



دانشگاه علم و صنعت ایران

دانشکده مهندسی کامپیوتر

عنوان درس:

ارزیابی کارایی سیستم‌های کامپیوتری Performance Evaluation of Computer Systems (PECS)

جلسه ۱۲: قوانین عملیاتی

مدرس:

محمد عبداللهی ازگمی

(Mohammad Abdollahi Azgomi)

azgomi@iust.ac.ir

فهرست مطالب

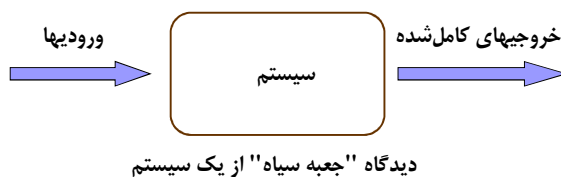
- مقدمه
- قانون بهره‌وری (utilization law)
- قانون لیتل (Little's law)
- قانون زمان پاسخ (response time law)
- قانون جریان اجباری (forced flow law)

مقدمه

- قوانین عملیاتی (operational laws) معادلات ساده‌ای هستند که به‌عنوان یک نمایش مجرد (abstract representation) یا مدل رفتار میانگین (model of the average behaviour) تقریباً برای همه سیستم‌ها قابل استفاده‌اند.
- یکی از مزیت‌های قوانین عملیاتی آن است که خیلی عمومی هستند و هیچگونه فرضی را در باره رفتار متغیرهای تصادفی توصیف‌کننده سیستم اعمال نمی‌کنند.
- این در مقابل تحلیل مارکوفی که متکی بر مفروضات خیلی قوی در باره تابع توزیع متغیرهای تصادفی مورد استفاده است.
- مزیت دیگر، قوانین سادگی آنها است.
- این بدان معنی است که قابلیت به‌کارگیری سریع و آسان را دارند.

مقدمه

- دلیل آنکه به این قوانین "عملیاتی" (operational) گفته می‌شود آن است که به‌طور مستقیم با اندازه‌گیری بدست می‌آیند.
- برخی فرضهای عملیاتی مورد استفاده قرار می‌گیرند، که با اندازه‌گیری قابل درستی‌یابی هستند.
- برای مثال: تعداد ورودیها = تعداد خروجیهای کامل شده
- کمیت‌های عملیاتی در طی یک مدت مشاهده محدود، به‌طور مستقیم اندازه‌گیری می‌شوند.
- در این قوانین دیدگاه "جعبه سیاه" از یک سیستم مورد مشاهده وجود دارد که در شکل زیر نشان داده شده است:



نمادها و پارامترهای مورد استفاده

- دوره (مدت زمان) مشاهده..... $T = \text{observation period}$
- تعداد ورودیها $A = \text{number of arrivals}$
- نرخ ورود $\lambda = \text{arrival rate} = A/T$
- تعداد خروجیهای کامل شده $C = \text{number of completions}$
- توان عملیاتی $X = \text{throughput} = C/T$
- میانگین کارها در سیستم $N \text{ or } L = \text{mean number in system}$
- مدت زمان مشغول بودن $B = \text{busy time}$
- بهره‌وری $\rho = \text{utilization} = B/T$
- میانگین زمان سرویس $S = \text{mean service time} = B/C$
- میانگین زمان پاسخ (زمان صرفشده) $R \text{ or } W = \text{mean time in system (response time)}$
- زمان تفکر یک کاربر ترمینال $Z = \text{think time of a terminal user}$
- تقاضای سرویس $D = \text{service demand}$
- نسبت ملاقات $V = \text{visit ratio}$

قانون بهره‌وری

- بهره‌وری (ρ) عبارت است از کسری از زمان که سیستم مشغول (busy) بوده است.
□ ρ همیشه مقداری بین صفر و یک دارد.

$$\rho = \frac{B}{T} = \frac{C}{T} \times \frac{B}{C} = XS$$

بهره‌وری یک سیستم (یا منبع) مساوی توان عملیاتی ضرب در زمان سرویس است.

