدن به چندین فرآیند، فرآیندهای پواسن با نرخ‌های متفاوت را پدید می‌آورد.

البته لازم به ذکر است در مورد تقسیم شدن یک فرآیند پواسن (به صورت تصادفی) محدودیتی وجود دارد. اگر یک فرآیند پواسن نرخ ارسال برنامه‌ها به دسته‌های تقسیم شده را مرتباً تغییر دهد، دسته‌های خروجی حاصل، فرآیند پواسن نخواهند بود. به طور نمونه یک فرآیند پواسن با نرخ λ را در نظر می‌گیریم. یک ورودی به طور مستقل با احتمال P به اولین فرآیند خروجی و با احتمال $q=1-p$ به دومین فرآیند خروجی تخصیص یافته است. $N_i(t)$ را احتمال تخصیص یافتن i ورودی در فاصله زمانی t به فرآیند خروجی اول، در نظر می‌گیریم. بنابراین:

$N_1(\delta(t))= P_r$ (یک ورودی در $\delta(t)$ رسیده و به فرآیند خروجی اول رفته)

$+P_r$ (چند ورودی در $\delta(t)$ رسیده و فقط یکی از آنها به فرآیند خروجی اول رفته)

$=P_r(\delta(t)$ یک ورودی در$).P_r($ورودی به فرآیند خروجی اول برود$)+O(\delta t)$

$= (\lambda\delta(t)+o(\delta t)).P+o(\delta t)$

$= P\lambda\,\delta(t)+o(\delta t)$

همچنین احتمال آنکه هیچ ورودی به فرآیند خروجی اولی اختصاص نیابد عبارتست از:

$N_0(\delta t)=P_r$ (هیچ ورودی در $\delta(t)$ نیاید)

$+P_r$ (یک ورودی در $\delta(t)$ بیاید و به فرآیند دوم برود)

$+P_r$ (چند ورودی در $\delta(t)$ بیایند و همه به فرآیند دوم بروند)

واضح است که برای $i \geq 2$ داریم؛ $Ni(\delta t)= o(\delta(t))$ و احتمالات به دست آمده برای یک فرایند پواسن با نرخ $P\lambda$ بدیهی می‌باشد.

مدل شبکهٔ صفی پشت سر هم، شامل صفهایی که هر یک از آنها می‌تواند چندین دسته ورودی داشته باشد و همچنین هر کدام به چند دسته خروجی میتواند تقسیم گردد. هر صف به صورت مستقل تحلیل

می‌شود. در این مدل تنها محدودیت آن است که در صف نمی‌توان بازخورد(Feed back) داشت. به عبارت دیگر یک برنامه نمی‌تواند صفی را ترک کند و دوباره به همان صف در شبکه برگردد.

۵-۳ شبکه‌های صفی دارای بازخورد

در این شبکه‌ها ممکن است یک مشتری صفی را که ترک کرده، دوباره به آن برگردد. البته توزیع ورودی نیازی نیست پواسن باشد. شکل ۵-۱ را در نظر بگیرید مشتری‌ها با توزیع پواسن وارد سیستم شده و سرویس با توزیع نمائی دریافت می‌کنند. سپس با احتمال p سیستم را ترک می‌نمایند و با احتمال $q=1-p$ از طریق بازخورد به صف برمی‌گردند.

با در نظر گرفتن دیاگرام، سیستم شبیه یک مدل M/M/۱، با نرخ ورود λ و نرخ سرویس دهی $p\mu$ رفتار می‌کند.

ورودی کل سیستم از دو بخش تشکیل شده است. یکی ورودی های خارجی که دارای توزیع پواسن می‌باشد و دیگری ورودی‌هایی که پس از دریافت سرویس، از طریق بازخورد (Feed-back) به سیستم برمی‌گردند. این دو ورودی مستقل از یکدیگر پردازش می‌شوند. چون ورودی‌های حاصل از بازخورد (Feed-back) فقط در لحظه‌های اتمام سرویس دهی تولید می‌شوند، نمی‌توانند از زمانهای سرویس‌دهی مستقل باشند.

فواصل زمانی بین مشتریانی که سرویس دهنده را ترک می‌کنند و آنهایی که بازخورد (Feed-back) می‌شوند، لزوماً توزیع نمائی نیست. البته آن مشتریانی که از سیستم خارج می‌شوند دارای توزیع پواسن می‌باشند.

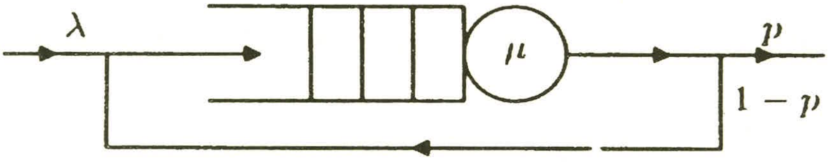
۵-۴ شبکهٔ صفی جکسون jackson

این شبکه شامل M صف M/M/۱ می‌باشد. مشتری‌ها از بیرون وارد سیستم شده و با نرخ λ_i به صف iام می‌روند. پس از گرفتن سرویس در صف i (که دارای توزیع نمائی با پارامتر μ_i می‌باشد) مشتریان، به احتمال P_{io} سیستم را ترک کرده و به احتمال P_{ij} به صف j می‌روند.

واضح است که:

$$\sum_{j=0}^{M} Pij = 1$$

سیستم کاملاً مارکف و هر حالت آن توسط بردار اعداد صحیح غیر منفی، که معرف طول صف در هر گره شبکه می‌باشد، تعیین می‌شود. بردار $n=(n_1,n_2,....,n_m)$ حالتی که تعداد n_1 مشتری در صف اول و n_2 مشتری در صف دوم و .... وجود دارند.



شکل ۵-۲

فرض می‌کنیم سیستم به حالت تعادل (Equiblirium) رسیده است (این حالتی است که شبکه در وضعیت تعادل ورود و خروج، طول صف و .... قرار دارد و جهت بررسی سیستم مهم می‌باشد).
احتمال آنکه n حالت تعادل سیستم باشد، با $\pi(n)$ نمایش داده میشود. معادلات تعادل به صورت زیر نوشته می‌شوند [۷].

$$\pi(n)\Big(\sum_{i=1}^{M}\lambda_i+\sum_{i=1}^{M}\delta(n_i)\,\mu_i\Big)=\sum_{i=1}^{M}\delta(n_i)\,\lambda_i\,\pi(n-e_i)$$
$$+\sum_{i=1}^{M}P_{io}\,\mu_i\,\pi(n-e_i)$$
$$+\,\delta\Big(\sum_{i=1}^{M}\sum_{j=1}^{M}(n-j)P_{ij}.\mu_i\pi(n+e_i-e_j)$$

$\delta(0)$ را تابع دلتای "کرونکر" می‌نامیم که قبلاً تعریف نموده‌ایم.

$$\delta(k)=\begin{cases}1 & k>0\\ \\ 0 & k\leq 0\end{cases}$$

ej برداری است که به جز j–امین عنصر آن که یک می‌باشد بقیه عناصر آن صفر هستند. دقت شود که عبارت $\delta(n_i)\;P_{ij}\,\mu_i\,\pi\;(n)$ در دو طرف معادله تعادل فوق ظاهر می‌شود و معرف مشتریانی است که گره i را ترک نموده و بلافاصله به همان گره i مجدداً بر می‌گردند. اکنون کل نرخ خروجی (Throughput) مربوط به گره i را می‌یابیم. این نرخ شامل مشتریانی است که از بیرون شبکه با نرخ λ_i به گره i رسیده و همچنین مشتریانی که پس از دریافت سرویس در گره j، به گره i منتقل شده‌اند (برای همه گره‌های j در شبکه).
اگر γj نرخ خروجی گره j باشد، پس نرخ مشتریانی که از گره j به گره i می‌رسند برابر است با $Pji\,\gamma j$.
بنابراین نرخ خروجی گره i (γi) بایستی معادلات ترافیک را ارضاء نماید یعنی:

$$\gamma_i=\lambda_i+\sum_{j=1}^{M}P_{ji}\,\gamma_j$$

با تعریف $\rho_i=\dfrac{\gamma_i}{\mu_i}$، احتمال حالت تعادل، $\Pi(n)$، به صورت زیر خواهد بود:

$$\pi(n)=\prod_{i=1}^{M}\rho^{ni}{}_{i}\;(1-\rho_i)$$

با جایگزینی این رابطه در معادلات تعادل و ساده سازی آن، مدل صف مذکور به سادگی تحلیل می‌شود. در این شبکهٔ صفی تمام صفهای M/M/1 مستقل از یکدیگر می‌باشند. این نکته مهم است که احتمال حالت تعادل وقتی بررسی می‌شود که جریانهای ورودی پواسن باشند ولی در سیستمهای واقعی جریانها پواسن نیستند.

احتمال وضعیت شبکه (درهر وضعیت ممکن) برابر حاصلضرب احتمالات وضعیت هر صف به صورت مجزا می‌باشد. این شبکه‌های فرم حاصلضرب را، شبکهٔ "Jackson" می‌نامند. در این شبکه‌ها جمعیت ثابت نیست و هنگام ورود به سیستم و یا خروج از آن می‌توان تغییر نماید.

۵-۵ شبکه های صفی بسته (Closed queueing networks)

شبکه‌های صفی که تعداد ثابتی مشتری بین صفهای مختلف آن در گردش می‌باشند، را شبکهٔ صفی بسته می‌گویند. مدلهایی نظیر "مدل تعمیر کار ماشین" و "مدل سرویس دهنده مرکزی در یک سیستم کامپیوتری از نمونه‌های شبکه صفی بسته می‌باشند. به طور مثال در مدل سرویس دهنده مرکزی که مدلی از سیستم‌های دسته‌ای با سطح ثابت چند برنامگی می‌باشد، هر برنامه پس از دریافت سرویس cpu، از دستگاههای ورودی خروجی (مانند دیسک و....) نیز سرویس دریافت می‌نماید (قبل از دریافت سرویس مجدد از cpu). در این سیستم‌ها کل تعداد برنامه‌ها ثابت بوده و چندین سرویس دهنده داریم به طوریکه وقتی یک job (برنامه) سیستم را ترک می‌کند، یک برنامه دیگر جایگزین آن می‌گردد. مدل مذکور در شکل ۵-۲ ذیل نمایش داده شده است.

p_i عبارت از احتمال دسترسی به دستگاه I/O شماره i–ام $(i=۲,... m)$ و $P_۱$، احتمال دسترسی مجدد به cpu می‌باشد.

تحلیل شبکه‌های صفی بسته را با شکل ذیل که دارای سه مرحله متفاوت سرویس دهی با توزیع نمائی می‌باشد، دنبال می‌کنیم. مجموعهٔ وضعیت سیستم را به صورت زیر نشان می‌دهیم.

$$S= \{N_۱,N_۲,N_۳\}$$

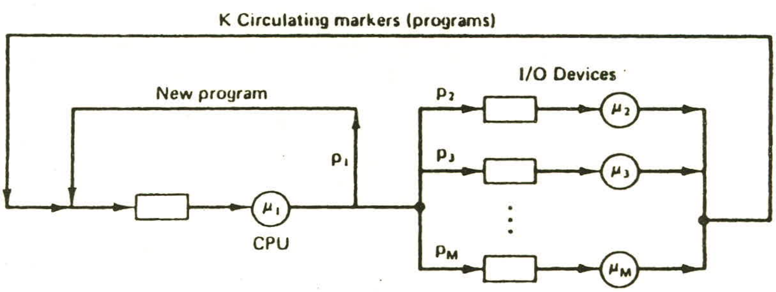
که Ni , $i=$ ۱ و ۲ و ۳ تعداد مشتری در صف i–ام می‌باشد.

اگر ۲ مشتری در سیستم موجود باشد در این صورت دیاگرام انتقال وضعیت سیستم به صورت زیر می‌باشد:

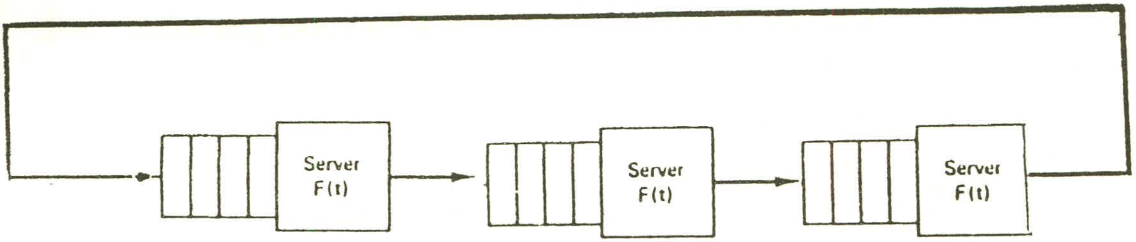
$S_i \to S_j$ نشان دهنده انتقال از مرحله i به j در سیستم می‌باشد.

با توجه به شکل ۵-۴ احتمال وضعیت یکنواختی سیستم از طریق معادلات تعادل جریان محاسبه می‌گردد. بنابراین برای وضعیتهای مختلف سیستم داریم:

$$\pi(۲,۰,۰)\mu_۱ = \pi(۱,۰,۱)\mu_۳$$
$$\pi(۱,۱,۰)(\mu_۱+\mu_۲) = \pi(۲,۰,۰)\mu_۱ - \pi(۰,۱,۱)\mu_۳$$
$$\pi(۰,۲,۰)\mu_۲ = \pi(۱,۱,۰)\mu_۱$$
$$\pi(۱,۰,۱)(\mu_۳+\mu_۱) = \pi(۱,۱,۰)\mu_۲ + \pi(۰,۰,۲)\mu_۳$$
$$\pi(۰,۱,۱)(\mu_۲+\mu_۳) = \pi(۰,۲,۰)\mu_۲ + \pi(۱,۰,۱)\mu_۱$$
$$\pi(۰,۰,۲)\mu_۳ = \pi(۰,۱,۱)\mu_۲$$



شکل ۵-۳ مدل سرویس دهنده مرکزی



شکل ۵-۴ شبکهٔ صفی بسته با سه مرحله

بطوریکه $\Pi\ (N_۱,N_۲,N_۳)$ برابر است با احتمال وضعیت $(N_۱,N_۲,N_۳)$.
با توجه به اینکه مجموع همه احتمالات برابر یک است، خواهیم داشت [۸]:

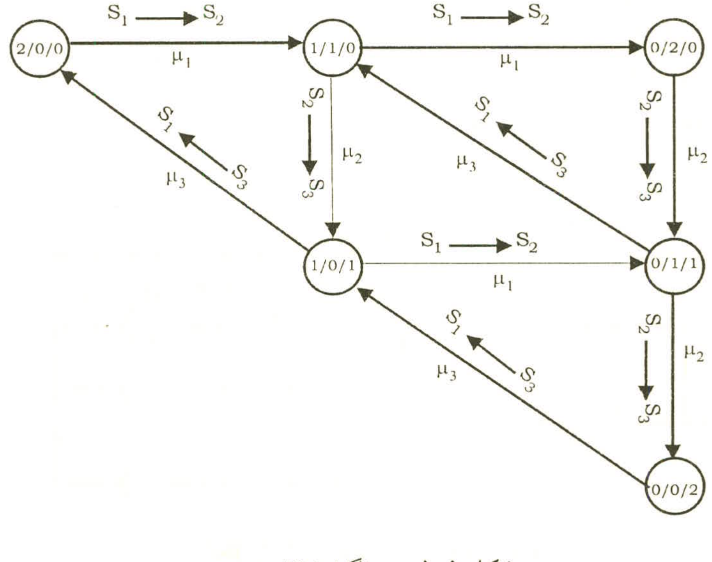
$$\Pi\ (N_۱,N_۲,N_۳)= K\left(\frac{۱}{\mu_۱}\right)^{N_۱}\left(\frac{۱}{\mu_۲}\right)^{N_۲}\left(\frac{۱}{\mu_۳}\right)^{N_۳}$$

در عبارت فوق K ثابت نرمال سازی است که تضمین می‌کند مجموع احتمالات برابر یک شود بنابراین داریم:

$$K=\frac{۱}{\sum_{i=۱}^{۲}\sum_{j=۰}^{۲}\sum_{k=۰}^{۲}\Pi\ (i,j,k)}$$

معادلات تعادل جریان را برای هر وضعیت به طور جداگانه حل می‌کنیم. با داشتن احتمالات وضعیت‌های سیستم، میانگین طول صف برای هر سرویس دهنده بدست می‌آید:

$$E\ (Nq)\ =\sum_{i=۱}^{\text{تعداد وضعیت}} iP\ (N=i)$$



شکل ۵-۵ دیاگرام انتقال وضعیت

چون احتمالات وضعیت در صفها، مانند وضعیت سیستم‌های صف مجزا نیست، پس نمی‌توان زمان انتظار در یک صف را بوسیلهٔ حاصلضرب ساده تعداد مشتریان و زمان سرویس دهی صف محاسبه نمود. ابتدا باید برای هر سیستم صف نرخ خروجی را محاسبه کنیم. سپس با استفاده از قضیه لیتل (Little's Law) که دریک شبکه صف متوسط تعداد مشتری N و متوسط نرخ ورود λ و متوسط زمان انتظار در سیستم W باشد خواهیم داشت [۸]:

$$N = \lambda . W$$

بنابراین از روی نرخ‌های خروجی بدست آمده، مقداری برای نرخ ورودی هر صف بدست می‌آوریم. بدین ترتیب نرخ خروجی در یک صف خاص توسط ضرب احتمال وجود یک مشتری در صف (یا مشغول بودن سرویس دهنده) در میانگین نرخ سرویس دهی بدست می‌آید.

یعنی: $\lambda_i = P$ (مشغول بودن سرویس دهنده i – اُم) $\times \mu$

اکنون با استفاده از نتایج قضیهٔ لیتل، زمان سبری شده در هر صف، توسط یک مشتری را محاسبه می‌کنیم:

$$Wq = \frac{E(N_i)}{\lambda_i}$$

کل زمان انتظار یک مشتری در سیستم برابر با مجموع همه زمانهای انتظار در صف است. همچنین میانگین نرخ خروجی نیز مستقیماً از قضیه لیتل بدست می‌آید. بنابراین برای ۲ مشتری داریم:

$$Wq = \frac{۲}{\lambda_{ave}}$$

در حالت کلی یک شبکه صفی بسته دلخواه با M صف و N مشتری در نظر می‌گیریم. همانطور که در شکل ۵-۵ دیده می‌شود، تمامی سرویس دهنده‌ها زمان سرویس دهی با توزیع نمائی دارند. هر مشتری در شبکه با یک احتمال صف خاصی را انتخاب می‌کند. پس داریم:

$p_{ij} =$ احتمال آنکه یک مشتری سرویس دهندهٔ i اُم را ترک کند و به صف jام برود

برای هر سرویس دهندهٔ j داریم:

$$\sum_j p_{ij} = 1$$

تعادل جریان (ورود و خروج مشتری) در صورتی برقرار می‌شود که:

$$\lambda_i = \sum_{i=۱}^{M} \lambda_i . P_{ij}$$

از طرف دیگر برای یک سرویس دهنده i نرخ خروجی مربوطه را به صورت زیر تعریف می‌کنیم:

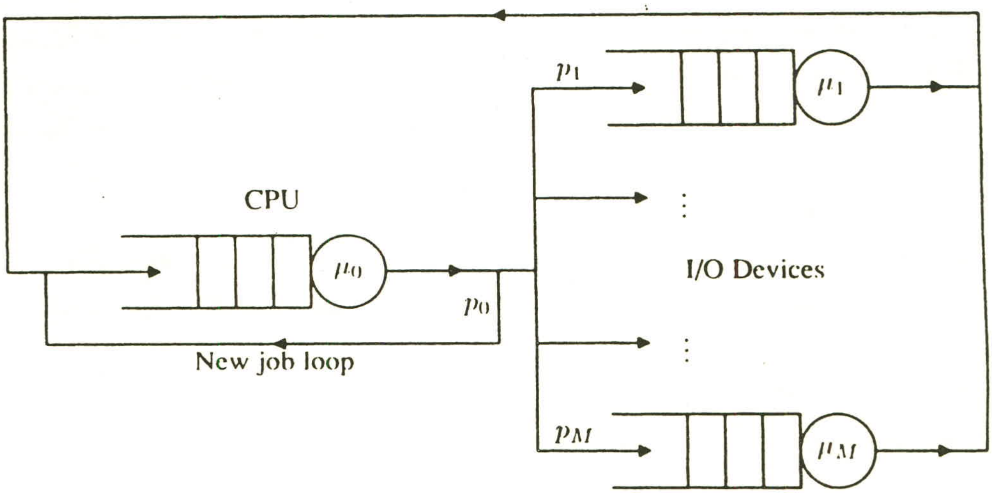
$$B(j) = \sum_{i=۱}^{M} B(i)\, P_{ij}$$

می‌توان به طور دلخواه یکی از عبارات B رابرابر یک قرارداده و بقیه جملات B را محاسبه کنیم. پس از پیدا کردن تمامی جملات، احتمال وضعیت یکنواختی سیستم به صورت زیر تعیین می‌گردد:

$$P[N_۱, N_۲, \ldots, N_M] = K \prod_{i=۱}^{M} \left[\frac{B(i)}{\mu_i}\right]^{Ni}$$

معادله B(j) با فرض برقراری تعادل جریان برای یک وضعیت خاص براحتی حل می‌گردد و سپس با حل

Closed queucing networks



شکل ۵-۶ یک شبکه صفی بسته نمونه

مجموعه معادلات، احتمالات را بدست می‌آوریم.

وضعیت S را به صورت $S=(K_1,K_2, \ldots K_M)$ در نظر می‌گیریم. اثرات ورود و خروج مشتری‌ها را در صف j اُم بررسی می‌کنیم. وضعیت A را با وضعیت S بصورت زیر تعریف می‌کنیم. در وضعیت A یک مشتری در صف i اُم بیشتر و یک مشتری در صف jاُم کمتر از وضعیت S وجود دارد. در حقیقت A یک وضعیت مجاور و نزدیک به S می‌باشد.

بدیهی است که نرخ ورود و نرخ خروج صف jاُم وضعیت S، برابر می‌باشد (سیستم در حالت یکنواخت). چون ممکن است بیش از یک وضعیت A داشته باشیم. پس باید همه وضعیت‌های مذکور نسبت به وضعیت S تعادل داشته باشند. با معادل ساختن جریانها داریم:

$$\sum_{i=1}^{M} P(A_i)\, \mu_i P_{ij} = P(S)\, \mu_j$$

با توجه به رابطهٔ وضعیت یکنواختی سیستم که در فوق نشان داده شده خواهیم داشت:

$$P(A_i) = K \prod_{j=1}^{M} \left[\frac{B(j)}{\mu_1}\right]^{N_i} \frac{B(i)}{\mu_i}$$

$$P(S) = K \prod_{j=1}^{M} \left[\frac{B(j)}{\mu_j}\right]^{N_j} \frac{B(j)}{\mu_j} ,$$

آخرین جمله عبارات مذکور $\left(\frac{B(j)}{\mu_j}, \frac{B(i)}{\mu_i}\right)$ نشاندهندهٔ وجود یک مشتری در سرویس دهنده‌های j,i می‌باشد.

با جایگزینی این عبارات در رابطه فوق و سپس ساده سازی خواهیم داشت:

$$\sum_{j=1}^{M} B(i)\, P_{ij} = B(j)$$

که همان نتیجه‌ای که در معادله B(j) بدیهی فرض کردیم. اکنون با داشتن این معادله، می‌توان یک مجموعه معادلات تولید کرد که با مقدار گذاری یکی از جملات B (با مقدار یک) به طور همزمان حل شوند. ثابت نرمال سازی را نیز می‌توان پیدا کرد و در نهایت با حل معادله احتمال وضعیت یکنواختی سیستم محاسبات را بدست آوریم.

با استفاده از معادله B(j) می‌توان میانگین طول صف را بدست آورد. حال اگر شبکهٔ بسته را در یک پریود طولانی زمان در نظر بگیریم، نرخ خروجی مربوطه بیان کننده تعداد دفعاتی که مشتری از سرویس دهنده مربوطه، سرویس دریافت نموده است. این نسبت مشخص کننده استفاده (Utilization) سرویس دهنده‌ها می‌باشد و تعیین می‌کند کدام سرویس دهنده بیشترین بهره را داشته، که به آن "آنالیز گلوگاه" نیز می‌گویند. مقدار کار انجام شده توسط سرویس دهنده i-اُم به صورت زیر تعریف می‌گردد:

$$\text{کار انجام شده توسط سرویس دهنده i-اُم} = \frac{B(i)}{\mu_i}$$

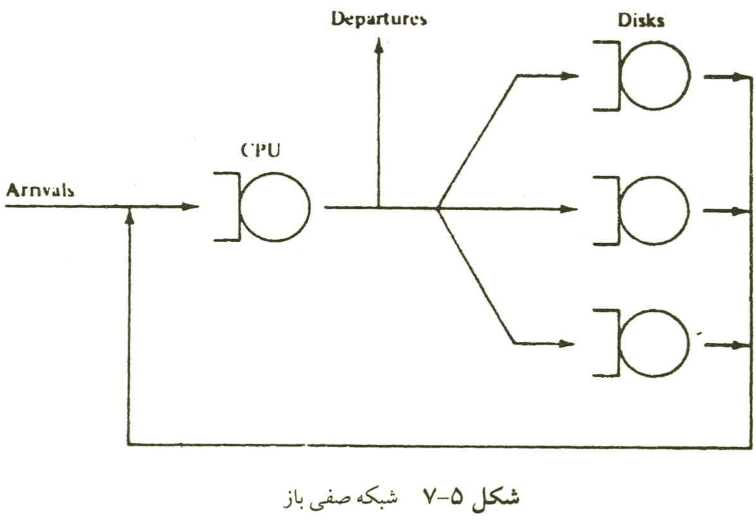
چون این مقدار بهره نسبی سرویس دهنده را تعیین می‌کند، پس سرویس دهنده با بالاترین بهره در حقیقت گلوگاه می‌باشد.

۵-۶ شبکه‌های صفی باز: (Open Queueing Networks)

شبکه‌هایی که مشتریان از یک منبع خارج به آن وارد شده و پس از دریافت سرویس از سیستم خارج گردند شبکه‌های صفی باز نامیده می‌شوند. می‌توان فرض کرد که فرآیند ورودی با احتمال هایی مختلف، جهت دریافت سرویس وارد صف مورد نظر گردیده و از سیستم خارج می‌شوند. شکل زیر یک شبکه باز نمونه را نشان می‌دهد.

