6. Kiszolgálási idő elemzése, Laplace-transzformáció

Kommunikációs hálózatok teljesítményének elemzése

Horváth Illés

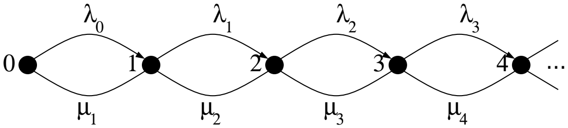
2024/10/09

Vázlat

(1) Emlékeztető: Markov-sorok
(2) Várható kiszolgálási idő részletesebben
(3) Laplace-transzformáció
(4) Kiszolgálási idő eloszlása általános Markov-sorra
(5) Racionális törtfüggvények inverz Laplace-transzformálása
(6) Kiszolgálási idő eloszlása terheléselosztásos rendszerben

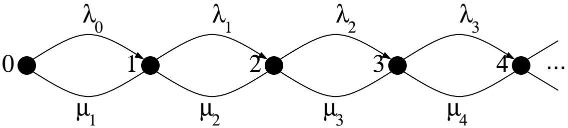
Markov-sorok, M/M/1 és M/M/1/K sor

Emlékeztető: általános Markov-sor.

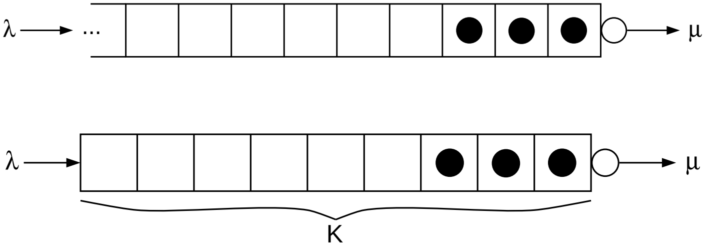


Markov-sorok, M/M/1 és M/M/1/K sor

Emlékeztető: általános Markov-sor.



Azon belül M/M/1 és M/M/1/K sorok: 1 szerver, FIFO (érkezési sorrend szerinti) kiszolgálás.



Kiszolgálási idő

Tipikusan használt teljesítményjellemző a kiszolgálási idő (vagy rendszerben töltött idő), ami egy igénynek a sorba való beérkezés pillanatától a kiszolgálás végéig eltelt idő.

Ezt általában stacionárius rendszerekre vizsgáljuk (azaz feltesszük, hogy a rendszer hosszú ideje fut).

Kiszolgálási idő

Tipikusan használt teljesítményjellemző a kiszolgálási idő (vagy rendszerben töltött idő), ami egy igénynek a sorba való beérkezés pillanatától a kiszolgálás végéig eltelt idő.

Ezt általában stacionárius rendszerekre vizsgáljuk (azaz feltesszük, hogy a rendszer hosszú ideje fut).

Már volt korábban:

- a kiszolgálási idő várható értéke egy véletlen igényre kiszámítható Little-formulával;
- néhány spec esetben (M/M/1, M/M/1/K sor) kiszámítható a kiszolgálási idő teljes eloszlása is.

Kiszolgálási idő

Tipikusan használt teljesítményjellemző a kiszolgálási idő (vagy rendszerben töltött idő), ami egy igénynek a sorba való beérkezés pillanatától a kiszolgálás végéig eltelt idő.

Ezt általában stacionárius rendszerekre vizsgáljuk (azaz feltesszük, hogy a rendszer hosszú ideje fut).

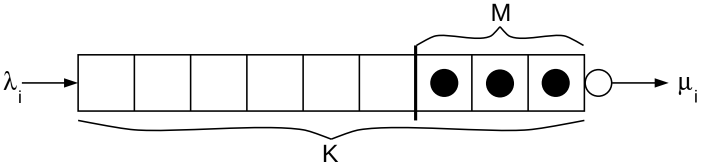
Már volt korábban:

- a kiszolgálási idő várható értéke egy véletlen igényre kiszámítható Little-formulával;
- néhány spec esetben (M/M/1, M/M/1/K sor) kiszámítható a kiszolgálási idő teljes eloszlása is.

Cél: a kiszolgálási idő teljes eloszlásának kiszámítása általános, véges bufferű Markov-sorra.