مثال اول از قانون بهره‌وری

■ یک دیسک ۵۰ درخواست در ثانیه را سرویس می‌دهد. هر درخواست نیازمند 0.005 ثانیه سرویس است. در این صورت:

■ بهره‌وری دیسک چقدر است؟

$$\rho = 50 \times 0.005 = 0.25 \text{ (25\%)}$$

■ حداکثر نرخ سرویس (توان عملیاتی) امکان‌پذیر چقدر است؟

$$\rho = 1 = (0.005) X$$

$$X = 200 \text{ requests/sec}$$

مثال دوم از قانون بهره‌وری

■ یک مسیریاب (router) در هر ثانیه ۱۰۰ بسته را مسیریابی نموده و به پیوند (link) بعدی ارسال می‌کند. زمان ارسال (یعنی زمانی که طول می‌کشد تا بسته مسیریابی شده و به پیوند بعدی فرستاده شود) به طور میانگین یک میلی ثانیه است.

□ بهره‌وری پیوند چقدر است؟

$$X = 100 \text{ packets/sec}$$

■ توان عملیاتی پیوند:

$$S = 0.001 \text{ sec}$$

■ زمان سرویس:

$$\rho = XS = 0.1 \text{ (10\%)}$$

■ آنگاه بهره‌وری:

□ ظرفیت پیوند (link capacity) چقدر است؟

$$\rho = XS \Rightarrow 1 = 0.001X \Rightarrow X = 1000$$

قانون لیتل

■ قانون لیتل (Little's Law) قانونی است که به هر سیستمی اعمال می‌شود، بدون توجه به اینکه فرآیند ورودیها به سیستم چیست یا آنکه در داخل سیستم چه اتفاقی می‌افتد.

■ فرض مهم برای قانون لیتل، تعادل جریان کاری (Job Flow Balanced) است: تعداد ورودیها مساوی تعداد خروجیهای کامل شده است.

■ برای این منظور باید داشته باشیم:

□ هیچ کار جدیدی در سیستم تولید نمی‌شود.

□ کارها هیچ وقت در سیستم از بین برده نمی‌شوند.

■ اگر کارها در سیستم از بین می‌روند (مثلاً به دلیل ظرفیت محدود)، این قانون با نرخ ورود تنظیم شده (که بخش از دست رفته از کل ورودیها کم می‌شوند) اعمال خواهد شد.

قانون لیتل

■ با برقراری شرایط فوق‌الذکر بر اساس قانون لیتل خواهیم داشت:

میانگین تعداد کارها در سیستم مساوی است با نرخ ورود ضرب در میانگین زمان پاسخ

(Mean number in system = arrival rate x mean response time)

$$N = \lambda R = X R$$

■ این قانون به‌طور شهودی قابل اثبات است که در ادامه ارائه می‌شود.

اثبات شهودی قانون لیتل

- فرض کنید که سیستم را برای بازه زمانی T مشاهده نموده و یک سابقه از زمانهای ورود و خروج هر کار مجزا تهیه نموده‌ایم.
- اگر T بزرگ باشد، تعداد ورودیها باید تقریباً برابر خروجیها باشد. اگر این تعداد را با A نشان دهیم، آنگاه خواهیم داشت:

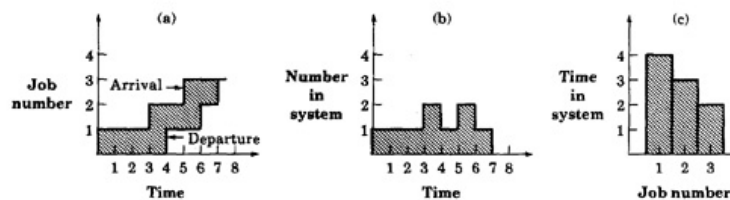
کل زمان / جمع کل ورودیها = نرخ ورود

$$\text{ArrivalRate} = \text{TotalArrivals} / \text{TotalTime}$$

$$\lambda = A/T$$

اثبات شهودی قانون لیتل

- در شکل زیر سه روش برای رسم نمودارهایی از داده‌های جمع‌آوری شده وجود دارد.
- شکل (a) جمع کل ورودیها و خروجیها را به‌طور جداگانه با توجه به کارهای ورودی و خروجی به عنوان تابعی از زمان نشان می‌دهد.
- اگر هر لحظه زمان نمودار خروج را از نمودار ورود کم کنیم، تعداد کارهای موجود در سیستم را در هر لحظه بدست خواهیم آورد، که در شکل (b) نشان داده شده است.
- از طرف دیگر، اگر زمان ورود را از زمان خروج برای هر کار کم کنیم، شکل (c) را برای زمان صرف شده در سیستم بدست خواهیم آورد.