فرض می‌کنیم توزیع زمان سرویس نمائی، و فرآیند ورود به سیستم پواسن باشد. در این صورت نرخ ورود به صف i-اُم برابر λ_i می‌باشد:

$$\lambda_i = P_{oi}\,\lambda \quad , \quad \lambda = \sum_{i=1}^{M} \lambda_i$$



شکل ۵-۷ شبکه صفی باز

با توصیف شبکه بسته که در بخش قبلی بررسی نمودیم، نرخ‌های خروجی $B(i)$ برای هر صف و سرویس‌دهنده i–اُم را به صورت زیر در نظر می‌گیریم:

$$B(i) = \sum_{j=1}^{M} B(j)\, P_{ji} \quad , \quad B(i) = \sum_{j=1}^{M} B(j)\, P_{ji} + \lambda_i$$

چون ورودی از یک منبع خارجی وارد سیستم می‌شود در نتیجه $\lambda = B(0)$ ، بنابراین برای بقیه $B(i)$ها، معادله فوق قابل محاسبه می‌باشد. با توجه به اینکه در شبکه‌های باز نرخ خروجی هر سرویس دهنده، بعبارتی دیگر، بکارگیری هر سرویس دهنده برابر خواهد بود با:

$$U_i = \frac{B(i)}{\mu_i}$$

همچنین احتمالات وضعیت یکنواختی سیستم برابر است با:

$$P(N_1, N_2, \ldots, N_M) = K \prod_{i=1}^{M} \left(\frac{B(i)}{\mu_i} \right)^{N_i}$$

$$= K \prod_{i=1}^{M} U_i^{N_i}$$

K ثابت نرمال سازی می‌باشد. تمامی احتمالات وضعیت را با یکدیگر جمع نموده و K را بدست می‌آوریم به عبارت دیگر داریم:

$$K = \prod_{i=1}^{M} (1 - U_i)$$

بنابراین احتمالات وضعیت یکنواختی عبارتست از:

$$P(N_1, N_2, \ldots, N_M) = \prod_{i=1}^{M} (1 - U_i)\, U_i^{N_i}$$

اگر به معادله $P_n = \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^n$، در سیستم‌های صف M/M/1 توجه کنیم، خواهیم دید عبارت فوق نتیجه حاصلضرب وضعیت‌های هر صف و سرویس دهنده، به صورت صف M/M/1 می‌باشد. این نتیجه را همانطور که قبلاً اشاره کردیم "تئوری جکسون" می‌نامند. به عبارت دیگر اگر فرآیند ورود به یک سیستم باز، پواسن باشد، تابع توزیع احتمال برای تعداد مشتریان در هر صف قابل محاسبه می‌باشد. هر چند که نرخ ورود به هر سرویس دهنده پواسن نباشد. "تئوری جکسون" در حقیقت تعیین می‌کند که هر صف، در یک شبکه صفی به صورت M/M/1، با نرخ ورودی $\alpha_i = \lambda_i + \sum_{j=1}^{M} P_{ji} \alpha_i$، رفتار می‌کند. بنابراین این مطلب درمورد شبکه‌های باز با صف‌های مجزای M/M/c می‌توان بکار برد. بطوریکه احتمالات وضعیت یکنواختی برای هر صف در شبکه بدست می‌آید. با معادلاتی که برای یک صف محزا با نرخ ورودی فوق‌الذکر (α_i) حل می‌کنیم، کل احتمال وضعیت‌های شبکه باز تعیین می‌شود.

۷-۵ تئوری BCMP:

در شبکه‌هایی که تاکنون مورد بررسی واقع شده، توزیع ورود مشتری و همچنین دریافت سرویس به ترتیب پواسن ونمائی بوده است. در صورتی که در سیستمهای کامپیوتری واقعی اینطور نیست. برنامه‌های مختلف زمانهای مختلف دریافت سرویس دارند. به طور مثال برنامه‌هایی از CPU سرویس زیادی و از I/O سرویس کمتری دریافت می‌کنند و بالعکس پردازشهایی، I/O زیادی احتیاج و CPU کمتری نیاز دارند. بنابراین بررسی این مسائل نیاز به تحلیل شبکه‌هایی از صف دارد که، دسته‌های مختلف ورودی، با سرویسهای مورد نظر، در سیستم موجود می‌باشند.

این فرآیند در تئوری BCMP قابل مطالعه و بررسی است [۹]. به طور کلی یک شبکه صفی از تعدادی گره که هرکدام دارای تعدادی صف و سرویس دهنده می‌باشد، تشکیل گردیده است. دسته‌های متعدد برنامه، مشتری و .... در ساختارهای باز و بسته شبکه، از هر گره به گره دیگری در گردش می‌باشند.
وضعیت مشتری، در یک شبکه صفی در حالت کلی به صورت ذیل تبیین می‌گردد:

۱- مشتری یا برنامه در یک صف مجزا با نظم FCFS و زمان سرویس نمائی موجود می‌باشد.
۲- چندین برنامه از دسته‌های مختلف وارد یک صف شده، و هرکدام به صورت مستقل پردازش می‌شوند (توزیع زمان سرویس مختلف دارند).
۳- برنامه‌ها بدون انتظار در صف، سرویس دریافت می‌کنند.
۴- برنامه‌هایی که دارای تقدم، در دریافت سرویس می‌باشند.

فرض کنیم سیستم شبکه دارای M گره می‌باشد. یک برنامه از دسته r-ام پس از دریافت سرویس از گره i-ام به احتمال $P_{ir,js}$ برای دریافت سرویس به گره j- ام می‌رود (لازم به ذکر است در این حالت یک برنامه در گره i- ام از دسته r و در گره j-ام از دسته s محسوب می‌شود زیرا همان برنامه در این دو گره رفتار متفاوت و سرویس دهی متفاوتی را خواهد داشت).

برنامه از دسته r، با احتمال $P_{ir,o}$ شبکه را ترک و با احتمال $P_{o,ir}$ به گره i-ام وارد می‌شود.
v_{ir} را تعداد دفعاتی که یک job از دسته r، گره i-ام را ملاقات می‌کند، در نظر می‌گیریم. اگر C را یک زنجیره از وضعیت گردش در شبکه (دسته، گره) در نظر بگیریم خواهیم داشت [۹]:

$$\sum_{i,r\varepsilon C} V_{ir}P_{ir,\,js}+P_{o,\,js}=V_{js} \qquad j,\,s\ \varepsilon\ C$$

جهت تحلیل احتمال وضعیت یکنواختی سیستم، بُردار

$$K_i=\left(K_{i1},K_{i2},\ldots,K_{iR}\right)$$

را در نظر می‌گیریم که در آن K_{ij} معرف تعداد برنامه از دسته j- ام در صف i-ام می‌باشد (R تعداد دسته‌های برنامه‌ها می‌باشد) بنابراین داریم:

$$K_i=\sum_{r=1}^{R}K_{ir}\quad,\quad K=\sum_{i=1}^{M}K_i$$

اکنون احتمال وضعیت یکنواختی سیستم برابر است با:

$$P_r\left(S=\left(K_1,K_2,\ldots,K_M\right)\right)=\frac{1}{G}d\,(s)\,f_1\left(K_1\right)\,f_2\left(K_2\right)\ldots f_m\left(K_m\right)$$

G عدد ثابت و d(s) تابعی که تعداد مشتری در سیستم را بیان می‌کند.

همچنین $f_i(k_i)$ به صورت زیر تعریف می‌گردد:

$$f_i(k_i)=\begin{cases} K_i!\prod_{r=1}^{R}\left(\frac{1}{K_{ir}!}\right)\left(\frac{V_{ir}}{\mu_{ir}}\right)^{K_{ir}} & \text{اگر گره i از نوع ۱، ۲ و ۴ باشد.(وضعیت مشتری موارد ۱، ۲ و ۴ فوق الذکر باشد)} \\ \prod_{r=1}^{R}\left(\frac{1}{K_{ir}!}\right)\left(\frac{V_{ir}}{\mu_{ir}}\right) & \text{اگر گره i از نوع ۳ باشد (مورد سوم فوق‌الذکر)} \end{cases}$$

مثال ۵-۱

مدل صفی زیر (سیستم چند برنامگی) را درنظر می‌گیریم:

فرض می‌کنیم فرآیند ورودی پواسن با نرخ λ باشد. همچنین توزیع زمان سرویس در cpu و I/O به صورت نمائی با میانگین $\frac{1}{\mu_1}$ و $\frac{1}{\mu_2}$ به ترتیب، می‌باشد. برنامهٔ ورودی پس از تکمیل سرویس cpu ، بـا احـتمال P از سیستم خارج می‌شود و با احتمال $1-P$ در صف سرویس I/O قرار مـی‌گیرد و ظـرفیت صـفها نـامحدود می‌باشد.

حل:

با توجه به روابط مذکور دربخش ۵-۶ (شبکه‌های صفی باز) این فصل و تئوری جکسون، مقادیر مورد نظر را محاسبه می‌کنیم، λ_1 را نرخ ورود به cpu در نظر گرفته بنابراین:

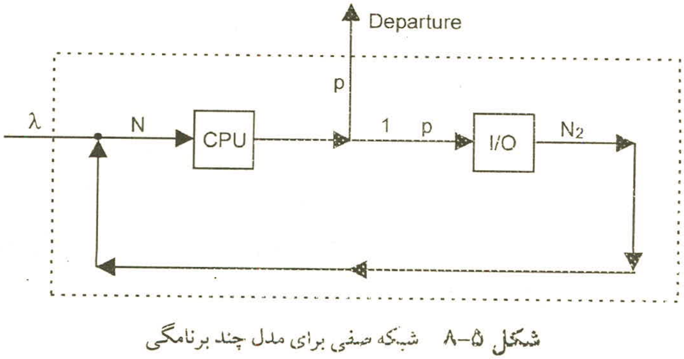
$$\lambda_1=\lambda+\lambda_2=\lambda+(1-P)\lambda_1$$

به عبارت دیگر:

$$\lambda_1=\frac{\lambda}{P}$$

بدیهی است که:

$$\lambda_2=(1-P)\lambda_1=(1-P)\frac{\lambda}{P}$$



شکل ۵-۸ شبکه صفی برای مدل چند برنامگی

بکارگیری سرویس دهنده‌ها (I/o,cpu) به ترتیب عبارتند از:

$$U_1=\frac{\lambda_1}{\mu_1}=\frac{\lambda}{P\mu_1}$$

$$U_2=\frac{\lambda_2}{\mu_2}=(1-p)\cdot\frac{\lambda}{p\mu_2}$$

میانگین نرخ خروجی سیستم برابر است با:

$$\lambda_T=P\lambda_1$$

با توجه به روابط مذکور در این فصل احتمال وجود n_1 برنامه در صف cpu و n_2 برنامه در صف I/O، عبارتست از:

$$P(n_1,n_2)=P(n_1)P(n_2)=(1-U_1)U_1^{n_1}(1-U_2)U_2^{n_2}$$

با استفاده از روابطی که در فصل ۴ ذکر شده، میانگین تعداد برنامه در سیستم برابر است با:

$$L=E(N)=E(N_1)+E(N_2)=\frac{U_1}{1-U_1}+\frac{U_2}{1-U_2}$$

میانگین زمان پاسخ با استفاده از قضیهٔ لیتل عبارتست از:

مثال ۵-۲

$$W=\frac{L}{\lambda}$$

فرض می‌کنیم تعدادی برنامه در مدل فوق، به طور میانگین ۴ ثانیه زمان cpu و به طور میانگین ۰/۲ ثانیه سرویس I/O مصرف نماید. در زمان سرویس cpu هر ۰/۲۵ ثانیه یک وقفه از I/O می‌آید. سیستم فوق جهت اجرای برنامه‌ها، ۱۰ ساعت در روز فعال می‌باشد که ۸۰۰۰ برنامه به طور متوسط پردازش می‌شوند. مطلوبست بکارگیری cpu و I/O، میانگین تعداد برنامه در صف cpu و در صف I/O، همچنین شامل خود برنامه در حال سرویس، میانگین تعداد برنامه‌ها در داخل شبکهٔ صفی و میانگین زمان پاسخ سیستم

حل:

$$\lambda=\frac{8000}{36000}=\frac{2}{9}\ \text{برنامه در ثانیه}$$

میانگین تعداد وقفه‌های I/O عبارتست از:

$$\frac{4}{0/25}=16 \text{ و } P=\frac{1}{16}$$

بنابراین داریم:

$$\lambda_1=\frac{\lambda}{P}=16\times\frac{2}{9}=\frac{32}{9}\ \text{برنامه در ثانیه}$$

$$\lambda_2=\frac{15}{16}\lambda_1=\frac{10}{3}\ \text{برنامه در ثانیه}$$

پس بکارگیری سرویس‌دهنده‌ها به ترتیب عبارتست از:

$$U_1=\lambda_1E(S_1)=\frac{32}{9}\times\frac{1}{4}=\frac{8}{9}$$

$$U_2=\lambda_2E(S_2)=\frac{10}{3}\times\frac{1}{5}=\frac{2}{3}$$

میانگین تعداد برنامه در صف Cpu و صف I/O به ترتیب برابر است با:

$$E(N_1) = \frac{U_1}{1-U_1} = 8 \text{ برنامه}$$

$$E(N_2) = \frac{U_2}{1-U_2} = 2 \text{ برنامه}$$

همچنین:

$$L = E(N) = E(N_1) + E(N_2) = 10 \text{ برنامه}$$

میانگین زمان پاسخ عبارتست از :

$$W = \frac{L}{\lambda} = 10 \times \frac{9}{2} = 45 \text{ ثانیه}$$

با توجه به اینکه یک برنامه به ۴ ثانیه سرویس Cpu و ۳ ثانیه (۱۵×۰/۲) زمان I/O نیاز دارد، زمان پاسخ ۴۵ ثانیه زیاد به نظر می‌رسد. برای بررسی این موضوع در نظر می‌گیریم که یک برنامه ۱۶ بار بین Cpu , I/O (به دلیل وقفه) انتقال می‌یابد و در هر بار انتقال داریم :

$$E(W_1) = \frac{E(S_1)}{(1-U_1)} = \frac{0/25}{\frac{1}{9}} = 2/25 \text{ ثانیه}$$

پس مجموع زمان سرویس Cpu برابر است با : ۱۶×۲/۲۵=۳۶ ثانیه

از طرفی ۱۵ بار انتقال برای دریافت سرویس I/O صورت می‌گیرد و در هر بار داریم:

$$E(W_2) = \frac{E(S_2)}{(1-U_2)} = \frac{\frac{1}{5}}{\frac{1}{3}} = 0/6 \text{ ثانیه}$$

بنابراین مجموع کل زمان I/O عبارتست از ۰/۶×۱۵ = ۹ ثانیه

و میانگین زمان پاسخ سیستم برابر است با : ۹+۳۶ = ۴۵ ثانیه

مثال ۵-۳

مدل Warn و Boyse [۱۰].

مدل مذکور، نمونهٔ جالبی از شبکه‌های صفی می‌باشد که جهت ارزیابی یک سیستم کامپیوتری با فرضهای ذیل، به کار می‌رود در این سیستم داریم:

۱- سطح ثابت چند برنامگی K موجود بوده که وقتی بار سنگین پردازش برای حافظهٔ اصلی داریم، صف آن هرگز خالی نمی‌شود.

۲- چندین Cpu یا یک Cpu در مدل، به عنوان یک سرویس دهنده، مورد مطالعه قرار می‌گیرد.

۳- K سرویس دهندهٔ I/O که به موازات یکدیگر سرویس دهی نموده و برای دریافت سرویس I/O صفی تشکیل نمی‌شود.

۴- در مورد سرویس Cpu و I/O یا هر دو، توزیع زمان سرویس نمائی است. یا آنکه هر دو دارای توزیع زمان سرویس ثابت می‌باشند.

۵- مدل‌بندی خواندن از I/O به صورت صفحه به صفحه می‌باشد بنابراین پس از آنکه یک صفحه کامل خوانده شد به صف Cpu جهت پردازش می‌رود.

۶- زمان بین دو کنش (Interaction) (زمان تفکر) (think time) از ترمینالهای کاربران، دارای توزیع عمومی با میانگین $E(t) = \frac{1}{\alpha}$ می‌باشد.

بدیهی است در سیستم‌های کامپیوتری واقعی، که توسط Boyse و Warn بررسی شد، به دلیل آنکه تعداد

مسیرهای داده برای I/O قابل دسترسی بود، فرضهای مذکور منطبق بر واقعیت بودند. اگر فرض (۳) (نبودن هیچ صفی برای سرویس I/O) برقرار نبود، میانگین زمان سرویس I/O را طوری تنظیم کنیم تا فرض مذکور برقرار شود.

اگر برای سرویس دهندهٔ I/O صف بزرگی تشکیل شود. این مدل قابل استفاده نبوده و باید از مدل سرویس دهندهٔ مرکزی (شکل ۵-۲) استفاده نمود.

در این جا هنگامی که یک برنامه در Cpu، جهت پردازش، به یک صفحه (فرض می‌شود برنامه در حافظهٔ اصلی به صفحاتی، تقسیم شده و هر بار با توجه به حجم حافظه تعدادی از آنها جهت پردازش، در حافظهٔ اصلی وجود دارند) نیاز پیدا کند اجرای برنامه قطع می‌گردد.

تا هنگامی که صفحهٔ مورد نیاز در حافظهٔ اصلی در دسترس قرار نگیرد برنامهٔ فوق قادر به اجرا نمی‌باشد (قطع سرویس Cpu)، به عبارت دیگر یک عمل I/O در این زمان شروع می‌شود. بنابراین عمل I/O و اجرای سرویس Cpu در این مورد منطبق (Overlaping) نمی‌باشد. ولی در بقیهٔ موارد فرض می‌شود که I/O بتواند با فعالیت Cpu همزمان باشد.

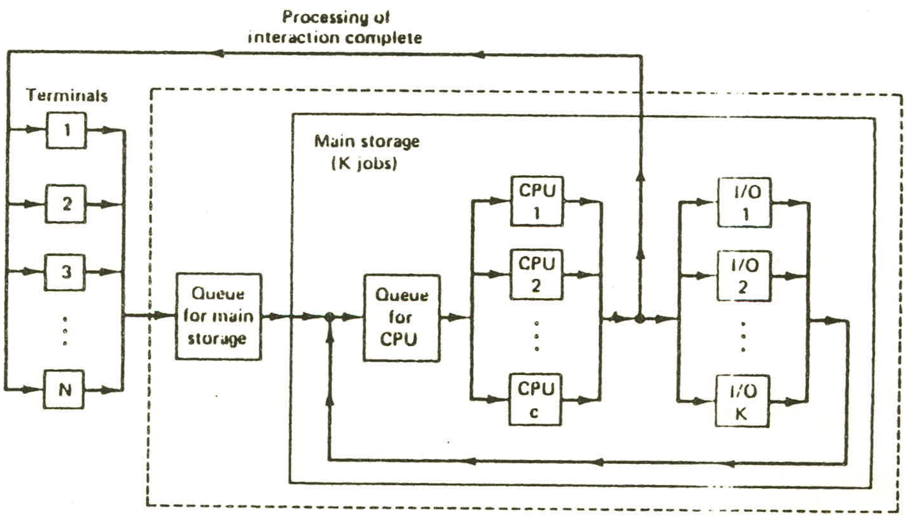
پارامترهای لازم جهت مشخص کردن مدل به شرح ذیل می‌باشند:

- E(S) میانگین استفادهٔ از Cpu .
- m میانگین تعداد فواصل زمانی که Cpu برای هر برنامه کاربر نیاز دارد.

بنابراین m.E(S) میانگین زمان Cpu برای هر برنامه کاربر می‌باشد.

- E(O) میانگین زمان سرویس یک تقاضای I/O.
- K سطح چندبرنامگی و یا تعداد سرویس‌دهنده‌های I/O، می‌باشد.
- N، تعداد ترمینالهای فعال و $E(t) = \frac{1}{\alpha}$ نیز متوسط زمان بین دو کنش از کاربر می‌باشد.

با پارامترهای فوق و با نمایش مدل (شکل ۵-۹) که در ذیل آمده به بررسی و تحلیل آن می‌پردازیم:



شکل ۵-۹ مدل Boyse, Warn

$$\rho = min \left\{ \frac{K}{C\left[۱ + \frac{E(O)}{E(S)} \right]} , ۱ \right\}$$

ρرا میانگین بکارگیری Cpu، λ_Tرا میانگین نرخ خروجی برنامه‌های کاربران و W را میانگین زمان پاسخ در نظر می‌گیریم. البته W، میانگین زمان ارسال یک تقاضا برای Cpu تا هنگام تکمیل می‌باشد. با این پارامترها اولین قدم جهت تحلیل محاسبهٔ ρبرداشته می‌شود. با وجود فرضهای مسئله می‌توان مدل تعمیرکار ماشین را درنظر بگیریم به طوریکه واحدهای I/O نقش ماشین تجهیزات و واحدهای Cpu نیز تعمیرکاران باشند. بنابراین مدل فوق یک سیستم صفی D/D/C/K/K می‌باشد. با توجه به تعریف E(S) و E(O) و سیستم صفی مذکور داریم [۱۰]:

در حالتی که زمانهای سرویس I/O , Cpu دارای توزیع نمائی باشند مدل فوق به صورت M/M/C/K/K مشخص می‌گردد که دراین صورت استفادهٔ بهینهٔ Cpu، ρ عبارتست از :

$$\rho = \frac{\lambda E(S)}{C}$$

با توجه به بخش ۱۰-۴ می‌توان نوشت:

$$\frac{P_n}{P_۰} = \begin{cases} \binom{k}{n} \left[\frac{E(S)}{E(O)}\right]^n , & n = ۱,۲,...,C \\ \frac{n!}{C!C^{n-c}} \binom{k}{n} \left[\frac{E(S)}{E(O)}\right]^n , & n = C+۱,...K \end{cases}$$

$$P_۰ = \left[۱ + \sum_{n=۱}^{k} \left(\frac{P_n}{P_۰}\right) \right]^{-۱}$$

$$Lq = \sum_{n=C+۱}^{k} (n-c)P_n$$

$$Wq = \frac{Lq\ (E(O) + E(S))}{K - Lq}$$

$$\lambda = \frac{K}{E(O) + Wq + E(S)}$$

اگر تعداد Cpu یکی باشد یعنی C=۱ در آن صورت داریم:

$$P_۰ = \left[\sum_{n=۰}^{k} \frac{K!}{(K-n)\ !} \left(\frac{E(S)}{E(O)}\right)^n \right]^{-۱}$$

در مورد میانگین نرخ خروجی داریم:

$$\lambda_T = \frac{C\rho}{m.E(S)}$$

با استفاده از تئوری Buzen [۱۱]:

$$W = \left(\frac{N.m.E(S)}{C\rho}\right) - E(t)$$

مثال ۵-۴

به عنوان نمونه اگر در مدل مذکور یک Cpu داشته باشیم و مقادیر مدل را برای مدت ۳۰ دقیقه در هنگامیکه بار پردازش سیستم بالاست، به شرح ذیل در نظر می‌گیریم:

تعداد ترمینال‌های فعال $N = 10$

$E(t) =$ ۴ ثانیه، (think time)

سطح ثابت چندبرنامگی $K = ۳$

تعداد پردازش شده: ۷۲۰

کل صفحاتی که خوانده شده: ۷۵۶۰۰

کل زمانی که Cpu صرف سیستم کرده: ثانیه ۸۲۷

کل زمانی که Cpu صرف اجرای برنامه کرده: ثانیه ۶۱۳

کل زمانی که Cpu در سیستم بررسی شده: ثانیه ۱۸۰۰

میانگین زمان سرویس I/O، ثانیه $E(O) = ۰/۰۳۸$

لازم به ذکر است هر فعل و انفعال (کنش) به جز اجرا توسط Cpu مدتی صرف کارهای سیستم، ارسال به ترمینال‌ها و ... می‌کند. برای یک مدت ۳۰ دقیقه‌ای ۶۱۳ ثانیه صرف برنامه و ۸۲۷ ثانیه صرف سیستم شده است. بنابراین کل زمان Cpu برای هر کنش عبارتست از:

$$mE(S) = \text{ثانیه } ۲ = \frac{(۶۱۳+۸۲۷)}{۷۲۰}$$

کل صفحات خوانده شده ۷۵۶۰۰ و کل کنش‌های پردازش شده ۷۲۰ بوده پس:

تعداد صفحات خوانده شده برای هر برنامه $\frac{۷۵۶۰۰}{۷۲۰} = ۱۰۵$

به عبارت دیگر $m = ۱۰۵$، پس میانگین زمان استفاده از Cpu بین دو عمل خواندن صفحه برابر است با:

$$E(S) = \frac{۲}{۱۰۵} = ۰/۰۱۹ \text{ ثانیه}$$

اکنون با توجه به مقادیر فوق داریم:

$$\rho = \frac{(۶۱۳+۸۲۷)}{۱۸۰۰} = ۰/۸ \text{ میزان بکارگیری Cpu، ثانیه}$$

$$\lambda_T = \frac{C\rho}{mE(S)} = \frac{۰/۸}{(۱۰۵\times۰/۰۱۹)} = ۰/۴ \text{ کنش در ثانیه}$$

$$W = \frac{(NmE(S))}{C\rho} - E(t) = \frac{۱۰}{\frac{۲}{۰/۸}} - ۴ = ۲۱ \text{ ثانیه}$$

در حالت توزیع نمائی داریم:

$$P_{\circ} = \left[۳! \sum_{n=۰}^{۳} \frac{۱}{(۳-n)!} \left(\frac{۰/۰۱۹}{۰/۰۳۸}\right)^n \right]^{-۱} = ۰/۲۱۰۵$$

$$\rho = ۱ - P_{\circ} = ۰/۷۸۹۵$$

$\lambda_t = \frac{\rho}{۲} = ۰/۳۹۴۸$ کنش در ثانیه

$W = (\frac{۲۰}{\rho}) - ۴ = ۲۱/۳۳$ ثانیه

برای زمان سرویس ثابت داریم:

$$\rho = \min \left\{ \frac{K}{۱ + \frac{E(o)}{E(s)}} , ۱ \right\} = \min \{ ۱,۱ \} = ۱$$

$\lambda_T = \frac{\rho}{۲} = ۰/۵$ کنش در ثانیه

$W = (\frac{۲۰}{\rho}) - ۴ = ۱۶$ ثانیه

۶ مدلسازی سیستم های کامپیوتری

۶-۱ مقدمه

در تحلیل و طراحی یک سیستم با مسائل مهمی روبرو می‌شویم که لازم است جهت پاسخگویی و حل آنها، درک کاملی از سیستم مذکور داشته باشیم. این درک و فهم از دو راه میسر می‌شود.