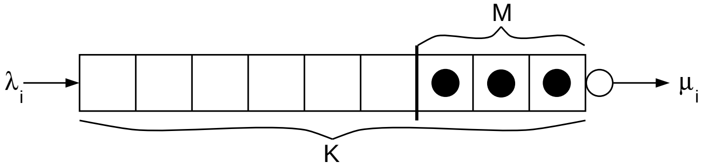
Általános Markov-sor

Tekintsünk egy általános véges bufferű Markov-sort. A buffer mérete K. Az i állapotban az érkezési ráta λ_i, a kiszolgálási ráta μ_i. Feltesszük, hogy a sorhossz eloszlása stacionárius.



Általános Markov-sor

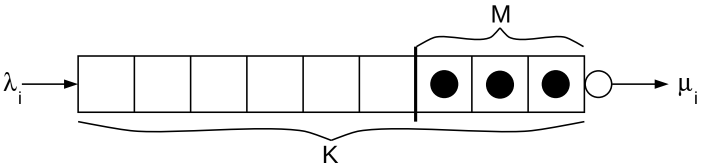
Tekintsünk egy általános véges bufferű Markov-sort. A buffer mérete K. Az i állapotban az érkezési ráta λ_i, a kiszolgálási ráta μ_i. Feltesszük, hogy a sorhossz eloszlása stacionárius.



A kiszolgálási elv FIFO helyett Limited Processor Sharing (LPS): megengedjük, hogy egyszerre több igényt szolgáljon ki a szerver párhuzamosan, maximum M-et.

Általános Markov-sor

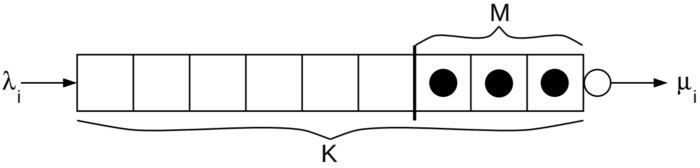
Tekintsünk egy általános véges bufferű Markov-sort. A buffer mérete K. Az i állapotban az érkezési ráta λ_i, a kiszolgálási ráta μ_i. Feltesszük, hogy a sorhossz eloszlása stacionárius.



Ilyenkor, ha legfeljebb M igény van a sorban, akkor a teljes μ_i kiszolgálási ráta egyenletesen oszlik meg a sorban lévő igények között. Ha M-nél több igény van a sorban, akkor az első M igény között oszlik meg a kiszolgálási ráta, a későbbi igények várakoznak.

Általános Markov-sor

Tekintsünk egy általános véges bufferű Markov-sort. A buffer mérete K. Az i állapotban az érkezési ráta λ_i, a kiszolgálási ráta μ_i. Feltesszük, hogy a sorhossz eloszlása stacionárius.



Ilyenkor, ha legfeljebb M igény van a sorban, akkor a teljes μ_i kiszolgálási ráta egyenletesen oszlik meg a sorban lévő igények között. Ha M-nél több igény van a sorban, akkor az első M igény között oszlik meg a kiszolgálási ráta, a későbbi igények várakoznak.

Speciálisan az $M = 1$ esetben visszakapjuk a FIFO kiszolgálást.

Laplace-transzformált

Előkészület: Laplace-transzformáció.

Laplace-transzformált

Előkészület: Laplace-transzformáció.

Egy $f : [0, \infty) \to \mathbb{R}$ függvény Laplace-transzformáltja

$$f^*(s) = \int_0^\infty e^{-st} f(t) \mathrm{d}t.$$

Egy $X \geq 0$ folytonos valószínűségi változó Laplace-transzformáltja

$$X^*(s) = E\left(e^{-sX}\right) = \int_0^\infty e^{-st} f(t) \mathrm{d}t,$$

ahol $f(t)$ az X sűrűségfüggvénye.

Laplace-transzformált

Előkészület: Laplace-transzformáció.

Egy $f : [0, \infty) \to \mathbb{R}$ függvény Laplace-transzformáltja

$$f^*(s) = \int_0^\infty e^{-st} f(t) \mathrm{d}t.$$

Egy $X \geq 0$ folytonos valószínűségi változó Laplace-transzformáltja

$$X^*(s) = E\left(e^{-sX}\right) = \int_0^\infty e^{-st} f(t) \mathrm{d}t,$$

ahol $f(t)$ az X sűrűségfüggvénye.

Ha f sűrűségfüggvény, akkor az integrál konvergens minden $\mathrm{Re}(z) \geq 0$ komplex számra.