اثبات شهودی قانون لیتل

- ناحیه هاشور خورده در هر سه شکل کل زمان صرف شده توسط کارها را نشان می‌دهید.
- از اینرو مساحت هر سه سطح هاشور زده مساوی هستند، که آنرا با J نشان می‌دهیم.
- با توجه به شکل (c) خواهیم داشت:

$$J/A = \text{میانگین زمان صرف شده در سیستم}$$

- با توجه به شکل (b) خواهیم داشت:

$$J/T = \text{میانگین تعداد کارها در سیستم}$$

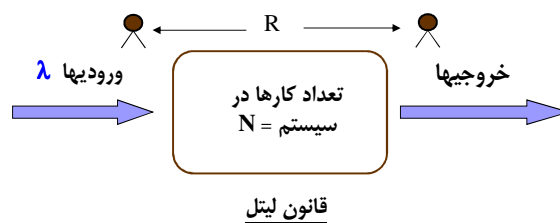
$$= \text{میانگین زمان صرف شده در سیستم} \times \text{نرخ ورود}$$

یا

$$N = \lambda R$$

- که این نتیجه، همان قانون لیتل است.

اثبات شهودی قانون لیتل



$$N = \lambda R \quad \text{یا} \quad L = \lambda W$$

- میانگین تعداد کارها در سیستم مساوی حاصل ضرب نرخ ورود (یا توان عملیاتی) در میانگین زمان اقامت مشتریان در سیستم

مثال اول قانون لیتل

■ در طی یک ساعت مشغولی، تقریباً ۲۰ مشتری به یک آرایشگاه وارد می‌شوند. هر مشتری به طور میانگین ۱۵ دقیقه را در آرایشگاه صرف می‌کند.

□ چه تعداد مشتری به طور همزمان در آرایشگاه حضور دارند؟

□ با توجه به قانون لیتل خواهیم داشت:

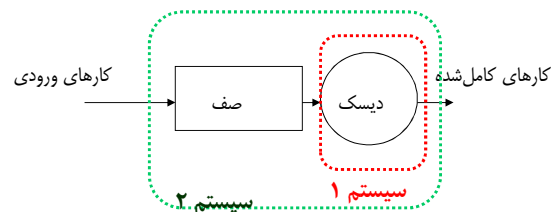
$$\square N = \lambda R$$

$$= 20 \text{ customers/hour} \times (15/60)$$

$$= 5 \text{ customers}$$

مثال دوم قانون لیتل

■ قانون لیتل به هر سیستم و زیرسیستمی از آن قابل اعمال است. در این مثال، اعمال این قانون را به دو بخش داخلی یک سیستم می‌بینیم.



■ یک دیسک می‌تواند ۵۰ درخواست را در ثانیه سرویس دهد. هر درخواست نیازمند ۵ میلی ثانیه زمان سرویس است. میانگین زمان پاسخگویی به هر درخواست هم ۲۰ میلی ثانیه است.