۱- درک مستقیم: از رفتار سیستم درک مستقیم داشته باشیم.

۲- درک تجربی: رفتار سیستم را توسط آزمایشهای گوناگون درشرایط مختلف مورد بررسی قرار دهیم و سپس آن را تحلیل کنیم.

روش درک مستقیم، سریع و قابل انعطاف بوده ولی قابلیت اطمینان زیادی ندارد. در درک تجربی سیستم، اطمینان زیادی حاصل شده ولی آزمایش و تجربه دشوار و غیر قابل انعطاف می‌باشد.

موضوع بحث ما روشی بنام مدلسازی، که بین دو روش فوق است. یک مدل در واقع چکیدهٔ یک سیستم می‌باشد که رفتارهای مهم سیستم را در بردارد. مدلسازی از روش تجربی آسان‌تر و از روش درک مستقیم مطمئن‌تر بوده و دارای انعطاف‌پذیری نیز است. یک سیستم ابتدا بوسیلهٔ یک مدل تعریف شده، سپس پارامتربندی و در نهایت تحت شرایط گوناگونی، رفتار آن سنجیده و ارزیابی می‌گردد.

درک مستقیم و تجربه به مدلسازی کمک زیادی می‌کند. مدلسازی که در این کتاب بحث شده کمی می‌باشد نه کیفی، بدین معنی که جهت ارزیابی سیستم مدل شده، با یکسری مقادیر (Measures) سروکار داریم. چون ارزیابی سیستم توسط مقادیر مذکور ملموس‌تر می‌باشد می‌توان گفت مدلسازی یک چارچوب برای جمع‌آوری، سازماندهی و سنجش اطلاعات در مورد یک سیستم کامپیوتری را، ارائه می‌نماید.

سیستم‌های کامپیوتری امروزه بیش از پیش پیشرفته و پیچیده شده‌اند و لازم است تکنیکها و ابزارهایی جهت بررسی رفتار این سیستم بیابیم. مسائلی از قبیل هزینه و کارایی، در تمام مدت عمر سیستم مطرح بوده و برای جوابگویی این مسائل و بررسی رفتار سیستم، به مدلسازی روی می‌آوریم. هزینه، کارایی و .... در مراحل ذیل از چرخهٔ عمر سیستم مطرح می‌شوند.

۱- مرحلهٔ طراحی واجرا: به عنوان مثال طراح یک سیستم کامپیوتری، جهت اتصال تعدادی ترمینال به یک کامپیوتر مرکزی (Mainframe) ، بایستی معماری‌ها و پروتکلهای مختلف را بررسی نماید. در شبکهٔ مثلاً طول بسته‌های پیغام را طوری در نظر بگیرد تا بهترین کارایی و کمترین هزینه را داشته باشد.

۲- مرحلهٔ استفاده و تکمیل سیستم: هنگام استفاده از یک سیستم کامپیوتری نیازهایی در زمان استفاده مطرح شده که طبعاً دارای هزینه و باید در مدل سیستم مد نظر قرار گیرد.

۳- مرحلهٔ تعیین اندازه: مثلاً در یک سیستم اطلاعات پزشکی که به صورت کامپیوتری طراحی شده باید توسط مدل، تخمینی از نرخهای ورودی داشته باشیم، تا بتوانیم زمان پاسخ مناسبی را برای سیستم تضمین نمائیم.

۴- مرحلهٔ تکمیل پیکره‌بندی و ظرفیت‌کاری (Work load) **سیستم** (work load)**:** یک شرکت بازرگانی که سفارشات و مبادلات کالا را با سیستم کامپیوتری انجام می‌دهد، با توجه به ظرفیت کارهای تجارتی خود، سیستم کامپیوتری موردنیاز را از نظر هزینه، منابع لازم و ... طراحی نموده و در نظر می‌گیرد. به طوری که با افزایش مبادلات و سفارشات کالا باید پیکره‌بندی سیستم نیز تکمیل‌تر گردد. این امر مستلزم آن است که تخمینی در مدل از ظرفیت‌کاری سیستم داشته باشیم.

خلاصه آنکه مدل، تصویر بسیار ساده و مختصری از حقیقت مسئله و به دو دلیل در اتخاذ تصمیم در مورد سیستم مؤثر می‌باشد. اول آنکه مسائل پیچیده را به صورتی آسان و قابل درک ارائه می‌نماید و دوم آنکه بااستفاده از مدل نتایج حاصل از اتخاذ تصمیم، قبل از اجرا قابل پیش بینی می‌باشد.

انواع مدل‌ها:

مدلها به انواع مختلفی طبقه‌بندی شده‌اند که معمول‌ترین آنها، شامل مدلهای فیزیکی، شماتیکی و ریاضی که به ترتیب مورد بررسی قرار می‌دهیم.

۱- مدل فیزیکی: مدلهای فیزیکی در مهندسی، موارد استفادهٔ زیادی دارد و طراحان قبل از ساختن یک هواپیما یا یک اتوبوس و یا هر وسیلهٔ جدید دیگری، ابتدا مدل فیزیکی آن را می‌سازند و پس از بررسی لازم، تصمیم به ساخت آن می‌گیرند.

برای مثال در کارخانجات هواپیماسازی، برای بررسی مسائل یک هواپیما، مدل کوچکی از آن تهیه کرده و در تونل باد آزمایش می‌کنند. البته این مدل کوچک بوده و تمامی صفات هواپیمای واقعی را ندارد، ولی در یک محیط قابل کنترل قادریم هرگونه آزمایش را با هزینهٔ کم انجام دهیم به عبارت دیگر می‌توان بدون تحمل هزینه‌های زیادی، تغییرات لازم را اعمال نمود و پس از حصول اطمینان از بی‌عیب بودن، اقدام به ساخت نمود.

۲- مدل شماتیکی: این مدل مختصرتر از مدل فیزیکی بوده و بوسیلهٔ چارت نمودار و یا تصویر نمایش

داده می‌شود.

۳- مدل ریاضی: این نوع مدل در مقایسهٔ با دو نوع قبلی، مختصرتر ولی درک آن مشکل‌تر می‌باشد. به عنوان مثال با مشاهدهٔ مدل فیزیکی یک هواپیما ساختمان آن رامی‌توان تشخیص داد، ولی درک مدلهای ریاضی که در آنها برای نمایش روابط، از فرمول و علائم استفاده می‌شود برای اشخاص عادی آسان نیست.

مدل ریاضی به صورت روابطی که ممکن است دارای متغیرهای قابل کنترل و متغیرهای غیرقابل کنترل باشد. البته برای محدودکردن معادلات یکسری فرضها نیز وارد مسئله می‌شود. در هر حال برای ساخت مدل توصیه می‌شود کلیهٔ اطلاعات موجود دقیقاً بررسی شود.

محاسن مدل:

۱- یک چهارچوب و پایه برای بررسی مسائل و مشکلات در اختیار محقق می‌گذارد و با استفاده از مدل می‌توان تنگناهای موجود را به آسانی تشخیص داد.

۲- از نظر هزینه به نفع سازمان می‌باشد.

۳- در اغلب اوقات نمی‌توان مدت زیادی صبر کرد و نتیجهٔ خاصی را بررسی نمود، پس با استفاده از مدل می‌توان عامل زمان را کوتاه و نتیجه را پیش‌بینی نمود.

۴- پس از اینکه مسئله به صورت مدل ریاضی نوشته شد، اولاً علائم قراردادی برای همهٔ زبانها یکسان و در واقع یک زبان مشترک بوجود می‌آید. دوم آنکه می‌توان عوامل را تغییر داده و روابط را بررسی کرد.

معایب مدل:

۱- در بعضی مواقع به دلیل پیچیدگی مدل، از آن استفاده نمی‌شود.

۲- مدل ممکن است آن قدر خلاصه شده باشد، که با حقیقت مسئله شباهتی نداشته باشد.

۳- در یک مدل نقاط حساس نسبت به دنیای حقیقی، مشابه و یک جور نیست.

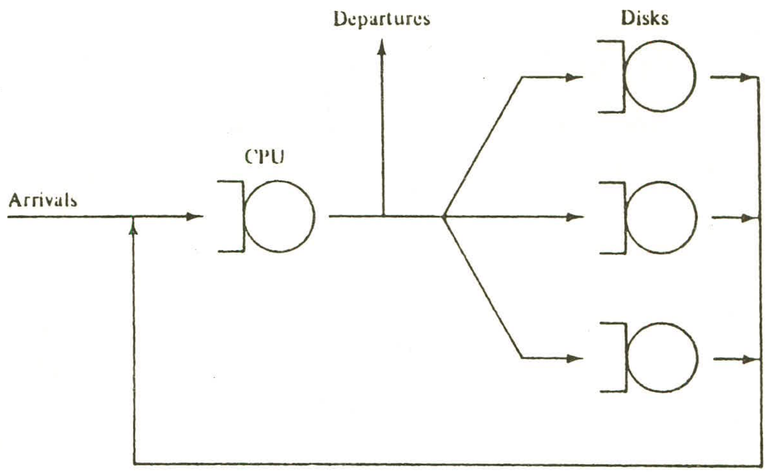
پس از آنکه مدلسازی را مطرح نمودیم حال به بررسی مدلسازی سیستم‌های کامپیوتری پرداخته و مدلهای مختلف را در فصلهای بعدی مورد بررسی قرار می‌دهیم.

۶-۲ مدل شبکهٔ صف

شبکهٔ صف که در فصل قبل بررسی نمودیم یک روش ویژه جهت مدلسازی سیستم‌های کامپیوتری است. در حقیقت یک شبکهٔ صف، مجموعه‌ای از مراکز سرویس که منابع سیستم را ارائه می‌نماید و تعدادی مشتری نیز که معرف کاربران می‌باشند تعریف می‌گردد. مراکز سرویس دونوع می‌باشد .

۱- منفرد ۲- چندگانه

همانطور که قبلاً بیان نمودیم نرخ ورود مشتری و مدت زمان سرویس پارامترهای مهمی در تجزیه و تحلیل سیستمهای صفی می‌باشند. در راستای اثربخشی سیستم، باید مدل (شبکهٔ صفی) را ارزیابی کرده و با حل معادلات ساده، مقادیر کارآئی نظیر طول صف، نرخ خروجی، زمان سپری شده در سیستم و ... را بیابیم. در مرکز سرویس چندگانه، به پارامترهای بیشتری نیاز داریم. مثلاً در شکل ۱-۶ یک Cpu و سه



شکل ۶-۱ مدل شبکهٔ صف از یک سیستم کامپیوتری با چندین سرویس دهنده

عدد دیسک منابع سیستم (سرویس دهنده‌ها) را تشکیل می‌دهند. شکل مذکور به صورت شبکهٔ صف می‌باشد که یک سیستم کامپیوتری را مدل نموده است.

پارامترهای مورد نیاز در اینجا عبارتند از:

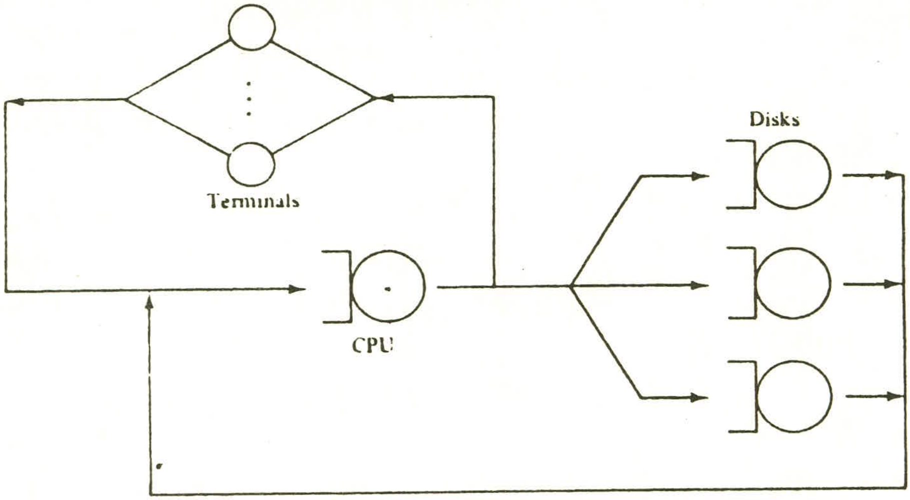
۱- نرخ ورود مشتری
۲- مقدار سرویس مورد نیاز
۳- مدت سرویس دهی برای هر مرکز به طور مجزا

لازم به ذکر است اکثر مقادیر کارآئی در مدل صف، به صورت میانگین می‌باشند، که باید درک صحیحی از این مفهوم شود. همانطور که گفتیم مدلسازی سه مرحله تعریف، پارامتربندی و ارزیابی دارد، که در مورد مدل شبکهٔ صف هریک از اینها را بررسی می‌کنیم.

۱- تعریف:

ارتباط مستقیم و متناظر بین یک سیستم کامپیوتری و مدل شبکهٔ صف توسط تعریف، برقرار می‌شود. مثلاً مرکز سرویس مدل شبکهٔ صف همان منابع سخت افزاری سیستم و صف نیز پردازشهای کاربران می‌باشد. همانطور که ذکر شد تعیین نرخ ورود مشتری مهم می‌باشد. در سیستم‌های کامپیوتری معمولاً، تعیین نرخ ورود مشتری، شدت ظرفیت‌کاری را نیز تعیین می‌کند. روش دیگر تعیین کل تعداد مشتری در سیستم، که شدت ظرفیت‌کاری را نیز مشخص می‌کند. کل تعداد مشتری در زمان معینی را در متوسط زمانی که هر مشتری بین دو تراکنش (Transaction) صرف فکرکردن می‌کند، ضرب نمائیم، نرخ ورود هم کنش در زمان مذکور بدست می‌آید (ظرفیت‌کاری پردازشهای) بنابراین اگر شکل ۶-۱ را براساس شدت ظرفیت‌کاری، اصلاح نمائیم به شکل ۶-۲ می‌رسیم.

در سیستم‌های کامپیوتری واقعی، جهت تعریف مدل لازم است ظرفیت‌کاری‌های مختلفی را درنظر بگیریم. از طرفی ممکن است چندین نوع مشتری در دسته‌های مختلف، در مدل موجود باشند و هریک



شکل ۶-۲ مدل شبکه صف با شدت ظرفیت کاری

ظرفیت‌کاری جداگانه‌ای داشته باشند که باید توسط راههای مذکور تعیین شوند. به عنوان مثال چهارنوع ظرفیت‌کاری داریم:

۱- پردازشهای تراکنشی.
۲- پردازشهایی که ماهیت دسته‌ای دارند.
۳- پردازشهایی به صورت پرسش و پاسخ در پایگاه داده.
۴- پردازشهایی جهت توسعهٔ برنامه به صورت فعل و انفعال.

پس باید مراکزسرویس مناسب جهت هردستهٔ پردازش تعیین کنیم.درمرحلهٔ پارامتربندی نیز برای هر گروه ظرفیت‌کاری مشخصی ارائه دهیم.مثلاً برای پردازشهای نوع (۱)نرخ ورود ۱۰ تقاضا در ثانیه، برای نوع (۲) سطح چندبرنامگی ثابت ۲، برای نوع (۳) تعداد ۲۵ کاربر بامتوسط زمان ۲ دقیقه تفکر بین دو تراکنش و برای پردازشهای نوع (۴) تعداد ۱۰ کاربر، هریک با متوسط ۱۵ ثانیه تفکر بین دو تراکنش، در نظر می‌گیریم. برای هر گروه مدت سرویس‌دهی در هر مرکز سرویس را تعیین می‌کنیم. بنابراین در کارآیی مدل، مقادیرارزیابی درکل (مثلاً کل بهرهٔ Cpu)و در هرگروه(مثلاً بهرهٔ Cpuبرای پردازش batch)را می‌توان محاسبه‌نمود.

۲- پارامتربندی:

در این مرحله به مشخصه‌های مدل که تعریف شده، پارامترهایی نسبت می‌دهیم تا بتوانیم رفتار مدل را با روابطی بیان نمائیم. مثلاً برای محاسبهٔ زمان سرویس دهی Cpu برای هر مشتری، دو پارامتر تعیین می‌کنیم. زمان مشغول بودن Cpu و تعداد تقاضاها جهت پردازش (دسته‌ای، هم کنش و ...) باشد. بنابراین متوسط زمان Cpu برای هر تقاضا عبارتست از :

$$\frac{\text{زمان مشغول بودن Cpu}}{\text{تعداد تقاضاها}}$$

یکی از توانائی‌های عمدهٔ مدل شبکهٔ صف، قابلیت اصلاح و تغییر رابطهٔ بین پارامترها می‌باشد. به عبارت دیگر با تغییر شرایط، روابط تغییر یافته و شرایط به سادگی بررسی می‌شوند.

به عنوان مثال در مدل شکل ۲-۶، نرخ ورودی برابر هر ۵ ثانیه یک پردازش و زمان سرویس Cpu را ۳ ثانیه و زمان سرویس دیسک‌ها را به ترتیب ۱، ۲و۳ ثانیه، در نظر می‌گیریم اکنون :

- اگر پردازشهای I/O را در بین دیسک‌ها به طور متعادل پخش کنیم چه می‌شود؟

$$\frac{۱+۲+۴}{۳} = ۲/۳۳$$

پس از ارزیابی مجدد، زمان پاسخ از ۳۲/۱ به ۲۰/۶ کاهش می‌یابد.

- اگر ظرفیت‌کاری ۲۰٪ کاهش یابد چه می‌شود؟

تعداد تقاضا در ثانیه که همان ظرفیت کاری را در ۰/۲ ضرب می‌کنیم:

$$۰/۲ \times ۱/۲ = ۰/۲۴$$

پس از ارزیابی مجدد مدل، زمان پاسخ از ۲۰/۶ به ۲۶/۶ افزایش می‌یابد

- اگر Cpu را ۳۰٪ سریعتر کنیم چه می‌شود؟

داریم $۲/۳۱ = \frac{۳}{۱/۳}$، که پس از ارزیابی مجدد زمان پاسخ از ۲۶/۶ به ۲۱ می‌رسد.

پس از اعمال تغییر در مدل، جهت تحلیل توجه زیادی لازم است. زیرا دقت مقادیر کارآئی که از مدل حاصل می‌شوند به مقادیر ظرفیت‌کاری و مدت زمان سرویس دهی بستگی دارد. همواره پیش بینی تأثیرات ناشی از تغییر پیکره‌بندی و ظرفیت‌کاری روی این پارامترها، کار ساده‌ای نمی‌باشد. به عنوان مثال در مدل مذکور فرض می‌کنیم دیسکها از نظر فیزیکی یکسان می‌باشند. هنگامی که اعمال I/O به طور متعادل روی دیسکها توزیع می‌شوند، زمان سرویس هر دیسک یک مقدار میانگین خواهدشد. بنابراین تأثیرات اولیه‌ای در پارامترهای مدل ایجاد می‌شود و تأثیرات ثانویهٔ بعدی نیز ممکن است رخ دهد. یعنی کل میزان جابجایی بازوی دیسک کاهش یافته و در نتیجه نیاز کاربران به سرویس دیسک در مجموع کاهش یابد.

بنابراین اگر تأثیرات ثانویه را پیش بینی کنیم می‌توان در هنگام ارزیابی مدل، مقادیر دقیقی جهت ارزیابی بدست آوریم وگرنه نتایج بدبینانه‌ای در انتظار خواهد بود. لازم به ذکر است تأثیرات اولیه حاصل از تغییرات مدل، هنگامی که در ارزیابی مدنظر قرار می‌گیرند، کمک می‌کند تا پیش‌بینی بهتر صورت پذیرد.

مدلهایی که چندین نوع مشتری دارند شرایط مختلفی را می‌پذیرند و تحلیل و پارامتربندی آنها مشکل‌تر می‌باشد. مدلهایی که یک کلاسی (یک نوع مشتری) هستند مطمئن‌تر و سریع‌تر می‌باشند. هنگامی که چندین نوع مشتری در مدل داریم، بااستفاده از ابزار کامپیوتری موجود، مشکل است به طور دقیق میزان مصرف منابع (سرویس دهی) توسط پردازشهای گوناگون، با ظرفیت‌کاری‌های مختلف، را تعیین کنیم، خصوصاً آنکه سربار (Overhead) سیستم و فعالیتهای جزئی I/O را به طور دقیق نمی‌توان ارزیابی نمود.

۳- ارزیابی:

به منظور ارزیابی مدلهای شبکه صف دو راه موجود است. اول آنکه حدود بالا و پایین مقادیر کارایی را محاسبه نمائیم. مثلاً حد بالا و پایین زمان پاسخ را برای یک مجموعهٔ خاصی از پارامترها تعیین کنیم (به طور مثال شدت ظرفیت‌کاری ، زمان سرویس‌دهی و ....). خاصیت این روش سادگی محاسبات مربوط به آن که حتی به طور دستی نیز انجام می‌شود و در درک سیستم نیز به ما کمک می‌کند.

راه دوم، محاسبهٔ دقیق مقادیر کارآئی (Per formance) می‌باشد. الگوریتم‌هایی که بدین منظور به کار می‌روند پیچیدگی زیادی دارند و مجبوریم از کامپیوتر کمک بگیریم. الگوریتم‌ها در این خصوص، هنگام

اجرا، تعداد مراکز سرویس دهنده و دسته‌های مشتریان (کلاسهای مشتریان) را تعیین می‌کند. البته زمان کارآئی مدل، به تعداد مشتریان در هر کلاس بستگی دارد. مثلاً یک مدل شبکهٔ صفی که دارای ۱۰۰ مرکز سرویس‌دهی و ۱۰ کلاس مشتری می‌باشد، در عرض چندثانیه توسط کامپیوتر ارزیابی می‌گردد. یک نرم‌افزار مدلسازی شبکهٔ صف در سطوح مختلف خود یک مدل شبکهٔ صفی را ارزیابی کرده و با روتین‌های مختلف، رابطهٔ بین یک زبان کامپیوتری و زبان مدل شبکهٔ صف رابرای کلیهٔ قسمتهای مدل، برقرار می‌کند. و در نهایت مقادیر پارامترهای مدل را از داده‌های سیستم استخراج می‌نماید. (براساس یک زبان سطح بالا زبان مدل شبکهٔ صف را بیان می‌نماید. مثلاً QNAP، یک نرم افزار مدلسازی شبکهٔ صف با زبان فرترن می‌باشد).

۶-۳ مدل شبکهٔ صف ابزار مناسب

همانطور که گفتیم، مدلها عموماً و مدل شبکهٔ صف خصوصاً، ابزار مهمی در تحلیل و طراحی سیستم‌های کامپیوتری می‌باشند. در بیشتر موارد مدل شبکهٔ صف یک تعادل بین دقت و کارایی در سیستم، ایجاد می‌کند. از نظر تجربه نشان داده شده که مدل شبکهٔ صف برای بهرهٔ و توان خروجی دقت قابل قبولی را ارائه می‌دهد. البته این دقت براساس نیاز کاربردهای مختلف می‌باشد. دقت در تعیین اجزاء دیگری که در پردازش تحلیلی سیستم کامپیوتری مؤثر می‌باشند (نظیر مشخصات ظرفیت‌کاری) بر دقت مذکور اثر می‌گذارد. از نظر کارایی و هزینه، در تمامی مراحل مدلسازی شبکهٔ صف، کارایی خوب و هزینهٔ کمی متحمل می‌شویم. تعریف مدل، با ارتباط متناظر بین خواص مدل شبکهٔ صف و سیستم کامپیوتری صورت می‌پذیرد. پارامتربندی نیز با نسبت دادن اعداد به پارامترها، به سادگی انجام شده و در نهایت، ارزیابی مدل با الگوریتم‌های کامپیوتری صورت می‌گیرد.

خلاصه آنکه مدل با دقت بالا ساخته می‌شود. البته برای دستیابی به دقت بیشتر، افزایش هزینه مهم و در کاربردهای مختلف باید از نظر کارایی و هزینه مدل، یک تعادل برقرار شود.

در مباحثی که انجام دادیم مجموعه‌ای از مدل‌های شبکهٔ صف را بیان نمودیم. شبکه‌های مورد بحث تفکیک‌پذیر بوده و برای نمایش سیستم‌های کامپیوتری مناسب می‌باشند.

سیستم‌هایی که صدها منبع و دهها نوع ظرفیت کاری مختلف با مشتریان گوناگون دارند، جهت تحلیل در شبکه‌های تفکیک‌پذیری که تاکنون بیان نمودیم، بایستی یکسری فرضهایی را بپذیرند و محدودیتهایی روی آنها اعمال گردد. عدم توجه به آنها نتایج نامناسب و اندازه گیری غلطی را در پی خواهند داشت.

در شبکه‌های عمومی صف مسائل به طور تحلیلی ارزیابی شده و خیلی از فرضها را در نظر نمی‌گیریم ولی باید توجه داشت هرچه اندازهٔ شبکهٔ بزرگتر شود الگوریتم‌های تحلیلی فضا و زمان زیادی را می‌طلبند و در نتیجه کارایی آنها کم خواهد شد.

۶-۴ مدل شبکهٔ صف و تئوری صف

مدل شبکهٔ صف می‌تواندبه صورت یک زیرمجموعهٔ کوچکی از تکنیکهای تئوری صف در نظر گرفته شود که این تکنیکها خصوصاً جهت مدلسازی سیستم‌های کامپیوتری، استفاده می‌شود. تکنیکهای پیشرفتهٔ ریاضی جهت تحلیل مدلهای ساخته شده توسط تئوری صف، ابداع شده که مقادیر کارآئی با تمام جزئیات را به صورت نسبی، توسط مدل بدست می‌آورند (مثلاً به صورت توزیعی و گسسته). در حقیقت

زیرمجموعه‌ای از تئوری صف الگوریتم‌هایی را ارائه می‌دهد که با برقرار ساختن ارتباط بین سیستم‌های کامپیوتری و زبان تئوری صف، براساس شناخت اصطلاحات سیستم‌های مذکور، با فهم از تئوری صف، به نحو موفقی آن را بکار می‌برند.