Tulajdonságok

Példa. Az $X \sim \mathrm{EXP}(\lambda)$ eloszlás Laplace-transzformáltja

$$X^*(s) = \int_0^\infty e^{-st} \lambda e^{-\lambda t} \mathrm{d}t = \lambda \int_0^\infty e^{-(\lambda+s)t} \mathrm{d}s = \frac{\lambda}{\lambda + s}.$$

Tulajdonságok

Példa. Az $X \sim \mathrm{EXP}(\lambda)$ eloszlás Laplace-transzformáltja

$$X^*(s) = \int_0^\infty e^{-st}\lambda e^{-\lambda t}\mathrm{d}t = \lambda \int_0^\infty e^{-(\lambda+s)t}\mathrm{d}s = \frac{\lambda}{\lambda+s}.$$

Lemma (Teljes várható érték tétel Laplace-transzformáltra)

Ha X, Y, Z valószínűségi változók f_X, f_Y, f_Z sűrűségfüggvénnyel, és

$$\mathbb{P}(Z = X) = p, \qquad P(Z = Y) = 1 - p,$$

akkor

$$f_Z^*(s) = pf_X^*(s) + (1-p)f_Y^*(s).$$

Tulajdonságok

Példa. Az $X \sim \mathrm{EXP}(\lambda)$ eloszlás Laplace-transzformáltja

$$X^*(s) = \int_0^\infty e^{-st}\lambda e^{-\lambda t}\mathrm{d}t = \lambda\int_0^\infty e^{-(\lambda+s)t}\mathrm{d}s = \frac{\lambda}{\lambda+s}.$$

Lemma (Teljes várható érték tétel Laplace-transzformáltra)

Ha X, Y, Z valószínűségi változók f_X, f_Y, f_Z sűrűségfüggvénnyel, és

$$\mathbb{P}(Z = X) = p, \qquad P(Z = Y) = 1 - p,$$

akkor

$$f_Z^*(s) = pf_X^*(s) + (1-p)f_Y^*(s).$$

Biz. (vázlat) A Laplace-transzformált lineáris: ha a, b valós konstansok, akkor

$$(af + bg)^*(s) = a \cdot f^*(s) + b \cdot g^*(s).$$

Tulajdonságok

További fontos tulajdonsága, hogy konvolúcióból (független összegből) szorzatot csinál.

Lemma

Ha X, Y valószínűségi változók függetlenek, és

$$Z = X + Y,$$

akkor

$$Z^*(s) = X^*(s) \cdot Y^*(s).$$

Tulajdonságok

További fontos tulajdonsága, hogy konvolúcióból (független összegből) szorzatot csinál.

Lemma

Ha X, Y valószínűségi változók függetlenek, és

$$Z = X + Y,$$

akkor

$$Z^*(s) = X^*(s) \cdot Y^*(s).$$

1. biz. Az $X^*(s) = E\left(e^{-sX}\right)$ alak alapján; ha X és Y függetlenek, akkor e^{-sX} és e^{-sY} is függetlenek, és

$$\underbrace{E\left(e^{-s(X+Y)}\right)}_{Z^*(s)} = E\left(e^{-sX}e^{-sY}\right) = \underbrace{E\left(e^{-sX}\right)}_{X^*(s)}\underbrace{E\left(e^{-sY}\right)}_{Y^*(s)}.$$

Tulajdonságok

2. biz. Ha $Z = X + Y$, akkor $f_Z = f_X \otimes f_Y$, és

$$f_Z^*(s) = \int_0^\infty f_Z(t)e^{-st}\mathrm{d}t = \int_0^\infty \left(\int_0^t f_Y(u)f_X(t-u)\mathrm{d}u\right) e^{-st}\mathrm{d}t =$$

$$= \int_0^\infty \int_u^\infty f_Y(u)f_X(t-u)e^{-st}\mathrm{d}t\mathrm{d}u.$$

Tulajdonságok

2. biz. Ha $Z = X + Y$, akkor $f_Z = f_X \otimes f_Y$, és

$$f_Z^*(s) = \int_0^\infty f_Z(t) e^{-st} \mathrm{d}t = \int_0^\infty \left(\int_0^t f_Y(u) f_X(t-u) \mathrm{d}u \right) e^{-st} \mathrm{d}t =$$
$$= \int_0^\infty \int_u^\infty f_Y(u) f_X(t-u) e^{-st} \mathrm{d}t \mathrm{d}u.$$