□ با در نظر گرفتن سیستم ۱ خواهیم داشت:

■ $X = 50, S = 0.005; \rho = 0.25$ (بهره‌وری دیسک)

■ در این حالت $S=R$ و $N=\rho$ است. چرا؟ (چون در دیسک صف نداریم). یعنی: $N=\rho=XS=XR$

□ اما با در نظر گرفتن سیستم ۲ خواهیم داشت:

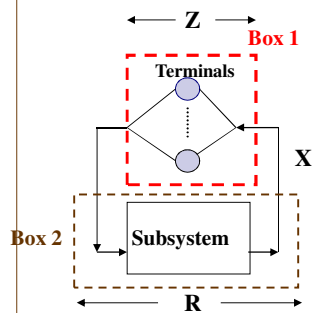
■ $X = 50, R = 0.02; N = XR$

■ $N = 1$ (تعداد کل کارها در سیستم) (queued + at server)

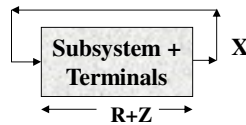
$$N_Q = N - \rho_{\text{disk}} = 1 - 0.25 = 0.75$$

■ تعداد کارهای منتظر چند تا است؟

قانون زمان پاسخ



شکل ۱: مدلی از یک سیستم اشتراک زمانی



شکل ۲: مدل ساده شده

■ قانون بعدی، قانون زمان پاسخ (response time law) است. این قانون به یک سیستم تعاملی (interactive) نظیر یک سیستم اشتراک زمانی (time-sharing) (که یک کامپیوتر مرکزی و تعدادی ترمینال) اعمال می‌شود که دارای پارامترهای زیر است:

□ N : میانگین تعداد ترمینالها

□ Z : میانگین زمان تفکر (think time)

□ R : میانگین زمان پاسخ سیستم

■ در شکل (۱) مدلی از یک سیستم اشتراک زمان ارائه شده است.

■ همچنین در شکل (۲) مدل ساده شده مدل (۱) آمده است.

■ با استفاده از قانون لیتل در مورد مدل شکل (۲)، قانون زمان پاسخ به صورت زیر بدست می‌آید:

$$N = X(R+Z) \Rightarrow R = (N/X) - Z$$

قانون زمان پاسخ

■ قانون لیتل هم به Box 1 و هم به Box 2 قابل اعمال است.

■ با اعمال قانون لیتل به Box 2 خواهیم داشت:

$$N_S = XR$$

□ که در آن N_S تعداد ترمینالهایی است که در حال سرویس یا در حال سعی برای بدست آوردن سرویس هستند.

■ با اعمال قانون لیتل به Box 1 خواهیم داشت:

$$N_T = XZ$$

□ که در آن N_T تعداد کاربرانی است که در حال فکر کردن هستند.

■ توجه کنید که داریم:

$$N = N_T + N_S$$

مثال اول قانون زمان پاسخ

■ در یک سیستم تعاملی دارای ۶۴ کاربر هستیم که هر کدام ۱۵ ثانیه تفکر می کنند. توان عملیاتی سیستم هم ۴ تعامل در ثانیه است. زمان پاسخ (R) چقدر است؟

□ $R = (N/X) - Z = 1 \text{ Sec}$

مثال دوم قانون زمان پاسخ

■ یک سرویس دهنده وب در طی مدت یک ساعت اندازه گیری شده است. در این مدت ۵۰ مشتری به آن متصل شده اند. میانگین زمان CPU مورد تقاضا برای هر درخواست ۵ میلی ثانیه اندازه گیری شده و بهره روی CPU 0.25 بوده است. میانگین زمان پاسخ برای درخواستها هم 0.25 ثانیه بوده است.

□ میانگین زمان تفکر مشتری بر حسب ثانیه چقدر است؟

$N = 50; S_{cpu} = 0.005 \text{ sec}; \rho_{cpu} = 0.25; R = 0.25 \text{ sec}$

■ با استفاده از قانون بهره روی خواهیم داشت:

$$X = \rho_{cpu} / S_{cpu} = 50 \text{ request/sec}$$

■ آنگاه با استفاده از قانون زمان پاسخ خواهیم داشت:

$$Z = (N/X) - R = 0.75 \text{ sec}$$

□ در هر زمان چند مشتری در حال تفکر هستند؟

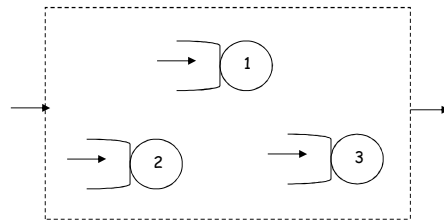
$$N_{cpu} = X R = 50 \times 0.25 = 12.5 \Rightarrow N_T = N - N_{cpu} = 50 - 12.5 = 37.5$$