۶-5 مدل شبکهٔ صف و شبیه‌سازی

قبل از هر مطلبی شبیه‌سازی عبارتست از تکنیکی که طرز عمل سیستم را در محیط خود در یک دورهٔ زمانی مجسم سازد.

یک مدل شبیه‌سازی دارای چهارجزء اساسی می‌باشد:

۱- عوامل: این عوامل در سیستم می‌تواند دائمی یا موقت باشد.

۲- متغیرها: مشخصه‌هایی هستند که عوامل با آنها اندازه گیری می‌شود.

۳- پارامترها: ضرائب معادلاتی که روابط بین متغیرهای مختلف را تعیین می‌کند.

۴- روابط عملی: روابطی که وابستگی‌های بین عوامل و متغیرهای مختلف را در مدل شبیه‌سازی تعریف می‌کند.

در مورد مدلهای شبیه‌سازی باید بگوئیم توانائی عمدهٔ آن، انعطاف‌پذیری می‌باشد. رفتار یک سیستم با شبیه‌سازی، در هر سطح دلخواهی می‌تواند بررسی شود و در نهایت جهت ارزیابی شبکه‌های صف مورد استفاده قرار گیرد. شبیه‌سازی یک سیستم نوعی الگوسازی از آن سیستم محسوب می‌شود.

اشکال عمدهٔ مدل شبیه‌سازی، هزینهٔ نسبتاً زیاد آن می‌باشد. در مرحلهٔ شبیه‌سازی با مسائلی چون نوشتن برنامه و رفع خطای برنامه‌های پیچیدهٔ کامپیوتر درگیر خواهیم بود، که خود هزینهٔ زیادی دربردارد. در موارد خاص، برای مدلسازی سیستم‌های کامپیوتری، از بسته‌های نرم‌افزاری استفاده می‌شود، که براساس توصیف مدل جهت شبیه‌سازی ساخته‌شده‌اند.

مدل شبیه‌سازی در مرحلهٔ پارامتربندی نیز هزینهٔ زیادی دارد. زیرا یک مدل با تمامی جزئیات، مستلزم تعداد زیادی پارامتر می‌باشد. در نهایت چون اجرای یک شبیه‌سازی، طبعاً منابع محاسباتی زیادی نیاز دارد، جهت ارزیابی نیز هزینهٔ زیادی مصرف می‌کنند، خصوصاً اگر قابلیت اطمینان زیادی بخواهیم. مدل شبکهٔ صف سطح مناسب و دقیقی برای کاربردهای مختلف طراحی و تحلیل سیستم‌های کامپیوتری، فراهم می‌نماید. بعضی موارد هزینهٔ ارزیابی یک مدل شبکهٔ سادهٔ صف، با استفاده از شبیه‌سازی به مراتب، بیشتر از هزینهٔ ارزیابی همان مدل توسط روشهای تحلیلی می‌باشد.

۷

تحلیل سیستم های کامپیوتری

۷-۱ مقدمه

در این فصل به ساده‌ترین روش ، جهت تحلیل سیستم‌های کامپیوتری با استفاده از مدل شبکهٔ صف می‌پردازیم. این روش، "تحلیل حدود" می‌باشد. با محاسبات مختصری امکان تعیین حد بالا و پایین نرخ خروجی سیستم و زمان پاسخ گویی سیستم، موجود می‌باشد. نرخ خروجی و زمان پاسخ سیستم، تابعی از شدت ظرفیت کاری سیستم می‌باشند (تعداد یا نرخ ورود مشتری).

- تکنیکهایی جهت بررسی ارزیابی بیان می‌کنیم که عبارتند از:

۱- حدود مجانبی

۲- حدود سیستم‌های متعادل

حدهای مجانبی بسادگی محاسبه و نسبت به حدهای سیستم متعادل برای گروه وسیعی از سیستم‌ها به کاربرده می‌شوند. در حالیکه حدهای سیستم متعادل دقیق‌تر و اطلاعات مفیدتری نسبت به حدهای مجانبی فراهم می‌نمایند. ویژگیهایی که تکنیکهای حدگیری را مفید و مناسب می‌سازند عبارتند از:

۱- فاکتورهای اولیه مؤثر در ارزیابی سیستم‌های کامپیوتری را، مقداردهی دهند. خصوصاً آنکه توسط این تکنیکها تأثیر (بحرانی) گلوگاه سیستم روشن شده و به صورت کمی تعیین می‌گردد.

۲- حدود به طور تحلیلی سریعاً محاسبه می‌شوند. بنابراین تحلیل حدها، به عنوان یک برش اولیه تکنیکهای مدلسازی مناسب می‌باشد. این تحلیل در مراحل اولیه بررسی یک سیستم، جهت حذف آلترناتیوهای نامناسب می‌تواند استفاده شود.

۳- غالباً توسط یک تحلیل حدی ، چندین مورد با هم می‌توانند بررسی شوند، که این تحلیل اطلاعات مفیدی در مورد همهٔ آنها فراهم می‌نماید.

تکنیکهای تحلیلی پیشرفته، نیاز به محاسبات بیشتر دارند. تکنیکهای حدی مفیدترین روش برای بررسی اندازهٔ سیستم هستند. هنگامی که طراحی در سطح وسیع و اندازهٔ سیستم مطرح می‌شود، از مشخصه‌های سیستم یک تخمین اولیه می‌زنیم. این تخمین اولیه موجب بی‌دقتی در درک سیستم می‌شود. بنابراین بررسی سریع حدود، جهت هدایت به تخمین‌های مقادیر ارزیابی، مناسب‌تر از تحلیل تفصیلی می‌باشد. بررسی اندازهٔ سیستم نوعاً، با پیکره‌بندی‌های مختلف مورد نظر، همراه می‌باشد. غالباً در یک سیستم منبع (مثلاً Cpu) نقش اصلی را دارا و مابقی اجزاء سیستم با توجه به توان منبع مذکور، پیکره‌بندی می‌شوند.

تحلیل حدی، دسته‌هایی از پیکره‌بندی‌های مختلف را به صورت آلترناتیو (قابل تغییر) که دارای یک منبع بحرانی می‌باشند، در نظر می‌گیرد. ولی در هر پیکره‌بندی از مرکز سرویس (سرویس‌دهنده) تقاضای سرویس متفاوتی موجود می‌باشد.

تکنیکهای حدی همچنین جهت تخمین بهرهٔ (Gain) ارزیابی بالقوهٔ آلترناتیوهای مختلف، که در سیستم می‌تواند رخ دهد، استفاده می‌شوند. در بخشهای بعدی نشان می‌دهیم چگونه گرافهای حدی می‌توانند حدود کاهش تقاضای سرویس مورد نیاز را در گلوگاه شبکه، تعیین نمایند. به طوریکه اهداف تعیین شدهٔ ارزیابی مدل، برآورد شوند (تقاضای سرویس در یک مرکز می‌تواند باانتقال بعضی‌ازکارهااز مرکزبه مراکزدیگر و یاباجایگزینی سرویس‌دهندهٔ سریعتری کاهش یابد).

تحلیل حدی بیشتر در مورد شبکه‌های تک کلاسی (یک نوع مشتری) مطرح می‌شوند. به عبارت دیگر تحلیل حدی، بیشتر برای بررسی ظرفیت گلوگاه مفید است. مدلهایی که در ادامهٔ بحث در نظر می‌گیریم، توسط پارامترهای ذیل توصیف می‌شوند.

K: تعداد مراکز سرویس دهنده.

D: مجموع تقاضای سرویس در سرویس دهنده‌ها.

Z: متوسط زمان تفکر بین دو تراکنش (اگر برنامه از نوع ترمینال باشد).

Dmax: بیشترین تقاضای سرویس در هر مرکز.

در مدلهایی که ظرفیت‌کاری آنها از نوع تراکنش می‌باشد، حدود نرخ خروجی، نشاندهندهٔ نرخ ورود برنامه پردازش شده است.

حدود زمان پاسخ‌دهی، به عبارت دیگر کمترین و بیشترین زمان پاسخ دهی ممکن که توسط یک مشتری تجربه می‌شود، به صورت تابعی از نرخ ورود سیستم می‌باشد. در مدلهایی با ظرفیت‌کاری از نوع دسته‌ای یا ترمینال، حدود معرف می‌نیمم و ماکزیمم زمان پاسخ‌دهی و نرخ خروجی ممکن، سیستم می‌باشند. بطوریکه تابعی از تعداد مشتریان سیستم ملحوظ می‌گردد.

حد بالای نرخ خروجی و حد پایین زمان پاسخ‌دهی را حدود خوش بینانه گوییم. زیرا بهترین کارایی ممکن را نشان می‌دهند. همچنین به حد پایین نرخ خروجی و حد بالای زمان پاسخ دهی، حدود بدبینانه می‌گوییم، زیرا بدترین کارایی ممکن را ارائه می‌دهند.

در بخش‌های بعدی فقط به حدود نرخ خروجی و زمان پاسخ سیستم اشاره می‌کنیم. ولی قواعد اساسی (نظیر قضیهٔ لیتل، فرضهای تعادل جریان و ....) می‌توانند جهت تبدیل این حدود به حدهای دیگر مقادیر ارزیابی، مورد استفاده قرار بگیرند (نظیر حدود نرخ خروجی هر مرکز سرویس، حدود بکارگیری (ρ)، مرکز سرویس و ....).

۷-۲ حدود مجانبی

تحلیل حدود مجانبی، حدهای خوش بینانه و بدبینانه‌ای را برای نرخ خروجی و زمان پاسخ، شبکه‌های صفی تک کلاسی فراهم می‌کند. این حدود با درنظرگرفتن شرایط فوق العادهٔ بار سبک و بار سنگین (به طور مجانب) بدست می‌آیند. اعتبار این حدود فقط به یک فرض بستگی دارد. آنکه میزان تقاضای سرویس یک مشتری در یک سرویس دهنده، به تعداد مشتری‌های جاری در سیستم و یا به محل سرویس دهندهٔ آنها، بستگی نداشته باشد.

اطلاعات فراهم شده توسط حدهای مجانبی، به نوع ظرفیت‌کاری سیستم بستگی دارد. باز (مثلاً نوع تراکنشی) یا بسته (مثلاً نوع دسته‌ای و ترمینال) را به ترتیب بررسی می‌کنیم.

۱- ظرفیت کاری مدل باز:

در مورد حدهای نرخ خروجی، ماکزیمم نرخ ورودی ممکن از مشتری‌ها، که سیستم می‌تواند آن را پردازش کند مشخص می‌گردد. اگر نرخ ورودی از این حد تجاوز کند، برنامه‌های ورودی مرتباً انباشته شده و تعداد برنامه‌های اجرا نشده رشد می‌کند. بنابراین در اجراهای طولانی، یک برنامه ورودی انتظار نامحدودی از نظر زمانی داشته (چون ممکن است هر تعداد برنامه دلخواه، از قبل در صف موجود باشند) و در این حالت سیستم اشباع شده است. پس حد نرخ خروجی، نرخی از ورودی است که پردازش ممکن را از حالت اشباع تفکیک می‌کند. برای تعیین حد نرخ خروجی از قانون بکارگیری استفاده می‌کنیم [۱۲].

در هر مرکز K $\quad U_k = X_k.S_k$

اگر نرخ ورود سیستم را λ در نظر بگیریم بنابراین: $\quad X_k = \lambda.V_k$

رابطهٔ فوق به صورت $U_k = \lambda.D_k$ تعریف می‌گردد که D_k میزان تقاضای سرویس در مرکز K می‌باشد. در مورد حد نرخ خروجی به اندازه‌ای که مراکز ظرفیت استفاده نشده دارند (یعنی $U_i<1$)، می‌توان در نرخ ورودی افزایش داشت. بنابراین هنگامی که همهٔ مراکز اشباع شوند ($U_i = ۱$) کل سیستم اشباع می‌شود. زیرا هیچ افزایش در نرخ ورودی مشتری‌ها رخ نمی‌دهد. پس حد نرخ خروجی وقتی که همهٔ مراکز اشباع شوند، کوچکترین نرخ ورودی (λ_{sat}) خواهد بود.

همانطور که می‌دانید مرکزی که با کمترین نرخ ورودی اشباع شود گلوگاه است (مرکزی با بیشترین میزان تقاضای سرویس). برای گلوگاه داریم [۱۲]:

$$U_{max}\lambda = \lambda.\ D_{max} \leq ۱$$

بنابراین:
$$\lambda_{sat} = \frac{۱}{D_{max}}$$

در نتیجه برای نرخهای ورودی بزرگتر یا مساوی $\frac{۱}{D_{max}}$ سیستم اشباع می‌شود. در حالی که سیستم قابلیت

پردازش نرخهای ورودی کمتر از $\frac{1}{D_{max}}$ را داراست.

حدهای مجانبی زمان پاسخ، کمترین و بیشترین زمان پاسخ ممکن برای مشتری را نشان می‌دهند. اگر λ نرخ ورودی سیستم در حالت $\lambda_{sat} < \lambda$ باشد (سیستم ناپایدار)، بررسی خود را به موردی محدود می‌کنیم که نرخ ورودی کمتر از حد نرخ خروجی است. در اینجا دو وضعیت فوق‌العاده پیش می‌آید.

۱- در بهترین وضعیت ممکن، هیچ مشتری در صورت وجود مشتری دیگر، وارد نمی‌شود و تأخیری در صف نداریم. بنابراین زمان پاسخ سیستم برای هر مشتری، کل تقاضای سرویس آن می‌باشد (D).

۲- در بدترین وضعیت ممکن، n مشتری با یکدیگر وارد سیستم شده که این عمل در $\frac{n}{\lambda}$ واحد زمان صورت می‌گیرد (نرخ ورودی سیستم $\lambda = \frac{n}{\frac{n}{\lambda}}$ می‌باشد). مشتری‌هایی که آخر وارد می‌شوند بدلیل وجود مشتری‌های جلوی زمان پاسخ زیادی دارند.

اگر اندازهٔ دستهٔ ورودی n باشد، مشتری‌ها مرتباً در حال افزایش می‌باشند. پس هر حد بدبینانهٔ مفروضی در مورد زمان پاسخ سیستم (با نرخ ورودی λ) می‌تواند طوری انتخاب شود که از حد بالا نیز تجاوز نموده و اندازهٔ دستهٔ ورودی نیز به حد کافی بزرگ باشد. بنابراین در مورد زمان پاسخ هیچ حد بدبینانه‌ای وجود ندارد (صرفنظر از آنکه مقدار نرخ ورودی λ چقدر ممکن است کوچک باشد). خوشبختانه حدهای زمان پاسخ و نرخ خروجی در مورد ظرفیت کاری از نوع بسته اطلاعات بیشتری را فراهم می‌آورند.

۲- ظرفیت‌کاری دسته‌ای و ترمینال:

شکل‌های ۱-۷.a و ۱-۷.b، حدهای مجانبی مربوط به نرخ خروجی و زمان پاسخ را در مورد ظرفیت‌کاری دسته‌ای و ترمینال نشان می‌دهد. مقادیر دقیق زمانهای پاسخ و نرخ خروجی باید در ناحیهٔ هاشور خوردهٔ شکلهای مذکور قرار بگیرند. منحنی‌ها موقعیت مقادیر مذکور را نشان می‌دهند.

برای بدست آوردن حدود نشان داده شده، ابتدا حدهای نرخ خروجی را در نظر می‌گیریم. سپس با استفاده از قضیهٔ little، آنها را به حدهای متناظر زمان پاسخ تبدیل می‌کنیم. تحلیل‌های مذکور برای ظرفیت‌کاری از نوع ترمینال می‌باشد. با صفر کردن Z، نتایج را برای کارهای دسته‌ای می‌توان بدست آورد.

هنگامی که بار پردازش سنگین (مشتری زیاد) باشد، در صورتی که مشتریان سیستم زیاد شوند N، بکارگیری تمام مراکز سرویس افزایش می‌یابد. اما واضح است که هیچ بهره‌ای از ۱ بیشتر نمی‌شود. برای هر مرکز سرویس K، با توجه به قانون بکارگیری (u) داریم:

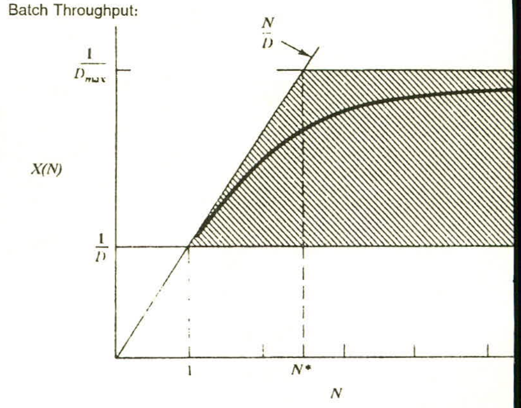
$$U_k\ (N) = X(N)\ .\ D_k \leq 1$$

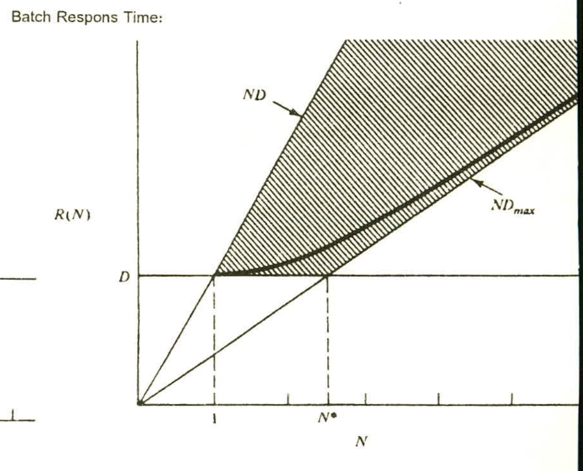
چون مرکز گلوگاه زودتر از همه اشباع می‌شود، بنابراین بیشتر از مراکز دیگر، نرخ خروجی سیستم را محدود می‌کند. پس نتیجه می‌گیریم :

$$X(N) \leq \frac{1}{D_{Max}}$$

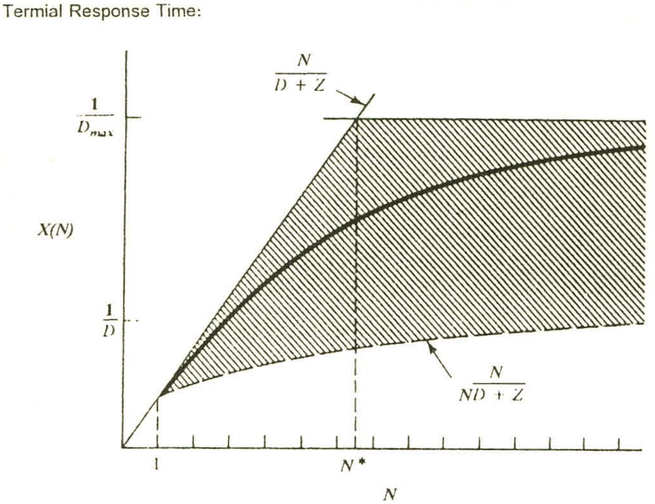
رابطه فوق بدیهی است. زیرا اگر هر مشتری در گلوگاه به طور متوسط D_{Max} واحد زمان، به سرویس نیاز داشته باشد، آنگاه در اجراهای طولانی، قطعاً مشتریان زودتر از هر D_{Max} واحد زمان، اجرا نمیشوند.

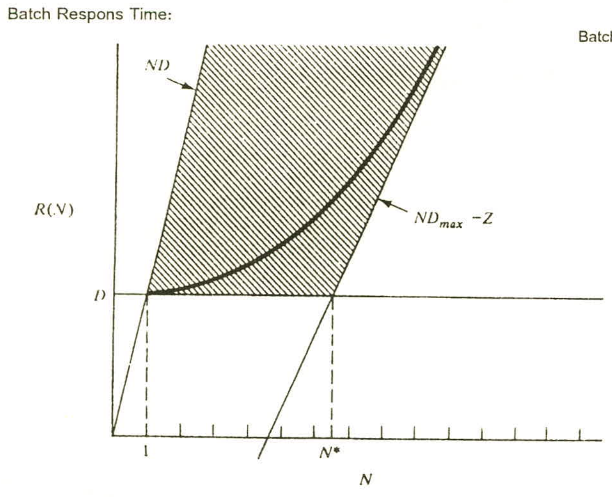
در وضعیت بار پردازش سبک (تعداد مشتریان کم) یک مشتری وارد سیستم شده و نرخ خروجی





شکل الف ۷-۱





شکل ب ۷-۱

$\frac{1}{(D + Z)}$ می‌باشد. زیرا هر فعل و انفعال شامل یک دورهٔ سرویس (بامیانگین $D = \sum_{k=1}^{K} D_k$) و یک زمان تفکر (با میانگین Z) می‌باشد. برای مشتریان بیشتری که به سیستم اضافه می‌شوند دو وضعیت حدی وجود دارد:

۱- کمترین نرخ خروجی وقتی رخ می‌دهد که هر مشتری اضافه شده پشت سر همهٔ مشتریانی که قبلاً وارد شده‌اند در صف منتظر بماند. در این حالت اگر N مشتری در سیستم باشد، $(N-1)D$ واحد زمانی در صف سپری می‌شود. D واحد زمانی در سرویس سپری شده و Z واحد زمانی صرف تفکّر مـی‌گردد. بنابراین نرخ خروجی هر مشتری $\frac{1}{(ND+Z)}$ و نرخ خروجی سیستم $\frac{N}{(ND+Z)}$ خواهد شد.

۲- بیشترین نرخ خروجی ممکن، وقتی رخ می‌دهد که هر مشتری اضافه شده، هیـچ تأخیری در صف نداشته باشد. D واحدزمانی در سرویس و Z واحدزمانی صرف تفکر می‌گردد. بنابراین نرخ خروجی

هر مشتری $\frac{1}{(D+Z)}$ و نرخ خروجی سیستم $\frac{N}{(D+Z)}$ می‌باشد.
مباحث فوق به صورت حدهای مجانبی روی نرخ خروجی سیستم خلاصه می‌شود [۱۲].

$$\frac{N}{N D + Z} \leq X (N) \leq \mathrm{Min}(\frac{1}{D_{Max}} \text{و} \frac{N}{D+Z})$$

دقت شود که حد خوش‌بینانه دو بخش دارد یکی در شرایط بار سنگین پردازشی و دیگری در شرایط بار سبک، بکار می‌رود.

همانطور که در شکل ۷-۱ دیدیم، تعداد خاصی مشتری (N^*) وجود دارد که برای تمامی Nهای کمتر از آن، شرایط بار سبک بکار رفته و برای Nهای بیشتر از آن، شرایط بار سنگین استفاده‌می‌شود. نقطه‌تلاقی مذکور، در محل برابری دوحد رخ می‌دهد یعنی:

$$N^* = \frac{D+Z}{D_{max}}$$

می‌توان حدود زمان پاسخ R(N) را توسط تبدیل حدود نرخ خروجی با استفاده از قضیهٔ لیتل بدست آورد. با استفاده از معادلهٔ قبلی داریم [۱۲]:

$$\frac{N}{ND + Z} \leq \frac{N}{R(N) + Z} \leq \min (\frac{1}{D_{max}} , \frac{N}{D + Z})$$

با معکوس کردن هر جزء این رابطه، حدهای R(N) به راحتی بدست می‌آیند.

$$\max (D_{max} , \frac{D + Z}{N}) \leq \frac{R(N) + Z}{N} \leq \frac{ND + Z}{N}$$

یا

$$\max (D , ND_{max} - Z) \leq R(N) \leq N.D$$

جدول ۱-۷ حدود مجانبی را به طور خلاصه نشان می‌دهد. الگوریتم ۱ نیز گامهای محاسبات حدود مجانبی را برای ظرفیت‌کاری مدل بسته‌ای و ترمینال نشان می‌دهد (محاسبات برای نوع مدل تراکنشی ناچیز می‌باشد). دقت شود که تمامی حدود به جز حد نرخ خروجی بدبینانه در ظرفیت‌کاری ترمینال، خطوط راست می‌باشند.

جدول ۷-۱ حدود مجانبی

حدود	نوع ظرفیت‌کاری	
$\frac{1}{D} \leq X(N) \leq \mathrm{Min} (\frac{N}{D} , \frac{1}{D_{max}})$	batch	X
$\frac{N}{ND + Z} \leq$ $\leq \min (\frac{N}{D + Z} , X(N) \frac{1}{D_{max}})$	terminal	X
$X(\lambda) \leq \frac{1}{D_{max}}$	transaction	
$\max (D , ND_{max}) \leq R(N) \leq ND$	batch	R
$\max (D,ND_{max}-Z) \leq R(N) \leq ND$	terminal	R
$D \leq R(\lambda)$	transaction	R

توسط عملیات ریاضی، حدهای مجانبی به صورت تابعی از تعداد مشتریان شبکه، محاسبه شده و D و D_{max} تعیین می‌شوند. مقدار محاسبات به تعداد مراکز سرویس در مدل و به محدودهٔ جمعیت مشتریان مورد نظر وابسته است.

الگوریتم ۱:

در مراحل ذیل فرض شده ظرفیت‌کاری از نوع terminal می‌باشد، برای نوع batch باید Z=0 شود.

۱- مقادیر $D=\sum_{k=1}^{K} D_k$ و $D_{max}=\max D_k$ را محاسبه کنید.

۲- نقاط تقاطع اجزاء حدود خوش بینانه را محاسبه کنید.

$$N^* = \frac{D + Z}{D_{max}}$$

۳- حدود نرخ خروجی را از میان نقاط مذکور عبور دهید.

حد خوش بینانه : $(0, 0), (1, \frac{1}{D + Z}), N \le N^*$

$(0, \frac{1}{D_{max}}), (1, \frac{1}{D_{max}}), N \ge N^*$

حد بدبینانه: این حد برای N به صوت خطی نیست و برای هر N مورد نظر جداگانه محاسبه می‌شود (معادلات جدول ۱-۷)

۴- حدود میانگین زمان پاسخ را از نقاط مذکور عبور دهید:

حد خوش بینانه : $(0, D), (1, D), N \le N^*$

$(0, -Z), (1, D_{max} - Z), N \ge N^*$

حد بدبینانه: $(0, 0), (1, D)$

۳-۷ استفادهٔ حدود مجانبی

در این بخش سه نمونه از کاربردهای حدود مجانبی را ارائه می‌دهیم.