$v = t - u$ változócserével

$$\int_0^\infty \int_u^\infty f_Y(u) f_X(t-u) e^{-st} \mathrm{d}t \mathrm{d}u =$$
$$= \int_0^\infty \int_0^\infty f_Y(u) f_X(v) e^{-s(v+u)} \mathrm{d}v \mathrm{d}u =$$
$$= \int_0^\infty f_Y(u) e^{-su} \mathrm{d}u \cdot \int_0^\infty f_X(v) e^{-sv} \mathrm{d}v =$$
$$= f_Y^*(S) \cdot f_X^*(S).$$

Inverz Laplace-transzformáció

A Laplace-transzformált nekünk most arra jó, hogy ha egy ismeretlen eloszlást elő tudunk állítani ismert eloszlásokból teljes várható érték és konvolúció segítségével, akkor az Laplace-transzformált tartományban egy olyan egyenletet jelent, aminek minden komponense ismert alakú.

Inverz Laplace-transzformáció

A Laplace-transzformált nekünk most arra jó, hogy ha egy ismeretlen eloszlást elő tudunk állítani ismert eloszlásokból teljes várható érték és konvolúció segítségével, akkor az Laplace-transzformált tartományban egy olyan egyenletet jelent, aminek minden komponense ismert alakú.

Laplace-transzformált tartományban megoldjuk az egyenletet, és a megoldást már csak vissza kell transzformálni idő tartományba.

Inverz Laplace-transzformáció

A Laplace-transzformált nekünk most arra jó, hogy ha egy ismeretlen eloszlást elő tudunk állítani ismert eloszlásokból teljes várható érték és konvolúció segítségével, akkor az Laplace-transzformált tartományban egy olyan egyenletet jelent, aminek minden komponense ismert alakú.

Laplace-transzformált tartományban megoldjuk az egyenletet, és a megoldást már csak vissza kell transzformálni idő tartományba.

Inverz Laplace transzformáció: bizonyos (szép/egyszerűbb) függvényekre meg lehet adni az inverz Laplace transzformáltat analitikusan.

Inverz Laplace-transzformáció

A Laplace-transzformált nekünk most arra jó, hogy ha egy ismeretlen eloszlást elő tudunk állítani ismert eloszlásokból teljes várható érték és konvolúció segítségével, akkor az Laplace-transzformált tartományban egy olyan egyenletet jelent, aminek minden komponense ismert alakú.

Laplace-transzformált tartományban megoldjuk az egyenletet, és a megoldást már csak vissza kell transzformálni idő tartományba.

Inverz Laplace transzformáció: bizonyos (szép/egyszerűbb) függvényekre meg lehet adni az inverz Laplace transzformáltat analitikusan.

Bonyolultabb függvényekre általában numerikusan invertáljuk; ehhez ajánlom:

http://inverselaplace.org/

Várható kiszolgálási idő részletesebben

Vissza a Markov-sorhoz. Tekintsünk egy kitüntetett igényt egy j hosszú sorban az i-edik pozícióban ($1 \leq i \leq j \leq K$). Jelölje $X_{i,j}$ a kiszolgálási idejét, és legyen

$$W_{i,j} = E(X_{i,j}).$$

Várható kiszolgálási idő részletesebben

Vissza a Markov-sorhoz. Tekintsünk egy kitüntetett igényt egy j hosszú sorban az i-edik pozícióban ($1 \leq i \leq j \leq K$). Jelölje $X_{i,j}$ a kiszolgálási idejét, és legyen

$$W_{i,j} = E(X_{i,j}).$$

A Little-formula megadja egy véletlen igény kiszolgálási idejének várható értékét; az összes $W_{i,j}$ ismerete ennél részletesebb információt jelent.

Várható kiszolgálási idő részletesebben

Vissza a Markov-sorhoz. Tekintsünk egy kitüntetett igényt egy j hosszú sorban az i-edik pozícióban ($1 \leq i \leq j \leq K$). Jelölje $X_{i,j}$ a kiszolgálási idejét, és legyen

$$W_{i,j} = E(X_{i,j}).$$

A Little-formula megadja egy véletlen igény kiszolgálási idejének várható értékét; az összes $W_{i,j}$ ismerete ennél részletesebb információt jelent.

Megnézzük, hogyan lehet a $W_{i,j}$-ket kiszámolni teljes várható érték tétellel.

Várható kiszolgálási idő részletesebben

A következő változások történhetnek a sorral:

Várható kiszolgálási idő részletesebben

A következő változások történhetnek a sorral:

- Érkezik egy igény a sor végére. Ennek rátája λ_j. Ilyenkor a kitüntetett igény továbbra is az i-edik pozícióban marad, de a sor hossza $(j+1)$-re változik.