قانون جریان اجباری

- یک سیستم از تعدادی منابع و یا لوازم ساخته شده است که هر یک از منابع داخلی می تواند یک سیستم (یا زیرسیستم) محسوب شود.
- طبق قوانین عملیاتی وقتی یک تقاضا از محیط دریافت می شود، یک کار در داخل سیستم تولید می شود. این کار ممکن است بین منابع داخلی سیستم گردش کند تا همه پردازش های لازم بر روی آن انجام شود.
- ورودیهای هر منبع می تواند به عنوان یک تقاضا (demand) تلقی شوند که یک کار را در داخل منبع تولید می کنند.

قانون جریان اجباری

- هر درخواست در سطح سیستم (system-level) ممکن است که نیازمند چند بار ملاقات (visit) یکی از منابع سیستم باشد.
- برای مثال یک تراکنش پایگاه داده ها ممکن است نیازمند چندین دسترسی به دیسک باشد.



- **نسبت ملاقات (visit ratio):** نسبت تعداد کارهای کامل شده زیرسیستم k (یا تعداد ملاقاتها) به تعداد کارهای کامل شده کل سیستم است:

$$V_k = C_k / C$$

قانون جریان اجباری

■ اگر V_k نسبت ملاقات منبع k ، X_k توان عملیاتی آن و X هم توان عملیاتی سیستم در مدت T باشد، آنگاه:

$$\square V_k = C_k/C \Rightarrow C_k = V_k C \Rightarrow C_k/T = V_k C/T \Rightarrow X_k = V_k X$$

□ با استفاده قانون بهره‌وری برای منبع k ، یعنی $\rho_k = X_k S_k$ خواهیم داشت:

$$\square \rho_k = X_k S_k = X V_k S_k$$

□ که در آن:

■ S_k میانگین زمان سرویس منبع k برای تقاضای وارده به آن منبع است، و

■ D_k تقاضای سرویس منبع k ، یعنی کل زمان سرویس مورد تقاضا از منبع K است. (مجموع زمانی که این منبع سرویس داده است).

□ بنا بر این داریم:

$$\square D_k = V_k S_k$$

$$\square \rho_k = X D_k$$

مثال اول قانون جریان اجباری

■ یک پایگاه داده‌ها برای مدت ۱۵ دقیقه مشاهده شده است. در طی این مدت CPU سرویس‌دهنده به مدت ۱۲ دقیقه مشغول بوده است. مشاهده شده است که هر تراکنش به‌طور میانگین به ۲ ملاقات CPU نیاز داشته است و سرویس مورد تقاضا در هر بار ملاقات یک میلی ثانیه بوده است.

□ توان عملیاتی سیستم (بر حسب تراکنش در ثانیه) چقدر است؟

$$T = 15 \text{ min}; B_{\text{CPU}} = 12 \text{ min}; V_{\text{CPU}} = 2; S_{\text{CPU}} = 1 \text{ ms}$$

$$\rho_{\text{CPU}} = B/T = 0.8$$

$$\rho_{\text{CPU}} = X V_{\text{CPU}} S_{\text{CPU}} \Rightarrow X = 0.8/(2 \times 0.001) = 400 \text{ transactions/sec}$$

مثال دوم قانون جریان اجباری

■ یک سیستم تعاملی را که دارای مشخصات زیر است در نظر بگیرید:

- تعداد ترمینالها: ۱۸
- میانگین زمان تفکر: ۱۰ ثانیه
- تعداد ملاقاتها به یک دیسک خاص در هر تعامل: ۲۰
- بهره‌وری دیسک فوق: 0.3
- میانگین درخواست سرویس در هر ملاقات دیسک: ۲۰ میلی ثانیه

■ به سؤالات زیر پاسخ دهید:

- توان عملیاتی سیستم چقدر است؟
- توان عملیاتی دیسک چقدر است؟
- زمان پاسخ سیستم چقدر است؟