۱- بررسی موردی

۲- تخمین تأثیر کاهش تقاضای سرویس در گلوگاه

۳- مثالی از تحلیل یک سیستم پس از اصلاح

۱- مطالعهٔ موردی:

موردی را که بررسی می‌کنیم مربوط به مراکز کامپیوتر یک شرکت بیمه می‌باشد که در ۲۰ نقطهٔ جغرافیایی پراکنده شده‌اند (کامپیوترها قدیمی از نوع IBM ۳۷۹۰ است). به دلیل آنکه این سیستم زمان پاسخ قابل قبولی ندارد، شرکت سیستم‌های IBM ۸۱۳۰ و IBM۸۱۴۰ را در یک دورهٔ سه ساله جایگزین خواهد نمود. پس از مشورت با فروشنده، شرکت به این نتیجه رسید که استفاده از سیستم ۸۱۳۰ در بهبود کارایی مؤثر و نسبت به ۳۷۹۰، فاکتور بهبودی برابر ۲ به ۱/۵ دارد. هنگام استفاده از سیستم ۸۱۴۰ فاکتور بهبودی از ۲ به ۳/۵ خواهد رسید (البته هیچ عبارت دقیق و روشنی برای فرموله کردن فاکتور بهبودی کارایی وجود ندارد) [۱۲].

یک مدلسازی تحلیلی جهت تعیین مراکزی که ۸۱۳۰ برای آنها کافی است، انجام می‌گیرد (هزینهٔ ۸۱۳۰

کمتر از ۸۱۴۰ می‌باشد). هر دو سیستم ۸۱۳۰ و ۸۱۴۰ دارای دیسک‌هایی هستند که سریعتر از ۳۷۹۰ می‌باشد. در مورد Cpu آنها، پردازندهٔ ۸۱۳۰ از ۳۷۹۰ کندتر و ۸۱۴۰ حدود ۱/۵ برابر سریعتر از ۳۷۹۰ می‌باشد. نتایج زیر در مورد تقاضای سرویس، حاصل از اندازه‌گیری سیستم ۳۷۹۰ و آزمایش الگوسازی سیستم‌های ۸۱۴۰ و ۸۱۳۰ می‌باشد.

سیستم	Cpu	disk
	تقاضای سرویس (ثانیه)	
۳۷۹۰ (مشاهده)	۴/۶	۴
۸۱۳۰ (تخمین)	۵/۱	۱/۹
۸۱۴۰ (تخمین)	۳/۱	۱/۹

با توجه به تقاضای سرویس، از یک مدل حدّی جهت ارزیابی سیستم‌های مذکور استفاده می‌کنیم. شکل ۷-۲ شبکهٔ صف مدل فوق را نشان می‌دهد (با توجه به شکل هر چند بعضی سیستم‌ها دو دیسک دارند ولی کنترل کنندهٔ دیسک اجازهٔ فعالیت همزمان به آنها نمی‌دهد. در شکل یک مرکز سرویس دهندهٔ دیسکی داریم).

پارامترهای سیستم عبارتند از :

K - تعدادمراکز سرویس (۲)

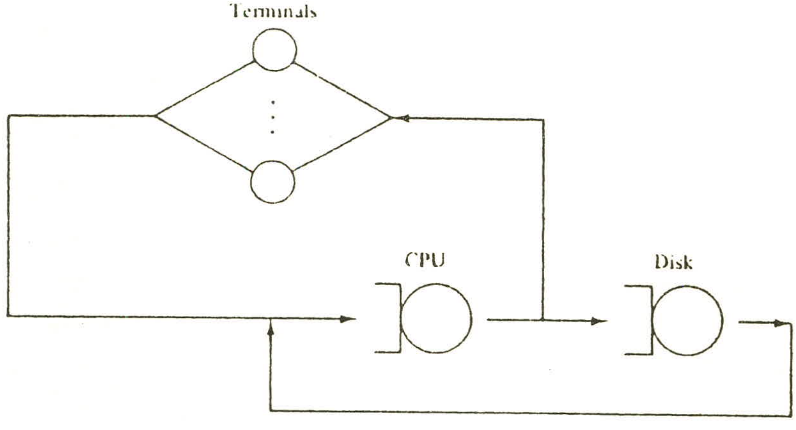
D_{max} - بیشترین تقاضای سرویس (۴/۶ ثانیه برای ۳۷۹۰، ۵/۱ برای ۸۱۳۰ و ۳/۱ برای ۸۱۴۰)

D- مجموع تقاضای سرویس (۸/۶، ۷ و ۵ به ترتیب)

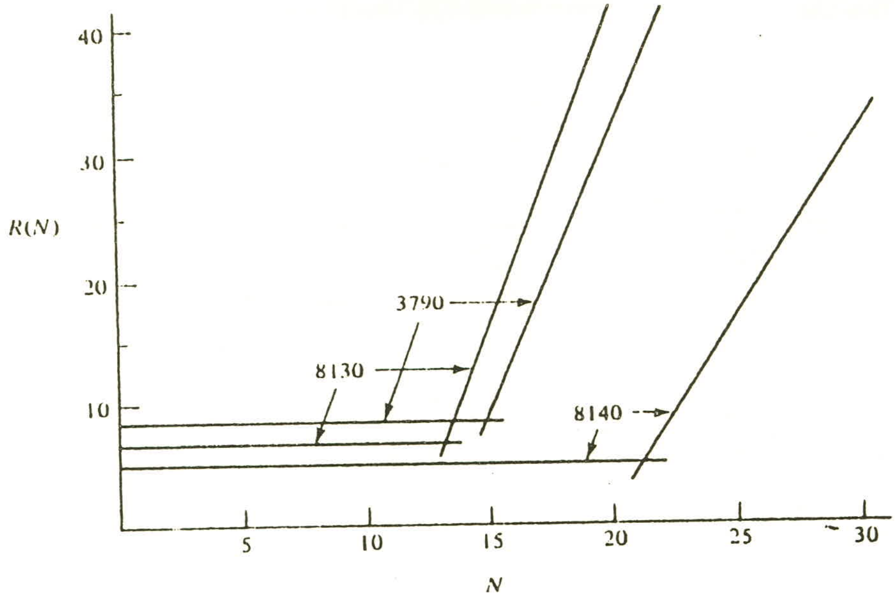
- نوع ظرفیتکاری مشتریان : ترمینال

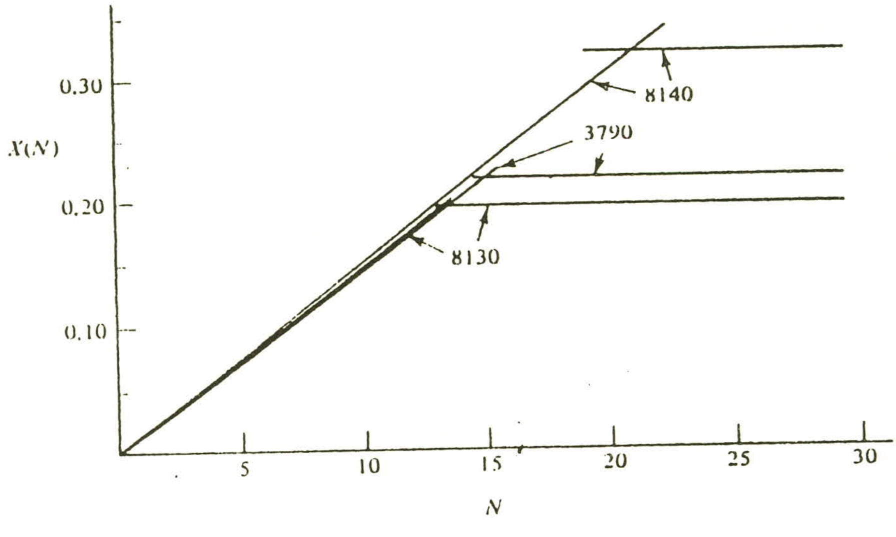
Z- میانگین زمان تفکر (یک تخمین ۶۰ ثانیه در نظر می‌گیریم).

با استفاده از الگوریتم ۱، حدهای مجانبی در شکل ۷-۳ برای مدل فوق رسم شده است.



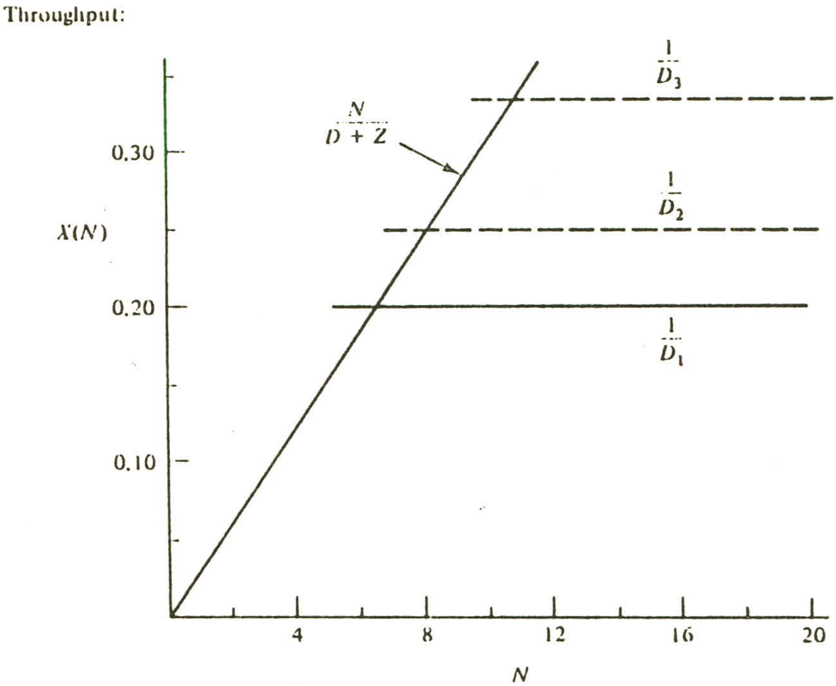
شکل ۷-۲ مطالعهٔ موردی

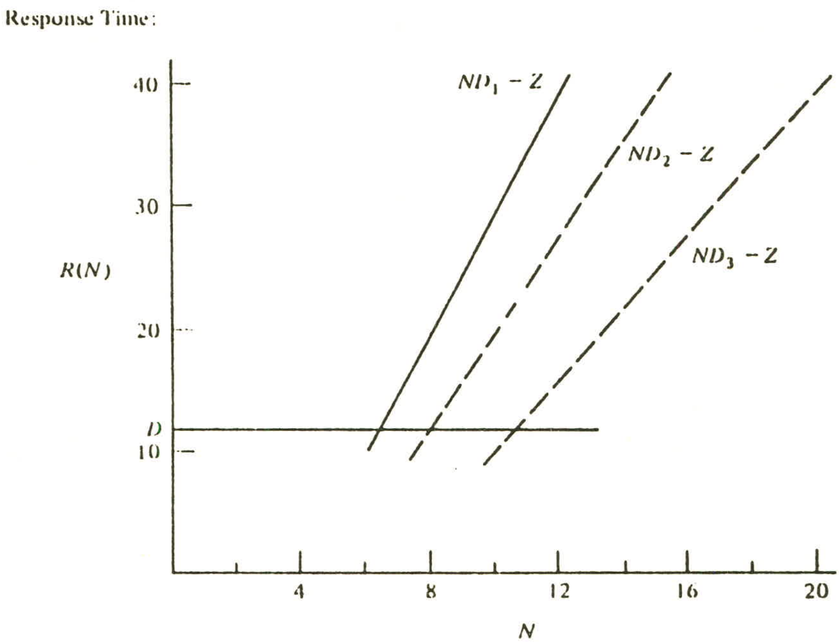




شکل ۷-۳

در بارهای سنگین، کارایی ۸۱۳۰ نسبت به ۳۷۹۰ کمتر است. چون Cpu ۸۱۳۰ کندتر و یک گلوگاه محسوب می‌شود. بنابراین هنگامی که تعداد ترمینالهای فعال از حد ماکزیمم تجاوز کند، به جای آنکه بهرهٔ کارایی از ۱/۵ به ۲ برسد، یک سیر نزولی خواهیم داشت. شکل مذکور بهبود کارایی را در حالتی که از ۳۷۹۰ به ۸۱۴۰ می‌رویم، نیز نشان می‌دهد. بررسی الگوسازی تأیید می‌کند که کارایی ۸۱۳۰ ضعیف‌تر از ۳۷۹۰ بوده البته در صورتی که تعداد ترمینالها از ۱۵ بیشتر شود.





شکل ۷-۴

بهرهٔ کارایی ۸۱۳۰ نسبت به ۳۷۹۰ در بارهای سبک نیز جزئی می‌باشد. بنابراین هیچ دلیلی برای تحقیق بیشتر در مورد ۸۱۳۰ وجود ندارد. نتیجه آنکه شرکت تصمیم می‌گیرد در تمامی مراکز ۸۱۴۰ نصب نماید. دقت کنیم که بدون تحلیل حدی ساده و مدلسازی ممکن بود شرکت سیستم ۸۱۳۰ را بدون انجام الگوسازی،

مورد استفاده قرار دهد و نتایج را باطل کند. در این مورد لزومی ندارد انواع ظرفیت‌کاری را بررسی نمائیم.

۲- تأثیر گلوگاه:

گلوگاه موجب می‌شود که نرخ خروجی تحت فشار قرار بگیرد و بیشتر از $\frac{1}{D_{max}}$ نشود. اگر با جایگزینی یک وسیلهٔ سریعتر یا انتقال بعضی پردازشها به وسایل و مراکز دیگر، D_{max} را کاهش دهیم حد نرخ خروجی، $\frac{1}{D_{max}}$، افزایش می‌یابد و در حقیقت گلوگاه کاهش می‌یابد.

یک محدودیت برای انجام این بهبود به مرکز سرویسی که، دومین تقاضای سرویس زیاد را دارا می‌باشد مربوط است. این مرکز را گلوگاه ثانویه می‌گوییم و با گلوگاه اولیه رقابت می‌کند. اکنون یک مدل با سه مرکز سرویس را در نظر می‌گیریم ($K = ۳$). ظرفیت‌کاری از نوع ترمینال، با میانگین زمان تفکر ۱۵ ثانیه ($Z = ۱۵$) و تقاضای سرویس در مراکز به ترتیب ۵ و ۴ و ۳ ثانیه درنظرمی‌گیریم ($D_۱ = ۵$, $D_۲ = ۴$, $D_۳ = ۳$).

شکل ۷-۴ حدهای مجانبی را برای مثال فوق نشان می‌دهد. بهبود کارایی توسط کاهش گلوگاه اولیه با گراف قابل مشاهده است. همانطور که بار مرکز گلوگاه کاهش می‌یابد، حد بار سنگین برای نرخ خروجی، افزایش می‌یابد. این در حالی است که حد بار سنگین برای میانگین زمان پاسخ، در جهت پایین محور کاهش می‌یابد (حول نقطهٔ (۰ و ۰) برای کارهای دسته‌ای و حول نقطهٔ ($Z - ۰$) برای کارهای ترمینال می‌باشد). مجانب‌های بار سبک نیز تغییر می‌کنند. اما نسبت به مجانبهای بار سنگین، حساسیت کمتری در مورد تقاضای سرویس در یک مرکز دارند.

با توجه به بهبود کارایی گلوگاه، در مراکز دیگر نیز بهبود حاصل می‌شود. با کاهش تقاضای سرویس در مراکز به جز گلوگاه، بهبود فقط در مجانب بار سبک حاصل شده است.

شکل ۷-۵ تأثیرات روی حدود مجانبی را به صورت مستقل، با دو برابر کردن سرعت (نصف کردن تقاضای سرویس) در گلوگاههای اولیه و ثانویه، با یکدیگر مقایسه می‌کند. همانطور که مشاهده می‌شود، فقط در بارهای سنگین، آنهم وقتی که نرخ تقاضا در گلوگاه اولیه کاهش یافته بهبود کارایی ظاهر می‌شود.

۳- تحلیل یک سیستم پس از اصلاح:

در این بخش جهت بازتاب اصلاح یک سیستم، از حدهای مجانبی استفاده کرده و مثالی را در نظر می‌گیریم. یک سیستم فعل و انفعالی با مقادیر اندازه‌گیری شدهٔ ذیل موجود است.

کل مدت زمان اندازه‌گیری مقادیر ثانیه $T = ۹۰۰$

مدت مشغول بودن Cpu ثانیه $B_۱ = ۴۰۰$

مدت مشغول بودن دیسک آهسته ثانیه $B_۲ = ۱۰۰$

مدت مشغول بودن دیسک سریع ثانیه $B_۳ = ۶۰۰$

تعداد برنامه‌های تکمیل شده $C_۱ = ۲۰۰$ job

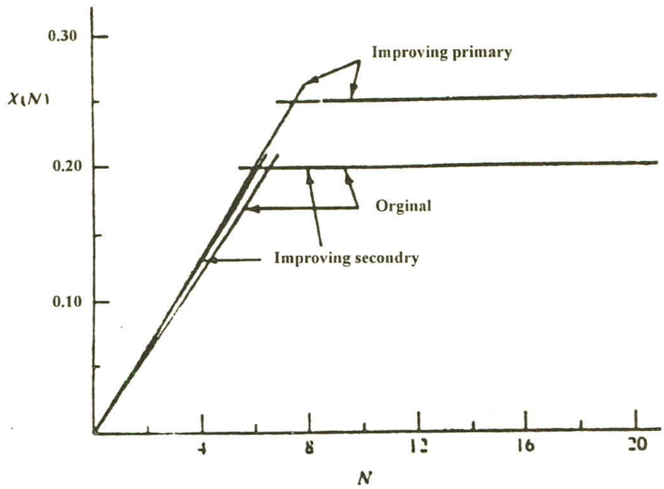
تعداد اعمال دیسک آهسته $C_۲ = ۲۰۰۰$

تعداد اعمال دیسک سریع $C_۳ = ۲۰۰۰۰$

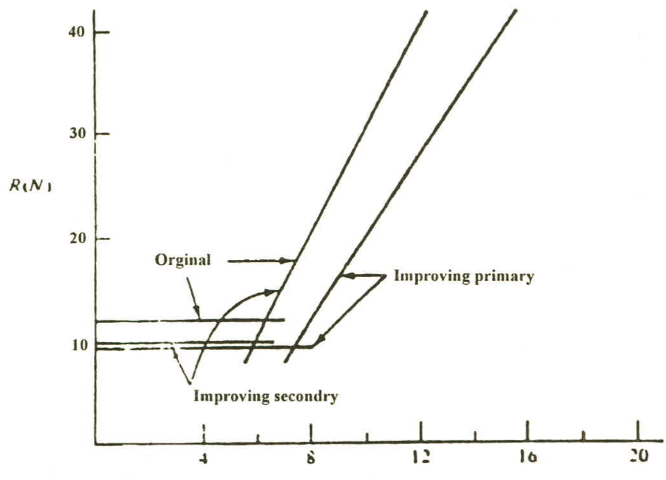
زمان تفکر ثانیه $Z = ۱۵$

کل مدت زمان سرویس مورد نیاز برنامه در مرکز ۱ ثانیه $D_۱ = ۲$

Throughput



Respons Time



شکل ۷-۵ تاثیرات روی حدود مجانبی

کل مدت زمان سرویس مورد نیاز برنامه در مرکز ۲ ثانیه $D_2=$ ۰/۵

کل مدت زمان سرویس مورد نیاز برنامه در مرکز ۳ ثانیه $D_3=$ ۳

تعداد ملاقات مربوط به دیسک‌ها به صورت $V_2=$ ۱۰ و $V_3=$ ۱۰۰ بوده و زمان‌های سرویس در هر ملاقات به صورت $S_2=$ ۰/۰۵ و $S_3=$ ۰/۰۳ می‌باشد.

چهار بهبود به شرح ذیل در سیستم می‌تواند پدیدار شود که انعکاس هر یک را در پارامترهای مدل بررسی می‌کنیم.

۱- Cpu را با Cpu دیگری با سرعت دوبرابر جایگزین کنیم.

۲- بعضی از فایلها را از دیسک سریعتر به دیسک آهسته انتقال دهیم (با متعادل کردن تقاضای آنها). در اینجا

فقط تأثیرات اولیه که تغییر سرعت دیسک می‌باشد را در نظر می‌گیریم. از تأثیرات ثانویه که در حقیقت اختلاف میانگین اندازهٔ بلاکهای انتقالی بین دو دیسک می‌باشد، صرفنظر می‌کنیم. تقاضای سرویس در دیسک جدید به صورت زیر نتیجه گیری می‌شود.

$$V_۲ + V_۳ = ۱۱۰ \quad \text{و} \quad S_۲ = ۰/۰۵ \;\text{و}\; S_۳ = ۰/۰۳$$

بنابراین، این رابطه به صورت زیر درمی‌آید.

$$\frac{V_۲S_۲}{۰/۰۵} + \frac{V_۳S_۳}{۰/۰۳} = ۱۱۰$$

چون می‌خواهیم داشته باشیم:

$$D_۲ = V_۲ \,.\, S_۲ = V_۳ \,.\, S_۳ = D_۳$$

بنابراین :

$$D_۲ = [\frac{۱}{۰/۰۵} + \frac{۱}{۰/۰۳}] = ۱۱۰ \quad \text{و} \quad D_۲ = D_۳ = ۲/۰۶$$

با تقسیم زمانهای سرویس مناسب، تعداد ملاقات جدید عبارتست از:

$$V_۲ = ۴۱ \,,\, V_۳ = ۶۹$$

۳- یک دیسک سریع دوم را اضافه کنیم (مرکز ۴)، که نصف بار دیسک مشغول کنونی را به عهده بگیرد. اینبار نیز تأثیرات اولیه را در نظر می‌گیریم:

$$D_۴ = ۱/۵ \quad , \quad D_۳ = ۱۰۵ \quad , \quad K = ۴$$

۴- سه تغییر با همدیگر اعمال می‌کنیم. Cpu سریعتر و یک بار متعادل بین دو دیسک سریع و یک دیسک آهسته توزیع می‌گردد. تقاضای سرویس به ترتیب برابر است با:

$$D_۱ = ۱ \qquad D_۳ = ۱/۲۷$$

$$D_۲ = ۱/۲۷ \qquad D_۴ = ۱/۲۷$$

می‌دانیم که $V_۲ + V_۳ + V_۴ = ۱۱۰$ برای اطمینان از اینکه $D_۲ = D_۳ = D_۴$ داریم :

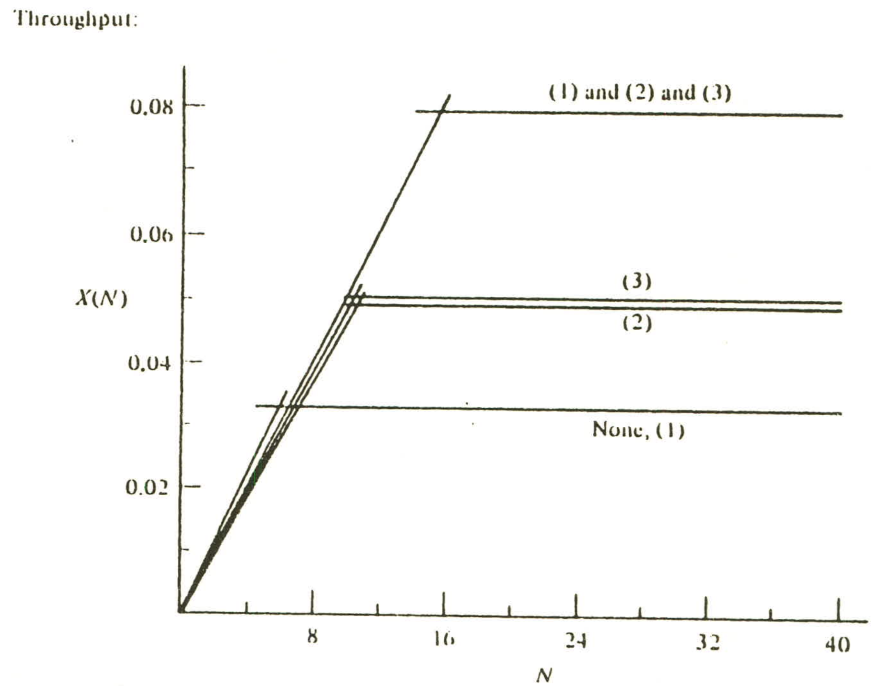
$$\frac{V_۲S_۲}{۰/۰۵} + \frac{V_۳S_۳}{۰/۰۳} + \frac{V_۴S_۴}{۰/۰۳} = ۱۱۰$$

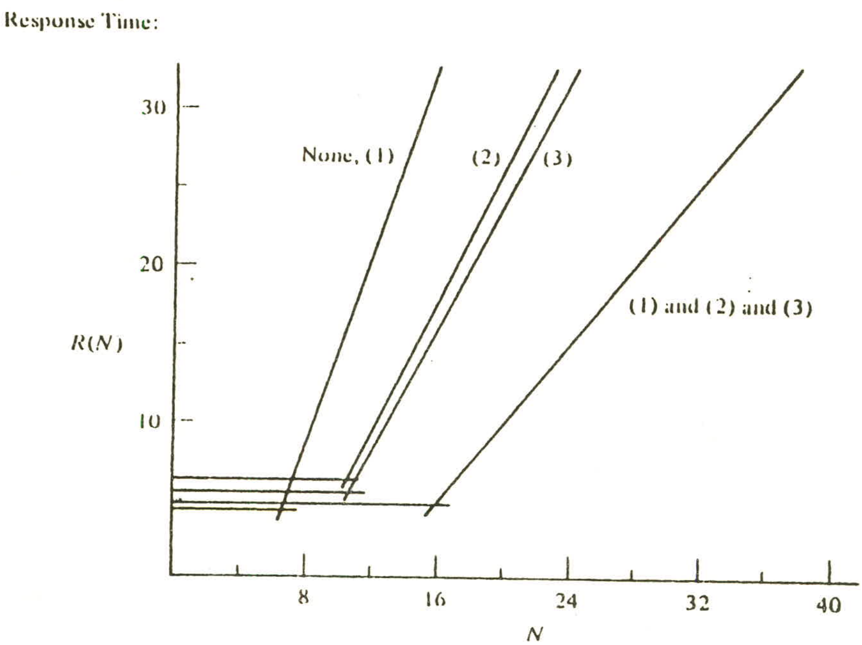
$$D_۲ [\frac{۱}{۰/۰۵} + \frac{۱}{۰/۰۳} + \frac{۱}{۰/۰۳}] = ۱۱۰$$

$$D_۲ = D_۳ = D_۴ = [\frac{۰/۰۰۱۵}{۰/۱۳}]\, ۱۱۰ = ۱/۲۷$$

شکل ۶-۷ حدود مجانبی را برای سیستم اصلی نشان می‌دهد (برچسب "None"). برای هر اصلاح به طور جداگانه شکل رسم شده است (برچسب‌های "۱" و "۲" و "۳"). برای ترکیب هر سه اصلاح باهمدیگر نیز، شکل دیگری رسم شده است (برچسب "۱ و ۲ و ۳").