Várható kiszolgálási idő részletesebben

A következő változások történhetnek a sorral:

- Érkezik egy igény a sor végére. Ennek rátája λ_j. Ilyenkor a kitüntetett igény továbbra is az i-edik pozícióban marad, de a sor hossza $(j+1)$-re változik.
- Történik egy kiszolgálás a sorban. Ennek rátája μ_j. Ha az a kitüntetett igény, akkor a kiszolgálása befejeződött. Ha egy másik igényt szolgált ki a szerver, akkor a kitüntetett igény $i > M$ esetén előrébb lép az $(i-1)$-edik pozícióba. $i \leq M$ esetén tulajdonképpen mindegy, előre lép-e, mivel az első M igény egyforma helyzetben van.

Várható kiszolgálási idő részletesebben

A következő változások történhetnek a sorral:

- Érkezik egy igény a sor végére. Ennek rátája λ_j. Ilyenkor a kitüntetett igény továbbra is az i-edik pozícióban marad, de a sor hossza $(j+1)$-re változik.
- Történik egy kiszolgálás a sorban. Ennek rátája μ_j. Ha az a kitüntetett igény, akkor a kiszolgálása befejeződött. Ha egy másik igényt szolgált ki a szerver, akkor a kitüntetett igény $i > M$ esetén előrébb lép az $(i-1)$-edik pozícióba. $i \leq M$ esetén tulajdonképpen mindegy, előre lép-e, mivel az első M igény egyforma helyzetben van.

A kétféle változás együttes rátája $\lambda_j + \mu_j$, így a következő állapotváltozásig eltelt idő $\mathrm{EXP}(\lambda_j + \mu_j)$, aminek a várható értéke $\frac{1}{\lambda_j+\mu_j}$, és a változáskor

- $\frac{\lambda_j}{\lambda_j+\mu_j}$ az érkezés valószínűsége, és
- $\frac{\mu_j}{\lambda_j+\mu_j}$ a kiszolgálás valószínűsége.

Várható kiszolgálási idő részletesebben

Legyen először $M < i \leq j < K$. Ilyenkor a kitüntetett igény nem kap kiszolgálást, hanem csak várakozik; a kiszolgálási idejének várható értéke $W_{i,j}$.

Várható kiszolgálási idő részletesebben

Legyen először $M < i \leq j < K$. Ilyenkor a kitüntetett igény nem kap kiszolgálást, hanem csak várakozik; a kiszolgálási idejének várható értéke $W_{i,j}$. A teljes várható érték tétel alapján

$$W_{i,j} = \frac{1}{\lambda_j + \mu_j} + \left(\frac{\lambda_j}{\lambda_j + \mu_j} W_{i,j+1} + \frac{\mu_j}{\lambda_j + \mu_j} W_{i-1,j-1} \right),$$

ahol $\frac{1}{\lambda_j + \mu_j}$ az első változásig eltelt idő várható értéke, és ...

Várható kiszolgálási idő részletesebben

Legyen először $M < i \leq j < K$. Ilyenkor a kitüntetett igény nem kap kiszolgálást, hanem csak várakozik; a kiszolgálási idejének várható értéke $W_{i,j}$. A teljes várható érték tétel alapján

$$W_{i,j} = \frac{1}{\lambda_j + \mu_j} + \left(\frac{\lambda_j}{\lambda_j + \mu_j} W_{i,j+1} + \frac{\mu_j}{\lambda_j + \mu_j} W_{i-1,j-1} \right),$$

ahol $\frac{1}{\lambda_j + \mu_j}$ az első változásig eltelt idő várható értéke, és ...

- ha érkezés történt, akkor a kitüntetett igény marad az i-edik helyen, de a sor hossza $j+1$-re nőtt, és onnan a kiszolgálásig szükséges további idő várható értéke $W_{i,j+1}$;
- ha kiszolgálás történt, akkor a kitüntetett igény pozíciója és a sor hossza is csökken 1-gyel, és a kiszolgálásig szükséges további idő várható értéke $W_{i-1,j-1}$.