در ابتدا شاید اولین تغییر از همه مهم‌تر جلوه کند. ولی همانطور که شکل نشان می‌دهد، چون دیسک سریع گلوگاه اصلی می‌باشد، تغییرات ۲ و ۳ اثرات بیشتری دارند. دقت شود که تغییر ۲ به اندازهٔ تغییر ۳، مدل را بهبود می‌بخشد، در حالیکه نیاز به سخت‌افزار اضافی ندارد. ترکیب کردن هر سه اصلاح نتایج مهمی را در پی دارد.





شکل ۷-۶

۷-٤ حدهای سیستم متعادل

با انجام محاسبات بیشتر، نسبت به حدود مجانبی، می‌توان حدهای دقیق‌تری را بدست آورد. چون این حدود براساس "سیستم‌های متعادل" پایه گذاری شده، "حدود سیستم متعادل" نام دارند. در این سیستم میزان تقاضای سرویس در هر مرکز یکسان می‌باشد یعنی:

$$D_1 = D_2 = D_3 = \dots = D_k$$

شکل ۱-۷a ، فرم عمومی حدهای سیستم متعادل را برای ظرفیت‌کاری دسته‌ای و شکل ۲-۷b برای ظرفیت کاری ترمینال نمایش می‌دهد (همراه با حدهای مجانبی). بررسی ویژگیهای سیستم متعادل سبب می‌شود که حدهای ارزیابی مکمل حدهای مجانبی تعیین شده و بدین ترتیب درک عمیق‌تری از رفتار سیستم داشته باشیم.

حدهای مذکور برای انواع ظرفیت‌کاری (transaction, terminal, batch) در جدول ۷-۲ داده شده است. تحلیل سیستم‌های متعادل یکی از تکنیکهایی که در فصل بعدی ارائه شده است. این تحلیل‌ها نیاز به فرضهایی دارد که باید در مورد مدل درنظر گرفته شوند (این فرضها را در فصل بعدی شرح می‌دهیم).

در مورد حدهای مجانبی تقاضای سرویس در یک مرکز برای یک مشتری به تعداد مشتری‌ها و یا به محل مراکز سرویس بستگی ندارد. ولی در مورد حدهای سیستم متعادل اینطور نیست. در این سیستم‌ها بکارگیری هر مرکز سرویس توسط رابطهٔ زیر بیان می‌گردد.

$$U_k(N) = \frac{N}{N + K - 1}$$

که در فصل بعدی آنرا مطرح خواهیم نمود. با توجه به قانون بکارگیری، نرخ خروجی سیستم عبارتست از:

$$X(N) = \frac{U_k}{D_k} = \frac{N}{N+K-1} \times \frac{1}{D_k}$$

که D_k میزان تقاضای سرویس در هر مرکز می‌باشد.

D_{max} ، D_{ave} ، D_{min} را به ترتیب ماکزیمم، میانگین و می‌نیمم تقاضای سرویس در مراکز مدل مورد بررسی در نظر می‌گیریم.

نرخ خروجی سیستم را به دو وضعیت یک سیستم متعادل محدود می‌کنیم. یک حالت با ماکزیمم نرخ تقاضای سرویس در هر مرکز و حالت دیگر با می‌نیمم نرخ تقاضای سرویس در هر مرکز، بنابراین:

$$\frac{N}{N+K-1} \times \frac{1}{D_{max}} \leq X(N) \leq \frac{N}{N+K-1} \times \frac{1}{D_{min}}$$

نامساوی فوق برای تمامی سیستم‌های متعادل که دارای K مرکز سرویس دهی و N مشتری و ماکزیمم تقاضای سرویس D_{max}، (کمترین نرخ خروجی) و همچنین می‌نیمم تقاضای سرویس D_{min} (بیشترین نرخ خروجی)، برقرار است.

متناظر با رابطهٔ فوق حدود مربوط به میانگین زمان پاسخ عبارتست از:

$$(N+K-1)\, D_{min} \leq R(N) \leq (N+K-1).\, D_{max}$$

برای دستیابی به حدهای دقیق‌تر، لازم است شرایطی را روی میزان تقاضای سرویس ماکزیمم و کل تقاضای سرویس اعمال کنیم. می‌دانیم که چنین سیستم‌ها تقاضای سرویس معینی دارند یعنی $D = \sum_{k=1}^{K} D_k$.
بنابراین دارای بالاترین نرخ خروجی و کمترین میانگین زمان پاسخ می‌باشند. این در حالی است که همهٔ تقاضاهای سرویس با یکدیگر برابر باشند یعنی:

$$D_k = \frac{D}{K} \qquad k = 1, \ldots, K$$

این مطلب تأیید می‌کند که نسبت افزایش تأخیر (در سرویس گرفتن) حاصله از افزایش بار، بیشتر از نسبت

جدول ۷-۲ خلاصه حدود سیستم‌های متعادل

	نوع ظرفیت‌کاری	حدود
X	**batch**	$\frac{N}{D+(N-1)D_{max}} \le X(N) \le \min\left(\frac{1}{D_{max}}, \frac{N}{D+(N-1)D_{ave}}\right)$
	terminal	$\frac{N}{D+Z+\frac{(N-1)D_{max}}{1+\frac{Z}{ND}}} \le X(N) \le \min\left(\frac{1}{D_{max}}, \frac{N}{D+Z+\frac{(N-1)D_{ave}}{1+\frac{Z}{D}}}\right)$
	transaction	$X(\lambda) \le \frac{1}{D_{max}}$
R	**batch**	$\max\left(ND_{max}, D+(N-1)D_{ave}\right) \le R(N) \le D+(N-1)D_{max}$
	terminal	$\max\left(ND_{max} - Z\ D+\frac{(N-1)D_{ave}}{1+\frac{Z}{D}}\right) \le R(N) \le D+\frac{(N-1)D_{max}}{1+\frac{Z}{ND}}$
	transaction	$\frac{D}{1-\lambda D_{ave}} \le R(\lambda) \le \frac{D}{1-\lambda D_{max}}$

کاهش تأخیر حاصله از کاهش بار، می‌باشد. بنابراین حدود مورد نظر توسط رابطهٔ زیر تعیین می‌شوند.

$$X(N) \le \frac{N}{N+K-1} \times \frac{1}{D_{ave}} = \frac{N}{D+(N-1)\,D_{ave}}$$

همچنین داریم:

$$R(N) \ge D + (N-1)\,D_{ave}$$

توجه شود که حد سیستم متعادل، حد مجانبی را در وضعیت بار سنگین، از وسط قطع می‌کند (نقطهٔ تلاقی را N^+ در نظرمی‌گیریم). ماورای این نقطه طبق تعریف حد سیستم متعادل با حد مجانبی منطبق بر یکدیگر می‌باشند.

به طور کلی تمامی سیستم‌هایی که کل تقاضای آنها D و ماکزیمم تقاضا D_{max} دارند دارای تعداد $\frac{D}{D_{max}}$ مرکز سرویس، با تقاضای D_{max} و کمترین نرخ خروجی و در بقیهٔ مراکز تقاضا صفر است (در حقیقت $\frac{D}{D_{max}}$ ممکن است عدد صحیح نباشد). پس حدهای بدبینانه عبارتند از:

$$\frac{N}{N+\frac{D}{D_{max}}-1} * \frac{1}{D_{max}} = \frac{N}{D+(N-1)\,D_{max}} \le X(N)$$

و

$$R(N) \le D + (N-1)\,D_{max}$$

جدول ۷-۲ حدود سیستم متعادل را برای ظرفیت‌کاریهای مختلف خلاصه کرده و الگوریتم ۲ نیز چگونگی محاسبهٔ حدود مذکور رابرای ظرفیت‌کاری دسته‌ای و ترمینال بیان می‌کند (محاسبات در مورد ظرفیت‌کاری تراکنشی ناچیز می‌باشد).

حدود برای ظرفیت‌کاری دسته‌ای، در مورد میانگین زمان پاسخ، خطوط راست می‌باشند همچنین حد خوش‌بینانه برای میانگین زمان پاسخ در مورد ظرفیت‌کاری ترمینال، نیز خط راست می‌باشد.

حدود سیستم‌های متعادل در مورد نرخ خروجی و حد بدبینانه این سیستم‌ها در مورد زمان پاسخ، برای ظرفیت‌کاری از نوع ترمینال خطی نمی‌باشد (نسبت به N) و باید برای هر مقدار خاص N، جداگانه محاسبه شود.

الگوریتم ۲ :

۱- با استفاده از الگوریتم ۱ حدهای مجانبی را محاسبه کنید.

۲- نقطهٔ تقاطع حد خوش بینانهٔ سیستم متعادل و حد خوش بینانهٔ مجانبی را تعیین کنید.

برای ظرفیت‌کاری batch :
$$N^{+} = \frac{D - D_{ave}}{D_{max} - D_{ave}}$$

برای ظرفیت‌کاری terminal:
$$N^{+} = \frac{(D+Z)^{2} - D.D_{ave}}{(D+Z)\,D_{max} - D.D_{ave}}$$

حد خوش بینانه سیستم متعادل فقط از ۱ تا N^{+} باید محاسبه شود و طبق تعریف از N^{+} به بعد، با حد مجانبی یکسان است.

۳- حدهای میانگین زمان پاسخ سیستم متعادل را حساب کنید. در مورد ظرفیت‌کاری batch حدها خطوطی هستند که از نقاط زیر می‌گذرند:

حد خوش بینانه : $(0, D - D_{ave})$, $(1, D)$

حد بدبینانه : $(0, D - D_{max})$, $(1, D)$

برای ظرفیت‌کاری terminal حدها خطوطی هستند که از نقاط زیر می‌گذرند:

حد خوش بینانه : $\left(0, D - \frac{D_{ave}}{1 + \frac{Z}{D}}\right)$, $(1, D)$

حد بدبینانه : این حد برای نوع terminal نسبت به N خطی نیست و با استفاده از معادلات جدول ۲ برای هر N خاص محاسبه می‌شود.

۴- حدود سیستم متعادل را در مورد نرخ خروجی برای تعداد N مورد نظر با استفاده، از معادلات جدول ۲ محاسبه کنید.

۸ ارزیابی سیستم های کامپیوتری

۸-۱ مقدمه

در این فصل ارزیابی مدل شبکهٔ صف با یک کلاس (یک نوع مشتری) را بررسی می‌کنیم. این مدل تخمین مقادیر کارآئی را بهتر از حدهای ساده فراهم می‌کند. مثلاً به جای آنکه نرخ خروجی یک سیستم با جمعیت معینی را بین ۱/۱ تا ۲ مشتری در دقیقه تعیین کند، مدل تک کلاسی، یک تخمین حقیقی از نرخ خروجی (مثلاً ۱/۷ مشتری در دقیقه) می‌زند.

در اینگونه مدل فرض می‌کنیم مشتریان از یکدیگر متمایز نیستند. لازم به ذکر است هر چند این مدل ساده می‌باشد ولی برای بیان دقیق یک سیستم واقعی مناسب نیست. وضعیتهایی که در این مدلها ممکن است بررسی شوند عبارتند از:

- **اطلاعات افزوده شده:** در بررسی حدی، اطلاعات کافی تهیه نمی‌شود ولی در جهت پیشرفت مدل با اطلاعات افزوده شده مدل یک کلاسی گام بعدی می‌باشد.
- **ظرفیت‌کاری یگانه (W.L):** سیستم‌های کامپیوتری، ممکن است جهت ارزیابی فقط یک ظرفیت‌کاری را اجرا نمایند. بنابراین لزومی جهت ارائه ظرفیت‌کاری دیگر به طور دقیق نیست.
- **ظرفیت‌های کاری یکنواخت:** در یک سیستم کامپیوتری، مؤلفه‌های مختلف ظرفیت‌کاری امکان دارد تقاضای سرویس مشابهی دارند. که به همین دلیل می‌توان مدل را خلاصه‌بندی کرده و همهٔ آنها را در یک کلاس مشتری در نظرگرفت. به طور عکس گاهی نمایش یک سیستم کامپیوتری توسط مدل یک کلاسی نامناسب است. این موارد نوعاً یا ناشی از وجود مؤلفه‌های مختلف ظرفیت کاری بوده (که نشان دهندهٔ منابع مختلف موجود است) و یا به دلیل فرضهای مدلسازی می‌باشد. بطوریکه ورودی‌ها یا خروجی‌ها در

مؤلفه‌های مختلف ظرفیت کاری، به طور جداگانه، قرار گرفته و نتوان آنها را تحت یک ظرفیت کاری مجموع، بررسی نمود.

– ظرفیت کاری‌های مجزای چندگانه: در یک سیستم که شامل ظرفیت کاری دسته‌ای و اشتراک زمانی (timesharing) می‌باشد، امکان دارد کارهای دسته‌ای به صورت cpu bound (بیشتر از Cpu سرویس می‌گیرند) باشد و کارهای ظریفت کاری دوّم (timesharing) به صورت I/O bound (بیشتر از I/O سرویس می‌گیرند) انجام شود. برای چنین سیستمی اگر یک مدل یک کلاسی در نظر بگیریم یک میانگین تعداد "برنامه" نیز مطرح می‌شود ولی چون برنامه‌ها در سیستم واقعی رفتار یکسانی ندارند، پس تصویر درستی از سیستم با مدل یک کلاسی ایجاد نمی‌شود.

– کلاس وابسته به ورودی: در یک سیستم مرکب (دسته‌ای /اشتراک زمانی)، فرض می‌کنیم که ظرف دو سال آینده کارهای اشتراک زمانی رشد ۱۰۰٪ داشته و کارهای دسته‌ای رشدی معادل ۱۰٪ داشته باشند. مدل یک کلاسی چون "میانگین" تعداد مشتریان را در نظر می‌گیرد، نمی‌تواند پارامترهای ورودی مربوط به مؤلفه‌های ظرفیت کاری (که نرخهای رشد متفاوتی دارند) در نظر بگیرد.

– کلاس وابسته به خروجی: مشابه مورد قبلی سیستم یک کلاسی "میانگین" برای خروجیهای مدل بیان می‌کند که با توجه به نوع اجرا در داخل سیستم، تفسیر مقادیر کارآیی مشکل می‌باشد. بنابراین برای خروجیهای مختلف، مدلی با چندین کلاس مورد نیاز است. سیستم‌ها ذاتاً دارای ویژگیهای مختلفی می‌باشند که جهت مدلسازی آنها، باید چندین کلاس در نظر گرفت، این مدلها را در مباحث بعدی مفصل بیان می‌کنیم.

۸-۲ نمایش ظرفیت کاری

نمایش ظرفیت کاری در مدل شبکهٔ صف یک کلاسی با دو ورودی مدل، مجموعه میزان تقاضای سرویس و شدت ظرفیت کاری، تعیین می‌شود.

در این مدل فرض می‌کنیم که برنامه‌ها یکسان اجرا شده و اختلاف آنها اثر مهمی روی کارایی سیستم ندارد. بنابراین برای محاسبهٔ مجموعهٔ تقاضای سرویس، به آسانی با یک مجموع یگانه سروکار داریم (در مدلهای چندکلاسی، تعداد کلاس مشتریان تعیین شده و برای هر کلاس تقاضای سرویس مجزایی محاسبه می‌گردد). تعیین ظرفیت کاری و شدت آن دارای دو مرحله می‌باشد. انتخاب یک نوع ظرفیت کاری مناسب (دسته‌ای، ترمینال و یا تراکنشی) و سپس نسبت دادن پارامترهای مناسب به شدت ظرفیت کاری تعیین شده. چون متناظر با ظرفیت کاری سیستم‌های کامپیوتری، در مدل شبکهٔ صف نیز سه نوع ظرفیت کاری وجود دارد، پس انتخاب نوع ظرفیت کاری براحتی صورت می‌گیرد.

همچنین سیستم‌های باز (کلاس با ظرفیت کاری از نوع تراکنشی) از سیستم بسته (کلاس از نوع دسته‌ای یا ترمینال) تفکیک می‌شوند. چون تعداد مشتری در سیستم‌های باز در هر زمان، هیچ محدودیتی ندارد و در سیستم بسته جمعیت مشتری محدود، بنابراین زمان پاسخ در سیستم باز بیشتر از سیستم بستهٔ متناظر می‌باشد (با نرخ خروجی یکسان). از این جهت برخلاف مدلهای بسته، طول صف در سیستم‌های باز،

فوق العاده بزرگ می‌شود. این اختلاف، زمانی که بعضی از سرویس دهنده‌ها در سیستم در حال اشباع می‌باشند، مهم‌تر جلوه می‌کند.

مسائل مذکور ما را بدان سمت می‌برد که پارامترهای شدت ظرفیت کاری را چگونه قرار دهیم. در مدلهای شبکهٔ صف شدت ظرفیت کاری، یک کمیت ثابت می‌باشد. در حالی که در یک سیستم کامپیوتری ممکن است تغییر کند. با این وجود، مدلهای شبکهٔ صف در وضعیتهای گوناگون برای سیستم کامپیوتری مدل مفیدی می‌باشند که از جمله:

- **وضعیت بار سنگین:** در این موارد رفتار سیستم را تحت بیشترین بار ممکن بررسی می‌کنیم. طبق فرض در بارهای سنگین، همواره برنامه‌هایی منتظر ورود به حافظه می‌باشند و هنگامی که یک برنامه، کاملاً اجرا شود، حافظهٔ آن آزاد شده و بلافاصله برنامه دیگری جایگزین آن می‌شود.

بنابراین در این سیستم‌ها، ظرفیت کاری به صورت دسته‌ای نمایش داده شده و تعداد ثابتی مشتری (که در حقیقت معادل بیشترین سطح چند برنامگی است) در سیستم همواره وجود دارد.

- **ظرفیت کاری غیرصحیح (حقیقی):** اندازه‌گیری داده‌های یک سیستم، در مورد "میانگین سطح چند برنامگی" نشان می‌دهد که این پارامتر ممکن است یک عدد صحیح نباشد. بعضی الگوریتم‌ها تعداد غیرصحیح (non-integer) جمعیت مشتری‌ها را در مدل شبکهٔ صف مجاز می‌دانند.

در موارد فوق مدل برای مقادیر صحیح ارزیابی شده و سپس حل برای مقادیر غیرصحیح به آن الحاق می‌کنیم. مثلاً اگر سطح چندبرنامگی اندازه‌گیری شده برابر ۴/۵ باشد، حل مدل با جمعیت دسته‌ای، ۴ و ۵، محاسبه شده و میانگین آن را برای ۴/۵ (سیستم مورد نظر) پیشنهاد می‌کنیم.

- **توزیع شدت ظرفیت کاری:** اندازه‌گیری داده‌های یک سیستم، توزیع شدت ظرفیت‌های کاری را نیز مشخص می‌کند. مثلاً شدت ظرفیت‌کاری زمانی‌که nکاربرترمینال به صورت فعال در سیستم هستند متناسب با P(N=n) بدست می‌آید.

توزیع مذکور جهت ارزیابی حل‌های بدست آمدهٔ یک مدل می‌تواند استفاده شود (با هر تعداد کاربر دلخواه) جدول ۸-۱ مثالی را ارائه می‌دهد.

$$U_{cpu} = \sum_{n=۱}^{۴} P\,(N=n)\; U_{cpu}\;(n) = ۰.۰۶۴۵$$

$$R = \sum_{n=۱}^{۴} \left[\frac{X\,(n)\;.P\,(N=n)}{\sum_{j=۱}^{۴} x\;(j)\;.P\;(N=j)} R\;(n) \right] = ۲.۴۹۲$$

جدول ۸-۱ محاسبهٔ پارامترهای مختلف ارزیابی

n	$P[N=n]$	$U_{CPU}(n)$	$X(n)$	$R(n)$
0	.1	0	0	0
1	.2	.032	.0525	.787
2	.3	.062	.1031	1.546
3	.3	.092	.1515	2.273
4	.1	.119	.1974	2.961

- بررسی اندازه (Sizing) :

حل مدل یک کلاسی سریع بدست آمده و بنابراین از آنها جهت ارزیابی مدلهای با تعداد ظرفیت‌کاریهای مختلف، استفاده‌می شود.ا کنون به‌عنوان‌مثال‌سؤالی‌را مطرح می‌کنیم. ماکزیمم نرخ ورودی از نوع تراکنش که با میانگین زمان پاسخ کمتراز ۳ ثانیه می‌تواند وارد شود، چقدر است؟

برای پاسخ به این سؤال در مدل اولیه، با تغییر در نرخ ورودی (... ،۱،۲ = λ) و بررسی زمانهای پاسخ بدست آمده، می‌توان به جواب رسید.

- بررسی دقیق و قوی: همواره به طور دقیق نمی‌توان میزان رشد ظرفیت کاری را پیش بینی نمود. بنابراین در مدل یک کلاسی جهت ارزیابی ظرفیت‌های کاری مختلف مورد نظر در آن، به تحلیلگر کمک می‌کند، هنگامی که ظرفیت کاری از مقادیر موردنظر فراتر رود، پیش بینی بهتری انجام دهد.

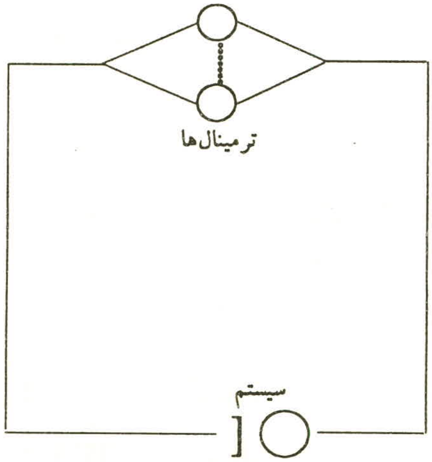
۸-۳ مطالعهٔ موردی [۱۲]

در این بخش سه نمونه از کاربردهای شبکهٔ صف یک کلاسی بررسی می‌شوند. ابتدا یک بررسی کلاسیکی برای یک مورد ساده که ارزیابی دقیقی را ارائه داده، انجام می‌دهیم. سپس تأثیر اعمال برخی اصلاحات سخت افزاری و نرم افزاری را بررسی نموده و در نهایت استفادهٔ مدل، جهت طرح ظرفیت‌کاری را بیان می‌نماییم.

۸-۳-۱ مدل برای سیستم فعل و انفعالی (interactive)

بررسی این مورد نشان می‌دهد که مدلهای آسان، می‌توانند پیش بینی‌های خوبی از کارایی سیستم‌ارائه‌دهد. سیستم‌مورد نظر یک IBM 7094 که به صورت اشتراک زمانی سازگار می‌باشد.

این سیستم فعل و انفعالی و براساس جایگزینی (swapping) کار می‌کند. در هرزمان فقط یک کاربر می‌تواند فعال باشد. همانطور که در شکل ۸-۱ دیده می‌شود کل سیستم (Cpu، حافظه، دیسک) به صورت برش



مدل سیستم فعل و انفعالی

شکل ۸-۱ مدل سیستم فعل و انفعالی

زمانی (time-slice) و برای کاربر به صورت یک سرویس دهندهٔ واحد عمل می‌کنند. هدف ما بررسی رفتار زمان پاسخ سیستم به صورت تابعی از تعداد کاربرها می‌باشد.

در ظرفیت کاری از نوع ترمینال می‌باشد. نمایش یک مرکز سرویس به دلیل آن است که در هر زمان فقط یک کاربر فعال، و حالت همزمانی در پردازش دیسک و Cpu نداریم. به همین دلیل میانگین زمان پاسخ در Cpu و یا دیسک‌ها با اهمیت، ولی میزان زمان سپری شده مهم‌تر می‌باشد.

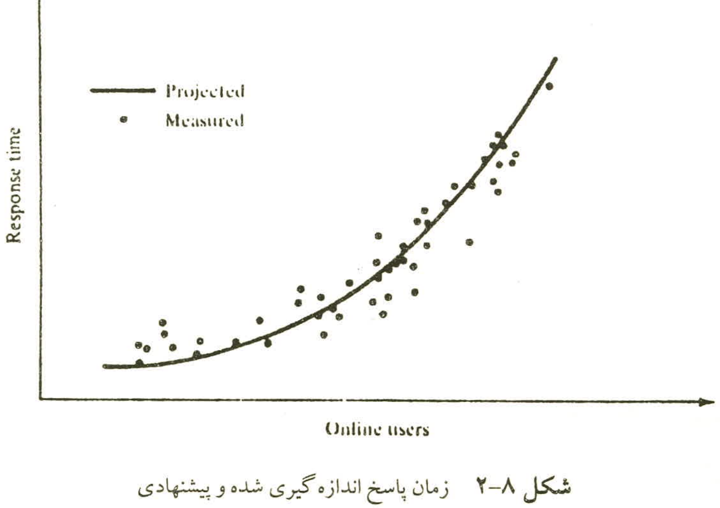
دقت شود که وجود یک مرکز سرویس واحد، مشکل حافظه را حل نموده است. زیرا در مدل، اگر برای Cpuو دیسک مراکز سرویس جدایی در نظر بگیریم اجازهٔ پردازش همزمان در مراکز را به مشتریان داده‌ایم و بنابراین مدل دقت کمتری خواهد داشت.

مدل با استفاده از اندازه گیری داده‌ها در طی مدت استفاده از سیستم، پارامتربندی می‌شود. این پارامترها شامل، میانگین زمان تفکر کاربر، میانگین زمان پردازش دیسک و Cpu، میانگین حافظهٔ مورد نیاز و ... می‌باشد. میزان تقاضای سرویس در مرکز سرویس، برابر با مجموع زمان پردازش و میزان سرویس مورد نیاز دیسک می‌باشد. تعداد مشتری‌های مدل قابل تغییر بوده و برای هر تعداد دلخواه، زمان پاسخ سیستم را تخمین می‌زنیم. شکل ۸-۲ مدل پیشنهادی را با مقادیر اندازه گیری شدهٔ زمان پاسخ مقایسه می‌کند.

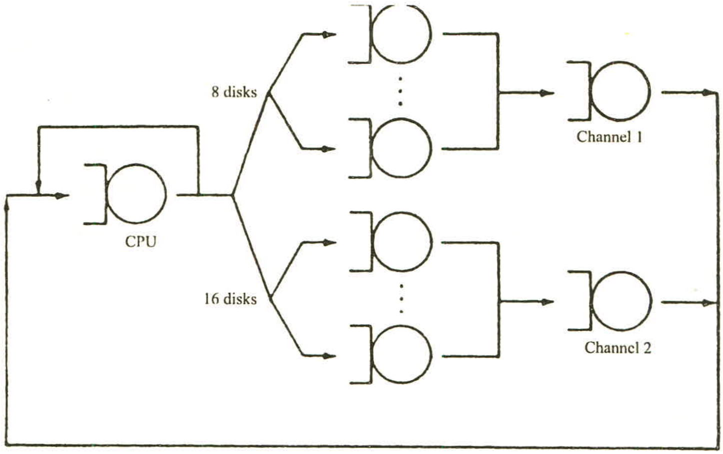
۸-۳-۲ بررسی اعمال اصلاحات روی مدل

در این جا از مدل یک کلاسی جهت ارزیابی اعمال اصلاحات در پیکره بندی سخت افزاری و نرم افزاری سیستم استفاده می‌کنیم. سیستم مورد نظر یک IBM 360 مدل 56j می‌باشد که دارای سه کانال به شرح زیر می‌باشد. کانالهای ۱ و ۲ به ترتیب ۸و ۱۶ دیسک از مدل IBM 2314 بوده و کانال سوم به یک DRUM متصل شده است. DRUM منحصراً توسط سیستم عامل استفاده می‌شود. مشتریان در مدل معرف برنامه‌های کاربران بوده و با DRUM برخوردی ندارند.

شکل ۸-۳مدل مذکور را نشان می‌دهد. با تعیین زمان‌های سرویس Cpu و دیسک‌ها و کانال‌ها، مدل طبق



شکل ۸-۲ زمان پاسخ اندازه گیری شده و پیشنهادی



شکل ۸-۳ مدل سیستم اصلاح شده

نوع ظرفیت‌کاری و شدت ظرفیت‌کاری، پارامتربندی می‌گردد. به دلیل مشخصه‌هایی کـه سـیستم داراست (مثلاً مراکز سرویس کانالها) دقت خوبی قابل دسترسی می‌باشد.

برای تخمین میزان تقاضای سرویس Cpu برای هر برنامه، کل زمان مشغول بودن Cpu (هم برای سیستم و هم برای کاربر) را بر تعداد برنامه‌هایی که در این فاصله اجرا شده‌اند، تـقسیم مـی‌کنیم. بـنابرایـن سـربار Cpu (Overhead) سیستم (نظیر زمانبندی بین Cpu و I/O) به طور معادل بین تمامی برنامه‌ها توزیع می‌شود.

پارامتربندی زیر سیستم‌های I/O (کانالها و دیسک‌ها) پیچیده‌تر می‌باشد. زیرا تکـنولوژی مـربوط بـه دیسک سیستم طوری است که طی مدت چرخش (rotat latency) هم دیسک و هم کانال مربوط به آن باید اشغال شوند و داده‌ها انتقال یابند. در حالی که جستجو در دیسک (چرخش هد دیسک) می‌توانست مستقل از کانال صورت بگیرد. حتی زمانیکه مشغول جستجوی داده در دیسک هستیم کانال متصل به آن نمی‌تواند سرویس دیگری ارائه دهد و مشغول فرض می‌شود.

در این مدل مجموع میانگین تأخیر دیسک (latency) و زمان انتقال داده‌هـا، بـه عـنوان مـیزان تـقاضای سرویس کانال در نظر گرفته شده است. و میانگین زمان جستجو (Seek) در دیسک، به عنوان مقدار تقاضای سرویس دیسک فرض می‌شود. بنابراین تمامی مؤلفه‌های زمان سرویس I/O دقیقاً یکبار ارائه شده‌اند (اگر هر سه مؤلفهٔ سرویس یعنی تأخیر چرخش، زمان انتقال داده‌ها از کانال و زمان جستجوی داده در دیسک، در مرکز سرویس دیسک ارائه شود، مشتریان مدل دو بار سرویس انتقال و تأخیر را پشت سـرگذرانـده و در مقادیر کارآیی خطای جدی پدیدمی‌آید).

اشکال عمده‌ای در ارائه مؤلفه‌های چندگانه سیستم I/O وجود دارد. یعنی برخلاف سیستم‌های واقعی، هیچ مشتری به طور همزمان هم در دیسک و هم در کانال فعالیت نمی‌کند. بنابراین در مدل به طور بالقوه یک امکان توازی مجازی وجود دارد. چون منطقاً یک مرکز دیسک که توسط یک برنامه به منظور انتقال و تأخیر در حال سرویس است، در همان زمان برای جستجوی داده توسط برنامه دیگری می‌تواند مورد استفاده قرار بگیرد. در حالت کلی این عدم دقت مهم، اما در مورد این مسئله خاص، تأثیر توازی بالقوه در مـدل نـاچیز است. زیرا میزان بکارگیری دیسک‌ها، متعادل و کل تعداد دیسک‌ها بیشتر از میانگین سطح چـند بـرنامگی

است. بنابراین احتمال آنکه یک مشتری سرویسی از یک دیسک را نیاز داشته باشد که آن دیسک قبلاً توسط مشتری دیگری اشغال شده کاهش، و در نتیجه توازی خیلی کم می‌شود.

اندازه گیری داده‌های سیستم نشان می‌دهد که در طی مدت اندازه گیری، میانگین سطح چندبرنامگی، تغییر می‌کند. در صورت ایجاد این تغییرات، مدل برای هر سطح چند برنامگی موجود، یکبار دیگر ارزیابی می‌شود. در اینجا هدف از مدلسازی، ارزیابی اثرات تغییرهای پیشنهاد شدهٔ ذیل بر روی سیستم می‌باشد.

- جایگزینی شش دیسک با تکنولوژی IBM 3330 به جای هشت دیسک 2314 روی یک کانال:
اثر این تغییر در مدل روی تقاضاهای سرویس کانال و مراکز سرویس دیسک انعکاس می‌یابد. 3330 سریعتر از 2314، داده‌ها را جستجو و انتقال می‌دهد. همچنین دارای قابلیت (RPS) (Rotational position sensing) می‌باشد که این سبب می‌شود، طی مدت تأخیر چرخشی (rotate letency) ، دیسک از کانال جدا شده‌و فقط وقتی‌که بخش (Sector) دادهٔ موردنظر زیرهدخواندن ونوشتن قرارگرفت،اتصال‌برقرار شود.

- حافظهٔ توسعهٔ یافتهٔ (Extended Service Center) (ESC): این حافظه را به جای حافظهٔ موجود می‌گذاریم و تأثیر این تغییر در مدل، همراه با کاهش میزان تقاضای سرویس Cpu می‌باشد.
- اجرای سیستم عامل بهبود یافته: با این تغییر انتظار می‌رود سربار کاهش یابد. بنابراین شاهد کاهش میزان تقاضای سرویس Cpu به دلیل پردازش بهتر سیستم عامل خواهیم بود.

ترکیبات مختلف از اعمال بهبودها و اثرات آنها روی بهرهٔ Cpu، توسط مدل منعکس می‌شود. استفاده U_{Cpu} یا میزان بکارگیری پردازنده به عنوان مقیاس ارزیابی به دو جهت صورت می‌گیرد. یکی آنکه U_{Cpu} به سادگی با کندتر شدن پردازنده، افزایش می‌یابد و دیگر آنکه مقیاسی جهت ملاکهای دیگر ارزیابی نظیر نرخ خروجی و زمان پاسخ سیستم می‌باشد.

بهبود سیستم عامل به تنهایی ۵٪ در U_{Cpu} افزایش ایجاد می‌کند. در همین ارتباط با جایگزینی حافظهٔ ESC، میزان بکارگیری Cpu (U_{Cpu}) به ۲۵٪ می‌رسد. اگر همراه با بهبود سیستم عامل، دیسکها را نیز تعویض کنیم به طور مشابه میزان بکارگیری ۲۵٪ خواهد بود. اگر این دو اصلاح اعمال شوند، اندازه گیری سیستم نشان می‌دهد که U_{Cpu} تا حدود ۲۰٪ افزایش دارد. می‌توان تصور کرد میزان تقاضای سرویس Cpu به دلیل یک تغییر غیرمنتظره در ظرفیت کاری کاهش یافته است. بنابراین مدل یک طرح نسبتاً دقیقی از رفتار واقعی سیستم را ارائه می‌کند [۱۲].

این مثال نشان می‌دهد که مدلهای کامل و ساده، جهت پاسخ به سؤالات کارایی مورد نظر، مفید هستند. در یک سیستم کامپیوتری فقط جنبه‌هایی که جهت ارزیابی و در نظر گرفتن اصلاحات، مهم بوده ارائه می‌شوند. به عنوان مثال هیچ نشانهٔ واضحی‌از حافظه‌در مدل ماوجودنداشت. درنهایت می‌توان نتیجه گرفت سادگی یک فایدهٔ بزرگ مدلهای شبکهٔ صف می‌باشد.

۸-۳-۳ طرح ظرفیت‌کاری [۱۲]

در این بخش، هدف ارزیابی زمان پاسخ به ازای پیش‌بینی ۳٪ رشد در حجم ظرفیت کاری سیستم موجود می‌باشد. سیستم مورد نظر یک Amhal 470 با ۸mb حافظهٔ اصلی و ۱۶ کانال و ۴۰ دیسک می‌باشد. سیستم عامل IBM VMS و پایگاه دادهٔ IMS در سیستم فوق جهت پردازش‌های ظرفیت‌کاری تراکنشی مورد استفاده

قرار می‌گیرند.

IMS برای پردازش پیغام‌ها، حافظهٔ اصلی را به ۵ قسمت تقسیم نموده ، به هر کدام یک کار را اختصاص داده و زمانبندی به عهدهٔ IMS است. اگر بیش از ۵ تقاضا همزمان ارسال گردد، برای دسترسی به یک ناحیهٔ از حافظه، کاربر در صف منتظر می‌ماند. تراکنش‌های مختلفی در سیستم وجود دارند ولی از نظر حجم، با نرخهای یکسانی افزایش می‌یابند. بنابراین یک مدل یک کلاسی جهت ارزیابی آنها کافی است. اگر تراکنش‌های مختلف، دارای نرخ رشد جداگانه‌ای باشند یک مدل با چند کلاس مورد نیاز خواهد بود. با مشخصات فوق الذکر شکل مدل در ذیل (شکل ۸-۴) رسم شده است.

برای حافظه فقط یک صف داریم و به نظر می‌رسد برای مدل نیز کافی باشد. لازم به ذکر است که حل یک مدل شبکهٔ باز با یک صف اشباع شدهٔ حافظه، دقیقاً معادل حل یک مدل بستهٔ متناظر با آن است. در صورتیکه کلاس باز مشتری‌ها با یک کلاس بستهٔ چند برنامگی، جایگزین شود. سطح چند برنامگی مذکور همان ماکزیمم تعداد برنامه‌هایی که به طور همزمان باید فعال باشند. مدل مذکور قابل تفکیک بوده و براحتی ارزیابی می‌شود.

- نرخ ورود مشتری‌ها معادل نرخ ورود تراکنش اندازه گیری شده از سیستم می‌باشد.
- تقاضای سرویس در Cpu عبارتست از :

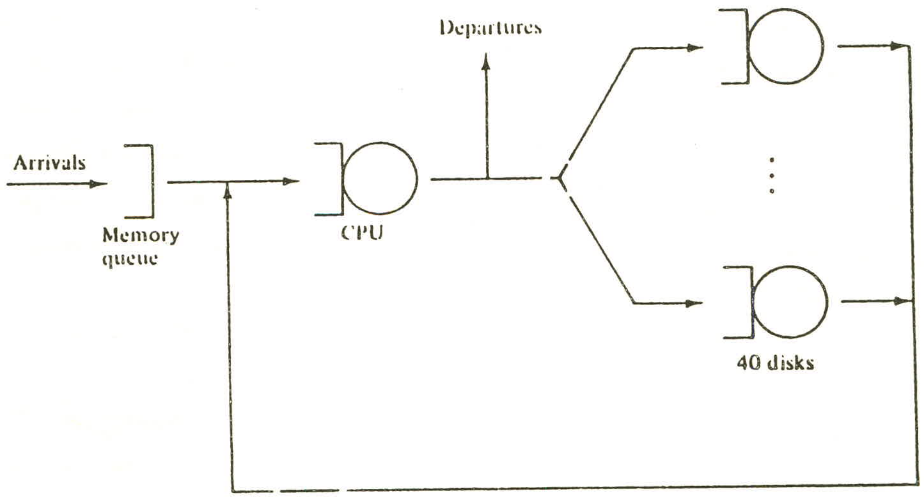
$$D_{cpu} = U_{cpu} \times \frac{T}{C}$$

که در آن U_{Cpu} میزان بکارگیری Cpu ، که طی مدت T (مدت اندازه گیری پارامترهای سیستم)، اندازه گیری شده است. همچنین C تعداد تراکنش‌هایی که طی این مدت تکمیل و اجرا شده‌اند.

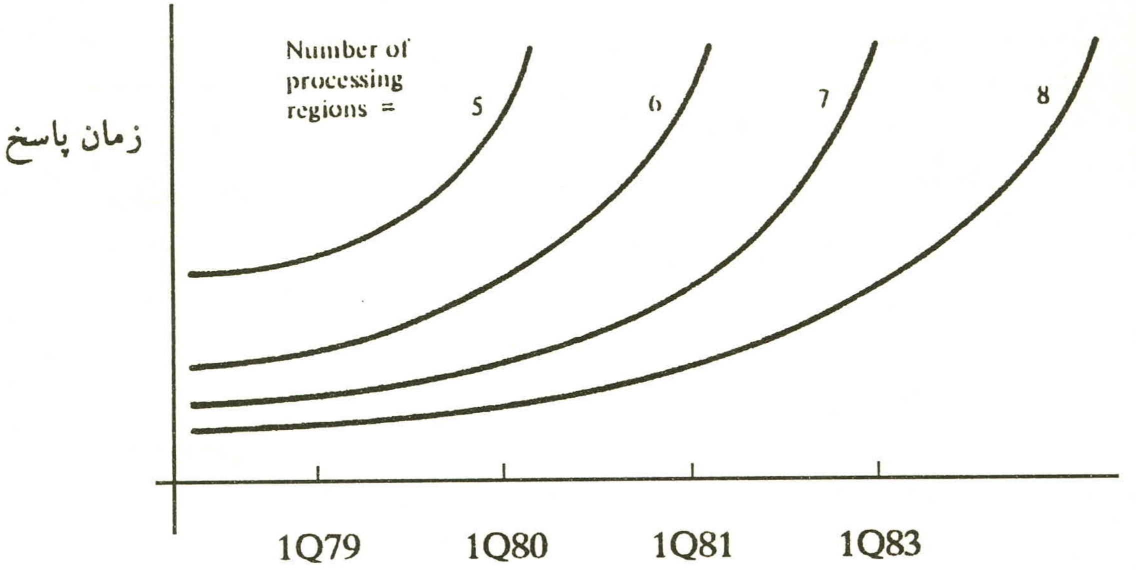
- تقاضای سرویس در هر مرکز سرویس K معادل است با :

$$D_k = U_k \cdot \frac{T}{C}$$

چون هم سربار و هم نیازهای سرویس جزء میزان تقاضای سرویس محسوب می‌شوند فرمول مذکور



شکل ۸-۴ مدل سیستم Amhal 470



شکل ۸-۵ زمانهای پاسخ پیشنهادی (سال و ربع (Quarter) سال)

را جهت محاسبه در نظر گرفته‌ایم [۱۲]. در مورد Cpu، هم زمان پردازش سیستم و هم کاربر و در مورددیسک‌ها تقاضای سرویس شامل زمان جستجو، زمان تأخیر چرخش، زمان انتقال داده و هر زمان تلف شده (به علت شرایط مسیر I/O) می‌باشد و فرمول‌های فوق براین اساس بیان شده‌اند. بدین ترتیب سربار کاملاً مفید واقع می‌شود. البته به شرطی که نسبت سربار به زمان پردازش مفید، حساس به اصلاحات مدل نباشد (بهبودهایی که در بالا ذکر کردیم).

مزیت عمدهٔ روش فوق سادگی آن جهت محاسبهٔ میزان تقاضای سرویس می‌باشد. به عنوان مثال لزومی ندارد هر مؤلفه زمان سرویس دیسک را تعیین کنیم. البته پیش بینی تغییرات در نسبت سربار به زمان سرویس مفید، بدون اطلاعات دقیق‌تر مدل نمی‌شود. با داشتن مجموعهٔ پارامترها، مدل را برای یافتن، زمان پاسخ ارزیابی می‌کنیم. شکل ۸-۵ گراف‌های زمان پاسخ مطرح شده را برای سالهای مختلف، برای ۴ اندازهٔ مختلف حافظه نشان می‌دهد که عبارتند از:

- پیکره‌بندی موجود جهت پشتیبانی ۵ ناحیهٔ پردازش پیغام کاربران
- توسعهٔ پیکره‌بندی برای ۸-۷-۶ ناحیهٔ پردازش پیغام کاربران

طبق بررسی فوق نتیجه می‌شود که با افزایش حافظه، سیستم برای ۲ سال دیگر نیز کفایت می‌کند.[۱۲]

۸-۴ روشهای حل

حل یک مدل شبکهٔ صف در حقیقت مجموعه‌ای از مقادیر ارزیابی می‌باشد که میانگین زمان رفتار مدل را بیان می‌کند. محاسبهٔ مقادیر مذکور برای شبکه‌های عمومی صف، پیچیده و پرهزینه می‌باشد.

برای تحلیل شبکه‌های صف تفکیک‌پذیر مختلف (مدلهای باز و بسته) روندهایی به ترتیب ذیل بیان می‌نماییم.

۸-۴-۱ تکنیک حل مدل باز

در مدلهای باز (مانند ظرفیت‌کاری از نوع تراکنش) یکی از مقادیر کلیدی خروجی یعنی نرخ خروجی سیستم، به عنوان یک ورودی داده می‌شود. دلیل این امر آن است که روشهای حل برای پایه این مدلها ساده می‌شود. حال فرمولهایی را برای محاسبهٔ مقادیر کارآیی ذکر می‌کنیم [۱۲].

- ظرفیت پردازش :

در یک مدل باز، نرخ خروجی و ورودی در حالت اشباع (λsat) همان ظرفیت پردازش می‌باشد. این مقدار توسط رابطهٔ زیر بیان می‌شود.

$$\lambda_{sat} = \frac{1}{\max D_k} = \frac{1}{D_{max}}$$

- نرخ خروجی :

طبق قانون جریان، اگر λمشتری در ثانیه وارد شبکهٔ صف شوند، نرخ خروجی سیستم باید λمشتری در ثانیه شود. همچنین اگر هر مشتری به طور متوسط نیاز به V_k ملاقات با دستگاه K داشته باشد، نرخ خروجی‌سیستم در دستگاه K باید $\lambda.V_k$در ثانیه باشد بنابراین:

$$X_k(\lambda) = \lambda.V_k$$

- میزان بکارگیری :

طبق قانون بکارگیری یک دستگاه معادل با نرخ خروجی آن ضرب در زمان سرویس است.

بنابراین خواهیم داشت:

$$U_k(\lambda) = X_k(\lambda).\ S_k = \lambda.D_k$$

در مورد مراکز تأخیر (تأخیر (Delay) به مراکزی گویند که سرویس آنها زمان اقامت در آنها و تأخیر صف نداریم) میزان بکارگیری به صورت میانگین تعداد مشتریان موجود در آنها می‌باشد.

- زمان اقامت (Residence):

کل زمان سپری شده توسط مشتری در مرکز K،$R_k(\lambda)$ گویند که شامل زمان صف و زمان دریافت سرویس می‌شود.

در مورد مراکز سرویس از نوع تأخیر، هیچ صفی وجود ندارد به طوریکه $R_k(\lambda)$ برابر حاصلضرب زمان سرویس در تعداد ملاقات می‌باشد.

$$R_k(\lambda) = V_k.S_k = D_k$$

اما برای مراکز صف ، R_k در واقع حاصل‌جمع کل زمان سپری شده در سرویس و کل زمان انتظار برای مشتریان دیگر جهت تکمیل سرویس می‌باشد بنابراین $R_k = V_k\ .\ S_k$ می‌شود.

مؤلفه دیگر زمان انتظار در صف، برای مشتریانی که از قبل در صف بوده‌اند، (در لحظهٔ ورود یک مشتری جدید) می‌باشد. حال $A_k(\lambda)$ را میانگین تعداد مشتری در صف، در لحظه ورود یک مشتری جدید، در نظر می‌گیریم. بنابراین مقدار کمّی طول صف عبارتست از:

$$L_q = V_k\,[A_k(\lambda)\ S_k]$$

فرض می‌کنیم زمان موردنظر برای تکمیل یک برنامه در سرویس، هنگامی که برنامه جدید وارد می‌شود، برابر با زمان سرویس دهی به برنامه باشد. بنابراین برای مراکز صف، زمان اقامت توسط رابطهٔ زیر مشخص می‌گردد:

$$R_k(\lambda)=V_k\,[S_k + S_k \cdot A_k(\lambda)] = D_k\,[1 + A_k(\lambda)]$$

در ساختار شبکه‌های تفکیک‌پذیر طول صفی که در لحظهٔ ورود توسط یک مشتری جدید، در مرکز K دیده شده، $A_k(\lambda)$، معادل با میانگین طول صف در زمان ورود مشتری مذکور، $Q_k(\lambda)$ است. دقت شود این مقدار به اندیس λ یعنی نرخ ورودی وابسته می‌باشد. بنابراین :

$$R_k(\lambda) = D_k\,[1 + Q_k(\lambda)]$$

با استفاده از قانون لیتل، رابطهٔ فوق را به صورت زیر می‌نویسیم:

$$R_k(\lambda) = D_k\,[1 + \lambda\, R_k(\lambda)]$$

رابطهٔ فوق نشان می‌دهد که وقتی $U_k(\lambda)$ به سمت صفر میل می‌کند داریم :

$$R_k(\lambda) \rightarrow D_k$$

همچنین وقتی $U_k(\lambda)$ به سمت یک میل می‌کند خواهیم داشت : $R_k(\lambda) \rightarrow \infty$

- طول صف :

با توجه به قانون لیتل داریم : $Q_k(\lambda) = \lambda \cdot R_k(\lambda)$

بنابراین:

$$Q_k(\lambda)= \begin{cases} U_k & \text{مراکز تأخیر} \\ \dfrac{U_k(\lambda)}{1 - U_k(\lambda)} & \text{مراکز صف} \end{cases}$$

- زمان پاسخ سیستم:

در حقیقت زمان پاسخ سیستم، مجموع زمانهای اقامت در همهٔ مراکز می‌باشد یعنی:

$$R(K) = \sum_{k=1}^{K} R_k(\lambda)(\lambda)$$

- میانگین تعداد مشتری در سیستم:

با استفاده از قانون لیتل و با جمع‌کردن طولهای صف در همهٔ مراکز داریم:

$$Q(\lambda) = \lambda \cdot R(\lambda) = \sum_{k=1}^{K} Q_k(\lambda)$$

الگوریتم ۱ فرمولهای مذکور را خلاصه‌بندی نموده است.