Várható kiszolgálási idő részletesebben

Legyen először $M < i \leq j < K$. Ilyenkor a kitüntetett igény nem kap kiszolgálást, hanem csak várakozik; a kiszolgálási idejének várható értéke $W_{i,j}$. A teljes várható érték tétel alapján

$$W_{i,j} = \frac{1}{\lambda_j + \mu_j} + \Big(\frac{\lambda_j}{\lambda_j + \mu_j} W_{i,j+1} + \frac{\mu_j}{\lambda_j + \mu_j} W_{i-1,j-1} \Big),$$

ahol $\frac{1}{\lambda_j + \mu_j}$ az első változásig eltelt idő várható értéke, és ...

- ha érkezés történt, akkor a kitüntetett igény marad az i-edik helyen, de a sor hossza $j+1$-re nőtt, és onnan a kiszolgálásig szükséges további idő várható értéke $W_{i,j+1}$;
- ha kiszolgálás történt, akkor a kitüntetett igény pozíciója és a sor hossza is csökken 1-gyel, és a kiszolgálásig szükséges további idő várható értéke $W_{i-1,j-1}$.

A cél: felírni hasonló egyenleteket az összes $W_{i,j}$-re, majd a kapott lineáris egyenletrendszert megoldani.

Várható kiszolgálási idő részletesebben

További esetek.

Ha $i \leq j \leq M$ és $j < K$, akkor egyrészt

$$W_{1,j} = \ldots W_{j,j},$$

hiszen az első j igény helyzete egyforma, másrészt

$$\begin{aligned} W_{i,j} =& \frac{1}{\lambda_j + \mu_j} + \\ &+ \left(\frac{\lambda_j}{\lambda_j + \mu_j} W_{i,j+1} + \frac{\mu_j}{\lambda_j + \mu_j} \frac{j-1}{j} W_{i-1,j-1} + \frac{\mu_j}{\lambda_j + \mu_j} \frac{1}{j} \cdot 0 \right), \end{aligned}$$

mivel az első j igényből $1/j$ eséllyel a kitüntetett igényt szolgálja ki a szerver, és olyankor az elhagyja a rendszert (0 további időt tölt a rendszerben).

Várható kiszolgálási idő részletesebben

Ha $i \leq M < j < K$, akkor is

$$W_{1,j} = \ldots W_{M,j},$$

hiszen az első M igény helyzete egyforma, másrészt

$$\begin{aligned} W_{i,j} = & \frac{1}{\lambda_j + \mu_j} + \\ & \left(\frac{\lambda_j}{\lambda_j + \mu_j} W_{i,j+1} + \frac{\mu_j}{\lambda_j + \mu_j} \frac{M-1}{M} W_{i-1,j-1} + \frac{\mu_j}{\lambda_j + \mu_j} \frac{1}{M} \cdot 0 \right), \end{aligned}$$

mivel ilyenkor az első M igény között oszlik meg a kiszolgálási ráta, és azon belül $1/M$ eséllyel a kitüntetett igényt szolgálja ki a szerver.

Várható kiszolgálási idő részletesebben

$j = K$, azaz tele buffer esetén nem történhet érkezés, csak kiszolgálás. Ha $i \leq M$, akkor is teljesül

$$W_{1,K} = \ldots W_{M,K},$$

továbbá

$$W_{i,K} = \frac{1}{\mu_j} + \left(\frac{M-1}{M} W_{i-1,M-1} + \frac{1}{M} \cdot 0\right),$$

Várható kiszolgálási idő részletesebben

$j = K$, azaz tele buffer esetén nem történhet érkezés, csak kiszolgálás. Ha $i \leq M$, akkor is teljesül

$$W_{1,K} = \dots W_{M,K},$$

továbbá

$$W_{i,K} = \frac{1}{\mu_j} + \left(\frac{M-1}{M} W_{i-1,M-1} + \frac{1}{M} \cdot 0 \right),$$

Hasonlóan tele buffer, csak $i > M$ esetén a kitüntetett igény biztosan nem kap kiszolgálást, így

$$W_{i,K} = \frac{1}{\mu_j} + W_{i-1,M-1}.$$

Várható kiszolgálási idő részletesebben

A fentiek megadnak a $W_{i,j}$ ismeretlenekre egy lineáris egyenletrendszert, amit meg tudunk oldani.

Várható kiszolgálási idő részletesebben

A fentiek megadnak a $W_{i,j}$ ismeretlenekre egy lineáris egyenletrendszert, amit meg tudunk oldani.

A $W_{i,j}$-k ismeretében megadható egy véletlen igény kiszolgálási idejének várható értéke, W is (amit egyébként a Little-formulából is ki tudunk számolni). Arra kell figyelnünk, hogy általános Markov-sorra az érkezési ráta függ az állapottól.