الگوریتم ۱ :

ظرفیت پردازش :	$\lambda_{sat} = \dfrac{1}{D_{max}}$
نرخ خروجی :	$X(\lambda) = \lambda$

بکارگیری (utilization):

$$U(\lambda) = \lambda \, . \, D_k$$

زمان اقامت :

$$R_k(\lambda) = \begin{cases} D_k & \text{مراکز تأخیر} \\ \dfrac{D_k}{1 - U_k(\lambda)} & \text{مراکز صف} \end{cases}$$

طول صف :

$$Q_k(\lambda) = \lambda . R_k(\lambda) = \begin{cases} U_k(\lambda) & \text{مراکز تأخیر} \\ \dfrac{U_k(\lambda)}{1 - U_k(\lambda)} & \text{مراکز صف} \end{cases}$$

زمان پاسخ سیستم :

$$R_k(\lambda) = \sum_{k=۱}^{K} R_k(\lambda)$$

میانگین تعداد افراد در سیستم :

$$Q(\lambda) = \lambda \, . \, R(\lambda) = \sum_{k=۱}^{K} Q_k(\lambda)$$

مثال ۸-۱ (مدل باز)

شکل ۸-۶ یک مثال ساده از مدل باز، با سه مرکز سرویس می‌باشد. مشخصات مدل عبارتند از:

$V_{cpu} = ۱۲۱$ | $\lambda = ۰/۳$ job در ثانیه | $V_{disk2} = ۵۰$

$S_{cpu} = ۰/۰۰۵$ | $V_{disk1} = ۷۰$ | $S_{disk2} = ۰/۰۲۷$

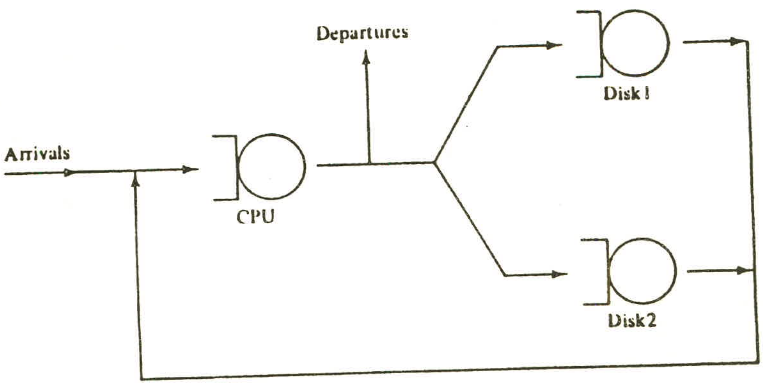
$D_{cpu} = ۰/۶۰۵$ | $S_{disk1} = ۰/۰۳$ | $D_{disk2} = ۱/۳۵$

$D_{disk1} = ۲/۱$

حل:

محاسبات مقادیر کارآیی طبق الگوریتم ۱ صورت گرفته و خروجی‌های مدل عبارتند از:

$$\lambda_{sat} = \frac{۱}{D_{max}} = \frac{۱}{۲/۱} = ۳۶/۳$$ تعداد ملاقات در ثانیه



شکل ۸-۶ ساختار مدل باز

$$X_{cpu}(۰/۳)=\lambda.\ V_{cpu}=(۰/۳)\ (۰/۶۰۵)=۰/۱۸۲$$

$$R_{cpu}\ (۰/۳)\ =\frac{D_{cpu}}{۱-U_{cpu}\ ۰/۳}=\frac{۰/۶۰۵}{۰/۸۱۸}=۰/۷۴۰$$

$$Q(۰/۳)=\lambda.R(\lambda)=(۰/۳)\ (۸/۶۸۵)=۲/۶۰۶$$ برنامه

$$Q_{cpu}\ (۰/۳)\ =\frac{U_{cpu}\ (۰/۳)}{۱-U_{cpu}\ (۰/۳)}=۰/۲۲۲$$ برنامه

$$R(۰/۳)=R_{cpu}(۰/۳)+R_{Disk1}(۰/۳)+_{Disk2}(۰/۳)=۰/۷۴+۵/۶۷۶+۲/۲۶۹=۸/۶۸۵$$ ثانیه

۸-٤-۲ تکنیک حل مدل بسته

روشی که برای ارزیابی شبکه‌های صفی بسته (مشتریانی از نوع batch یا terminal) مورد استفاده قرار می‌گیرد بنام "تحلیل میانی" یا(MVA) آنالیز مقدار میانگین (Meam Value Analysis) مشهور است.

روش MVA بر سه معادله استوار است که عبارتند از:

-ابتدا قانون لیتل را برای کل شبکهٔ صف به کار می‌بریم یعنی:

$$X(N)\ =\frac{N}{Z+\sum_{k=۱}^{K}R_k\ (N)} \qquad (۱)$$

که در رابطهٔ قوق N تعداد مشتری موجود در شبکه و X(N) نرخ خروجی سیستم و $R_k(N)$ زمان اقامت در مرکز K می‌باشد (اگر نوع مشتریان دسته‌ای باشد Z=0 است).

دقت شود که نرخ خروجی سیستم هنگامی که زمانهای اقامت دستگاه‌های سرویس دهنده ($R_k(N)$) داده شده، از روی دادهٔ ورودی محاسبه می‌شود.

-قانون لیتل را برای هر مرکز سرویس به طور جداگانه به کار می‌بریم یعنی:

$$Q_k(N)\ =\ X(N)\ .\ R_k(N) \qquad (۲)$$

البته اگر بخواهیم قانون لیتل را برای محاسبهٔ طول صف استفاده کنیم باید زمانهای اقامت مشخص باشند.

-معادلات زمان اقامت مرکز سرویس را به کار می‌بریم بنابراین:

$$R_k(N)=\begin{cases} D_k & \text{مراکز تأخیر} \\ [1\ +\ A_k(N)]\ D_k & \text{مراکز صف} \end{cases} \qquad (۳)$$

که در آن $A_k(N)$، متوسط تعداد مشتری در مرکز K، در هنگام ورود یک مشتری جدید می‌باشد.

توجه شود که مانند شبکه‌های باز، اینجا نیز کلید محاسبات مقادیر کارآیی، مجموعهٔ $A_k(N)$ است. در صورت وجود این مجموعه، $R_k(N)$ توسط X(N) و $Q_k(N)$ می‌تواند محاسبه شود.

در مورد شبکه‌های باز می‌توانستیم میانگین طول صف در زمان ورود مشتری جدید $Q_k(N)$ را جایگزین طول صف لحظهٔ ورود $A_k(N)$ نمائیم. ولی در مورد شبکه‌های بسته این عمل غیرممکن است. چون $A_k(N)$ در شبکهٔ بسته معادل $Q_k(N)$ نیست. به عنوان مثال شبکه‌ای با ۲ مرکز سرویس و یک مشتری در نظر بگیرید بطوریکه تقاضای سرویس در هر مرکز ۱ ثانیه باشد. چون فقط یک مشتری داریم، میانگین طول صف در زمان ورود در مراکز مذکور، همان میزان بکارگیری مراکز می‌باشد بنابراین :

$$Q_۱(۱) = Q_۲(۱) = \frac{۱}{۲}$$

طولهای صف لحظهٔ ورود یعنی $A_۱(۱)$ و $A_۲(۱)$ هر دو صفر هستند. چون با وجود فقط یک مشتری در شبکه، امکان ندارد که مشتری دیگری پشت سر این مشتری در صف موجود باشد. در حالت کلی وجه تمایز آن است که طولهای صف در لحظهٔ ورود براساس تعداد مشتری در حال ورود در آن مرکز محاسبه می‌شود. به همین دلیل این مشتریان خودشان در صف نمی‌توانند باشند. در حالیکه میانگین طول صفها در لحظهٔ ورود مشتری جدید، به طور تصادفی در زمانهای انتخابی (که همهٔ مشتری‌ها بالقوه می‌توانند در صف باشند) محاسبه می‌گردد.

همانطور که قبلاً اشاره کردیم برای ارزیابی مدل نیاز به محاسبهٔ $A_k(N)$ در همان ابتدا داریم. برای این منظور دو روش وجود دارد:

۱- تکنیک دقیق   ۲- تکنیک تقریبی

تقاوت این دو روش بیشتر به چگونگی برقراری ارتباط حل و مدل دارد و کمتر به خود سیستم مرتبط است. دقت حل وابستگی زیادی به دقت پارامتربندی مدل دارد و بستگی به انتخاب هر یک از روشهای فوق ندارد. حال هر دو تکنیک را به ترتیب بررسی می‌کنیم.

۸-۴-۲-۱ تکنیک حل دقیق

این تکنیک در مورد روش MVA به دو دلیل مهم می‌باشد. اول آنکه پایهٔ تکنیکهای حل تقریبی می‌باشد و دوم آنکه هیچ حد معینی برای بی دقتی تکنیکهای تقریبی وجود ندارد. به طور نمونه برای درصد کمی از حل‌های صحیح، نسبتاً این تکنیکها دقت دارند ولی تضمینی وجود ندارد که در هر شرایط خاصی نتایج درست باشد.

تکنیکهای حل دقیق محاسبهٔ طول صف در لحظهٔ ورود $A_k(N)$ را به طور دقیق با بکاربردن معادلات (۱) و (۳) بخش قبلی محاسبه می‌کنند.

یک ویژگی شبکه‌های تفکیک‌پذیر بسته که باعث استفاده از این روش شده، آن است که $A_k(N)$ دارای شکل خاصی می‌باشد یعنی :

$$A_k(N) = Q_k(N-1) \qquad (۴)$$

به عبارت دیگر هنگامی که N مشتری در شبکه موجود، طول صف دیده شده توسط مشتری جدید الورود برابر میانگین طول صف در لحظهٔ ورود با یک مشتری کمتر در شبکه خواهد بود. دلیل این امر آن است که در لحظه‌ای که یک مشتری وارد یک مرکز می‌شود قطعاً نمی‌تواند قبلاً در آن صف باشد. پس فقط $N-1$ مشتری موجود ممکن است همراه با مشتری جدید در سیستم حضور داشته باشند.

روش حل دقیق، که در الگوریتم ۲ نشان داده شده، در حقیقت کاربرد مکرر معادلات (۱) - (۴) را ارائه می‌کند. معادلات مذکور نرخ خروجی سیستم، زمان‌های اقامت دستگاه (Device) هر مرکز و میانگین طول صف هر مرکز در لحظهٔ ورود مشتری جدید با وجود n مشتری را از روی میانگین طول صف در لحظهٔ ورود با $n-1$ مشتری (که داده شده) محاسبه می‌کنند.

عملیات تکرار از آن جایی شروع می‌شود که همهٔ طولهای صف صفر بوده و هیچ مشتری در شبکه وجود ندارد. محاسبات کمّی برای یک مشتری لازم و سپس از این نتایج برای تعداد ۲ مشتری استفاده می‌کنیم. چون میانگین طول صف با یک مشتری، معادل میانگین طول صف لحظهٔ ورود برای ۲ مشتری می‌باشد. با تکرارهای صحیح معادلات مذکور به حل شبکه برای تعداد دلخواه N می‌رسیم.

الگوریتم ۲ : تکنیک حل دقیق MVA

برای $k=1$ تا $k=K$ قرار دهید: $Q_k = 0$

برای $n=1$ تا $n=N$ مراحل زیر را انجام دهید:

شروع: برای $k=1$ تا K قرار دهید،

$$R_k = \begin{cases} D_k & \text{مراکز تأخیر} \\ D_k\,(1 + Q_k) & \text{مراکز صف} \end{cases}$$

$$X = \frac{N}{Z + \sum_{k=1}^{K} R_k}$$

برای $k=1$ تا K قرار دهید $Q_k = X.R_k$

پایان.

برای حل کامل شبکه با N مشتری، باید معادلات (۱) - (۴) را N بار بکار ببریم. چون در هر بار، برای هر مرکز K، حلقه‌های تکرار انجام می‌شوند. بنابراین هزینهٔ محاسبات به صورت حاصلضرب N در K رشد می‌کند. فضای مورد نیاز K محل می‌باشد. چون مقادیر کارآیی برای شبکه در حالت n مشتری، پس از استفاده جهت محاسبهٔ مقادیر کارآیی در حالت $n+1$ مشتری، نگهداری نمی‌شود. بنابراین هیچ هزینهٔ اضافی برای بدست آوردن حل‌های میانی وجود ندارد (از یک مشتری تا $N-1$ مشتری حالات میانی است). این یک ویژگی مهم برای این روش محسوب می‌شود. پس از پایان اجرای الگوریتم ۲، مقادیر Q_K , X , R_K برای هر N، بلافاصله در دسترس می‌باشد. با استفاده از قانون لیتل خروجیهای دیگر مدل بدست می‌آید که عبارتند از:

X : نرخ خروجی سیستم

$N-XZ$: میانگین تعداد مشتری در سیستم

$\frac{N}{X} - Z$: زمان پاسخ سیستم

$X.D_k$: بکارگیری دستگاه K

R_k : زمان اقامت در دستگاه K

$X.V_k$: نرخ خروجی دستگاه K

Q_k : طول صف دستگاه K

مثال ۸-۲ (حل دقیق مدل بسته):

جدول ۲-۸ محاسبات مدل قبلی را نشان می‌دهد که به جای برنامه‌های تراکنش برنامه‌های ترمینال جایگزین شده‌اند.

جدول ۸-۲ محاسبات آنالیز مقدار میانگین دقیق

////	K	n=۰	n=۱	n=۲	n=۳
R_k	cpu	–	۰/۶۰۵	۰/۶۲۴	۰/۶۴۴
	دیسک ۱	–	۲/۳۳۱	۲/۳۳۱	۲/۶۰۵
	دیسک ۲	–	۱/۴۴۶	۱/۴۴۶	۱/۵۵۱
X		–	۰/۰۵۲۵	۰/۱۰۳۱	۰/۱۵۱۵
Q_k	cpu	۰	۰/۰۳۱۸	۰/۰۶۴۳	۰/۰۹۷۶
	دیسک ۱	۰	۰/۱۱۰۲	۰/۲۴۰۳	۰/۳۹۴۷
	دیسک ۲	۰	۰/۰۷۰۸	۰/۱۴۹۰	۰/۲۳۵۰

سه مرکز با تقاضای سرویس D_{cpu} = ۰/۶۰۵ و D_{Disk1} = ۲/۱ و D_{Disk2} = ۱/۳۵ ثانیه وجود دارد. تعداد مشتریان ترمینال N=۳ می‌باشد. میانگین زمان تفکر Z=۱۵ ثانیه می‌باشد.

الگوریتم با حل دقیق برای شبکه با N=۰ مشتری آغاز می‌شود و $R_k(N)$ ، $X(N)$ ، $Q_k(N)$ را برای هر تعداد دلخواه N محاسبه نموده تا به N=۳ برسد.

با بررسی جدول فوق متوجه می‌شویم که مجموع طولهای صف در هر سه مرکز با تعداد کل مشتری‌ها یکسان نیست. دلیل آن است که، با ظرفیت کاری از نوع ترمینال مواجه هستیم که تعدادی از مشتریان در حال تفکر می‌باشند. (الگوریتم ۲ در این مورد اجازه می‌دهد که زمان تفکر در یکی از معادلاتش وارد شود).

اگر از کل جمعیت مشتریان N، میانگین تعداد مشتری در سیستم $(Q = N - XZ)$ را کم کنیم، میانگین تعداد مشتریانی که تفکر می‌کنند (XZ) بدست می‌آید. این مقدار برای نوع دسته‌ای معادل صفر است.

خروجی‌های مدل برای N=۳ عبارتند از :

حل:

$$X(۳) = ۰/۱۵۲$$

$$R(۳) = \frac{۳}{۰/۱۵۲} - ۱۵ = ۴/۷۴$$

$$Q(۳) = N - X(۳).Z = ۳ - (۰/۱۵۲)(۱۵) = ۰/۷۲$$

$Q_{cpu}(۳) = ۰/۰۹۸$

$X_{cpu}(۳) = X(۳).V_{cpu} = (۰/۱۵۲)\ (۱۲۱) = ۱۸/۳۹$

$U_{cpu}(۳) = X(۳).\ D_{cpu} = (۰/۱۵۲)\ (۰/۶۰۵) = ۰/۰۹۲$

$R_{cpu}(۳) = ۰/۶۴$

۸-٤-۲-۲ روش حل تقریبی

کلید حل دقیق به روش MVA، معادلهٔ (۴) می‌باشد که طول صف لحظهٔ ورود را برای n مشتری، براساس میانگین طول صف در زمان ورود با n−1 مشتری، محاسبه می‌کند.

در معادلهٔ (۴) تقریب زیر را جایگزین می‌کنیم. با تکرار الگوریتم تعدادی تابع h مناسب بدست می‌آید که اساس روش حل تقریبی است.

$$A_k(N) \approx h[Q_k(N)]$$

تابع h به مقادیر دیگری از $Q_k(N)$ نیز بستگی دارد. با تقریب زدن، بخشی از وابستگی به N را نشان می‌دهیم. بنابراین برای سادگی از این نماد استفاده کرده و می‌گوییم که نیاز اصلی ما، مقادیر $Q_k(N)$ می‌باشد. دقت الگوریتم به دقت تابع h مورد استفاده بستگی دارد. روش مذکور در الگوریتم ۳ بیان شده است. به سادگی دیده می‌شود که زمان و فضای مورد نیاز این الگوریتم، به تعداد مراکز وابسته بوده و مستقل از جمعیت مشتریان مدل می‌باشد. (البته تعداد تکرار مورد نیاز جهت همگرایی امکان دارد از جمعیت مدل متأثر گردد). این مطلب یک بهبود مهم برای روش MVA دقیق محسوب می‌شود.

الگوریتم ۳: تکنیک حل تقریبی MVA

۱- برای همهٔ مراکز K قرار دهید: $Q_k(N) = \frac{N}{K}$

۲- برای همهٔ مراکز K تقریب بزنید: $A_k(N) = h[Q_k(N)]$

۳- از معادلات (۱) و (۲) و (۳) برای محاسبهٔ یک مجموعهٔ جدید $Q_k(N)$ استفاده کنید.

۴- اگر $Q_k(N)$ بدست آمده از (۳) در محدودهٔ تلورانس (مثلاً ۰/۱٪) ورودی‌های مرحلهٔ (۲) نباشد به مرحلهٔ (۲) رفته و با $Q_k(N)$ جدید تکرار می‌کنیم.

نکتهٔ مورد بحث این تکنیک سریع، تابع h می‌باشد. تابع h معینی که برای همهٔ شبکه‌های تفکیک‌پذیر، دارای دقت کافی باشد، وجود ندارد. به طور مثال، یک تقریب قابل توجه و ساده به صورت زیر می‌باشد:

$$A_k(N) = Q_k(N-۱) \approx h[Q_k(N)] \equiv \frac{N-۱}{N} \cdot Q_k(N) \qquad (۵)$$

معادلهٔ (۵) طول صف لحظهٔ ورود را با تقریب مقدار دقیق طول صف با یک مشتری کمتر، تخمین می‌زند. این تقریب براساس این فرض است که نسبت‌های $\frac{Q_k(N)}{N}$ و $\frac{Q_k(N-۱)}{N-۱}$ برای همهٔ مقادیر K با یکدیگر برابر می‌باشند. به عبارت دیگر مقدار کاهش طول هر صف به ازای یک مشتری، معادل مقدار افزایش طول هر صف به ازای یک مشتری اضافه، خواهد بود.

در حالت کلی این فرض کاملاً درست است. در موارد خاص برای مقادیر بزرگ N به طور مجانبی درست و برای مدل یک مشتری کاملاً درست نیست (چون این فرض تخمین می‌زند که طول لحظهٔ ورود صفر است).

تجربه نشان داده که این فرض نتایج خوب و قابل توجهی، برای جمعیت بین ۱ تا N دارد. چون فرآیند تحلیل سیستم‌های کامپیوتری دارای تفاوتهای ذاتی می‌باشد (مثلاً دقت مقادیر پارامترها)، برخی خطاها در محدودهٔ این تفاوتها جای گرفته، بنابراین روش تقریب MVA به عنوان یک حل عمومی قابل قبول خواهد بود [۱۲].

مثال ۸-۳ (حل تقریبی مدل بسته)

در مورد مثال قبلی به جای حل دقیق از روش حل تقریبی استفاده کرده و تقریبهای مناسبی برای طول صف مراکز بدست می‌آید که در جدول ذیل آمده است. ملاک خاتمهٔ تکرار مراحل، اختلاف کمتر از ۰/۰۰۱ در طول صف می‌باشد. جهت مقایسه، حل دقیق نیز در جدول ۳-۸ نشان داده شده است.

مدل از نوع ترمینال بوده و ابتدا با توزیع یکسان مشتری‌ها در سه مرکز کار را شروع می‌کنیم. همانطور که مراحل تکرار می‌شوند، مشتریان به تدریج ناپدید می‌شوند. در پایان تکرار الگوریتم، اختلاف بین کل جمعیت مشتریان و مجموع طول صفها در مراکز، معرف میانگین تعداد مشتریانی است که در حال فکر کردن می‌باشند.

جدول ۸-۳ محاسبات آنالیز مقدار میانگین تقریبی

تکرار	Q_{cpu}	Q_{Disk1}	Q_{Disk2}	X	R
۰	۱/۰	۱/۰	۱/۰	-	-
۱	۰/۱۳۹۰	۰/۴۸۲۶	۰/۳۱۰۲	۰/۱۳۷۹	۶/۷۵۸۳
۲	۰/۹۸۸	۰/۴۱۵۰	۰/۲۴۳۶	۰/۱۴۹۵	۵/۰۶۵۹
۳	۰/۰۹۷۲	۰/۴۰۴۳	۰/۲۳۶۶	۰/۱۵۰۸	۴/۸۹۵۰
۴	۰/۰۹۷۲	۰/۴۰۲۴	۰/۲۳۵۹	۰/۱۵۱۰	۴/۸۷۳۲
۵	۰/۰۹۷۳	۰/۴۰۲۱	۰/۲۳۵۹	۰/۱۵۱۰	۴/۸۷۰۰
حل دقیق	۰/۰۹۷۶	۰/۳۹۴۷	۰/۲۳۵۰	۰/۱۵۱۵	۴/۸۰۲۰

۸-۵ مدلهای شبکه صف تفکیک پذیر

با اعمال محدودیتهایی روی رفتار مراکز سرویس و مشتریان، زیرمجموعه‌ای از مدلهای شبکهٔ صف حاصل می‌شود که به آنها مدلهای شبکهٔ صف تفکیک‌پذیر (Separable) گویند. اصطلاح تفکیک‌پذیر به این دلیل است که هر مرکز سرویس به طور جداگانه از بقیهٔ مراکز، مورد ارزیابی قرار می‌گیرد. سپس حل نهایی سیستم از ترکیب حل‌های مذکور بدست می‌آید. پنج فرض در مورد رفتار یک مدل اگر برقرار باشد، آن مدل حتماً تفکیک‌پذیر خواهد بود این فرضها عبارتند از:

- **تعادل جریان مرکز سرویس:** در هر مرکز سرویس، تعداد ورودی‌ها به مرکز، برابر تعداد خروجی‌های آن مرکز می‌باشد.

- **رفتار یک گام:** در سیستم دو برنامه دقیقاً در یک زمان "تغییر حالت" (منظور پایان پردازش در یک مرکز یا ورود به سیستم) نمی‌دهند. سیستم‌های واقعی، غالباً رفتار یک گام (One step behavior) را به طور قطعی نشان می‌دهند.

سه فرض دیگر "فرضهای همگن" نامیده می‌شوند. که در هر مورد فرض می‌کنیم تعدادی از کمیتها صرفنظر از جای مشتریان در شبکه، همگن و یکنواخت می‌باشند.

- **یکنواختی مسیر حرکت:** رفتار مشتری در مدل توسط تقاضای سرویس آن مشخص می‌شود. بطور دقیق‌تر در حقیقت شامل الگوهای مسیریابی برنامه‌ها، که همان الگوی مراکز ملاقات شده می‌باشد تعریف می‌گردد. یکنواختی مسیر حرکت هنگامی برقرار می‌شود که برای تمامی jها و kها، مقدار زمانی که یک برنامه‌پس از دریافت کامل سرویس در مرکز j برای حرکت (به طور مستقیم) به مرکز k صرف می‌کند، از طول صف در هر مرکز مستقل باشد. یک نکتهٔ جالب در مدلهای تفکیک‌پذیر آن است که الگوی مسیر حرکت برنامه‌ها، به مقادیر ارزیابی ارتباطی ندارد و بنابراین از آنها چشم‌پوشی می‌شود.

- **یکنواختی دستگاه سرویس دهنده:** نرخ اجرایی برنامه‌ها در یک مرکز سرویس، ممکن است با تعداد برنامه‌ها در آن مرکز تغییر کند. اما به تعداد مشتریان شبکه و یا جایابی آنها در درون شبکه بستگی ندارد.

- **یکنواختی ورودی‌ها از خارج:** زمانهای ورود مشتریان از خارج شبکه به تعداد یا جایابی مشتریان درون شبکه بستگی ندارد.

فرضهای فوق برای شبکه‌های تفکیک‌پذیر کافی هستند. اما الگوریتم‌های حل که ارائه داده‌ایم نیاز به فرض اضافه‌ای دارند. در حقیقت شکل قوی‌تری از فرض یکنواختی دستگاه سرویس دهنده می‌باشد که عبارتست از:

- **یکنواختی زمان سرویس:** در زمان مشغول بودن مرکز سرویس، نرخ اجرای برنامه‌ها باید مستقل از تعداد مشتریان آن مرکز باشد. همچنین به تعداد یا جایابی مشتریان درون شبکه نیز وابسته نباشد.

فرض یکنواختی دستگاه سرویس دهنده، اجازه می‌دهد که نرخ اجرا در یک مرکز با طول صف آن مرکز تغییر نماید. مراکزی که این ویژگی را دارند "وابسته به بار" نامیده می‌شوند. یک مرکز تأخیر، نمونهٔ ساده‌ای از مرکز وابسته به بار می‌باشد. چون نرخ اجرا متناسب با تعداد مشتری در مرکز افزایش می‌یابد.

یکنواختی زمان سرویس موجب می‌شود که نرخ اجرا از طول صف مستقل باشد که به این مراکز "مستقل از بار" گویند. مراکز صفی که به آنها اشاره کردیم، وابسته به بار می‌باشند. گونه‌های خاصی از الگوریتم MVA که بیان شد، برای شبکه‌های شامل مراکز مستقل از بار و مراکز تأخیر قابل استفاده می‌باشد.

فرضهای فوق جهت اثبات ریاضی حل‌های دقیق الگوریتم ۲ ضروری می‌باشند. ولی در عمل برای مدلهای تفکیک‌پذیر لزومی ندارد برقرار باشند. تجربه نشان داده که دقت مدلهای شبکهٔ صف در صورت عدم رعایت این فرضها نیز فوق العاده قوی می‌باشد. در حالی که هیچ سیستم کامپیوتری واقعی، فرضهای یکنواختی را ارضاء نمی‌کند. تخطی از این فرضها بی دقتی عمده‌ای در مدلسازی پدید نمی‌آورد.

مسائلی که از نظر اعتبار نتایج یک مدل، باآنها درگیر هستیم عمدتاً ناشی از عدم دقت کافی در تعیین مشخصات سیستم در مدل می‌باشد. که در نهایت منجر به بی‌دقتی در مقادیر پارامترهای تقاضای سرویس و شدت ظرفیت کاری خواهد شد. مسئله مهم محدودیتهایی است که توسط فرضها، جهت قابلیت تفکیک‌پذیری روی ساختار مدل ایجاد می‌شوند. این امر باعث می‌شود جنبه‌هایی از سیستم کامپیوتری که

جهت بررسی ارزیابی مهم هستند ارائه نشوند (به عنوان مثال، مدلسازی مسائل حافظه یا اولویت در زمانبندی). در این موارد مدلهایی مورد نظر هستند که دارای ساختار ساده و در نتیجه ارزیابی ساده‌ای باشند. این ساختار در حقیقت همان شبکه‌های تفکیک‌پذیر است. این مدلها ابزار مفیدی جهت مدلسازی سیستم‌های کامپیوتری و اساس ساخت بلوکهایی که مدلهای مفصل‌تر را تشکیل می‌دهند می‌باشند.