Várható kiszolgálási idő részletesebben

A fentiek megadnak a $W_{i,j}$ ismeretlenekre egy lineáris egyenletrendszert, amit meg tudunk oldani.

A $W_{i,j}$-k ismeretében megadható egy véletlen igény kiszolgálási idejének várható értéke, W is (amit egyébként a Little-formulából is ki tudunk számolni). Arra kell figyelnünk, hogy általános Markov-sorra az érkezési ráta függ az állapottól.

Ha a sorhossz stacionárius eloszlása $v_{\mathrm{st}} = (x_0\, x_1\, \dots\, x_K)$, akkor az effektív érkezési ráta

$$\lambda_e = x_0\lambda_0 + x_1\lambda_1 + \cdots + x_{K-1}\lambda_{K-1},$$

Várható kiszolgálási idő részletesebben

A fentiek megadnak a $W_{i,j}$ ismeretlenekre egy lineáris egyenletrendszert, amit meg tudunk oldani.

A $W_{i,j}$-k ismeretében megadható egy véletlen igény kiszolgálási idejének várható értéke, W is (amit egyébként a Little-formulából is ki tudunk számolni). Arra kell figyelnünk, hogy általános Markov-sorra az érkezési ráta függ az állapottól.

Ha a sorhossz stacionárius eloszlása $v_{\text{st}} = (x_0\, x_1\, \ldots\, x_K)$, akkor az effektív érkezési ráta

$$\lambda_e = x_0\lambda_0 + x_1\lambda_1 + \cdots + x_{K-1}\lambda_{K-1},$$

$$P(\text{egy véletlen igény } i \text{ hosszú sorba érkezik}) = \frac{x_i\lambda_i}{\lambda_e},$$

végül

$$W = \frac{x_0\lambda_0}{\lambda_e}W_{1,1} + \frac{x_1\lambda_1}{\lambda_e}W_{2,2} + \cdots + \frac{x_{K-1}\lambda_{K-1}}{\lambda_e}W_{K,K}.$$

Kiszolgálási idő eloszlás általános Markov-sorra

Tekintsünk egy kitüntetett igényt egy j hosszú sorban az i-edik pozícióban ($1 \leq i \leq j \leq K$). $X_{i,j}$ a kiszolgálási ideje.

Kiszolgálási idő eloszlás általános Markov-sorra

Tekintsünk egy kitüntetett igényt egy j hosszú sorban az i-edik pozícióban ($1 \leq i \leq j \leq K$). $X_{i,j}$ a kiszolgálási ideje.

Legyen $X_{i,j}$ sűrűségfüggvénye $f_{i,j}(t)$, Laplace-transzformáltja $f^*_{i,j}(s)$.

Kiszolgálási idő eloszlás általános Markov-sorra

Tekintsünk egy kitüntetett igényt egy j hosszú sorban az i-edik pozícióban $(1 \leq i \leq j \leq K)$. $X_{i,j}$ a kiszolgálási ideje.

Legyen $X_{i,j}$ sűrűségfüggvénye $f_{i,j}(t)$, Laplace-transzformáltja $f^*_{i,j}(s)$.

A cél: a korábbi $W_{i,j}$-khez hasonló egyenleteket felírni, csak ezúttal $f^*_{i,j}(s)$-re, majd a kapott egyenletrendszert megoldani.

Kiszolgálási idő eloszlás általános Markov-sorra

Tekintsünk egy kitüntetett igényt egy j hosszú sorban az i-edik pozícióban $(1 \leq i \leq j \leq K)$. $X_{i,j}$ a kiszolgálási ideje.

Legyen $X_{i,j}$ sűrűségfüggvénye $f_{i,j}(t)$, Laplace-transzformáltja $f^*_{i,j}(s)$.

A cél: a korábbi $W_{i,j}$-khez hasonló egyenleteket felírni, csak ezúttal $f^*_{i,j}(s)$-re, majd a kapott egyenletrendszert megoldani.

A következő változások történhetnek a sorral:

- Érkezik egy igény a sor végére λ_j rátával, vagy
- történik egy kiszolgálás a sorban μ_j rátával.

Kiszolgálási idő eloszlás általános Markov-sorra

A kétféle változás együttes rátája $\lambda_j + \mu_j$, így a következő állapotváltozásig eltelt idő $\mathrm{EXP}(\lambda_j + \mu_j)$ eloszlású, ennek Laplace-transzformáltja $\frac{\lambda_j+\mu_j}{\lambda_j+\mu_j+s}$.

Kiszolgálási idő eloszlás általános Markov-sorra

A kétféle változás együttes rátája $\lambda_j + \mu_j$, így a következő állapotváltozásig eltelt idő $\mathrm{EXP}(\lambda_j + \mu_j)$ eloszlású, ennek Laplace-transzformáltja $\frac{\lambda_j+\mu_j}{\lambda_j+\mu_j+s}$.

A változáskor $\frac{\lambda_j}{\lambda_j+\mu_j}$ az érkezés valószínűsége és $\frac{\mu_j}{\lambda_j+\mu_j}$ a kiszolgálás valószínűsége.

Kiszolgálási idő eloszlás általános Markov-sorra

A kétféle változás együttes rátája $\lambda_j + \mu_j$, így a következő állapotváltozásig eltelt idő $\mathrm{EXP}(\lambda_j + \mu_j)$ eloszlású, ennek Laplace-transzformáltja $\frac{\lambda_j+\mu_j}{\lambda_j+\mu_j+s}$.

A változáskor $\frac{\lambda_j}{\lambda_j+\mu_j}$ az érkezés valószínűsége és $\frac{\mu_j}{\lambda_j+\mu_j}$ a kiszolgálás valószínűsége.

Legyen először $M < i \le j < K$. A fentiek alapján a Laplace-transzformáltra teljesül a következő egyenlet:

$$f^*_{i,j}(s) = \frac{\lambda_j + \mu_j}{\lambda_j + \mu_j + s}\left(\frac{\lambda_j}{\lambda_j + \mu_j} f^*_{i,j+1}(s) + \frac{\mu_j}{\lambda_j + \mu_j} f^*_{i-1,j-1}(s)\right)$$

Kiszolgálási idő eloszlás általános Markov-sorra

A kétféle változás együttes rátája $\lambda_j + \mu_j$, így a következő állapotváltozásig eltelt idő $\mathrm{EXP}(\lambda_j + \mu_j)$ eloszlású, ennek Laplace-transzformáltja $\frac{\lambda_j+\mu_j}{\lambda_j+\mu_j+s}$.

A változáskor $\frac{\lambda_j}{\lambda_j+\mu_j}$ az érkezés valószínűsége és $\frac{\mu_j}{\lambda_j+\mu_j}$ a kiszolgálás valószínűsége.

Legyen először $M < i \leq j < K$. A fentiek alapján a Laplace-transzformáltra teljesül a következő egyenlet:

$$f_{i,j}^*(s) = \frac{\lambda_j + \mu_j}{\lambda_j + \mu_j + s}\left(\frac{\lambda_j}{\lambda_j + \mu_j} f_{i,j+1}^*(s) + \frac{\mu_j}{\lambda_j + \mu_j} f_{i-1,j-1}^*(s)\right)$$

Kiszolgálási idő eloszlás általános Markov-sorra

További esetek.

Ha $i \leq j \leq M$ és $j < K$, akkor egyrészt

$$f_{1,j}^*(s) = \ldots f_{j,j}^*(s),$$

hiszen az első j igény helyzete egyforma, másrészt

$$f_{i,j}^*(s) = \frac{\lambda_j + \mu_j}{\lambda_j + \mu_j + s} \times \\ \left(\frac{\lambda_j}{\lambda_j + \mu_j} f_{i,j+1}^*(s) + \frac{\mu_j}{\lambda_j + \mu_j} \frac{j-1}{j} f_{i-1,j-1}^*(s) + \frac{\mu_j}{\lambda_j + \mu_j} \frac{1}{j} \cdot 1 \right),$$

mivel az első j igényből $1/j$ eséllyel a kitüntetett igényt szolgálja ki a szerver, és olyankor az elhagyja a rendszert (0 további időt tölt a rendszerben, aminek a Laplace-transzformáltja 1).

Kiszolgálási idő eloszlás általános Markov-sorra

Ha $i \leq M < j < K$, akkor is

$$f_{1,j}^*(s) = \dots f_{M,j}^*(s),$$

hiszen az első M igény helyzete egyforma, másrészt

$$f_{i,j}^*(s) = \frac{\lambda_j + \mu_j}{\lambda_j + \mu_j + s} \times \Big(\frac{\lambda_j}{\lambda_j + \mu_j} f_{i,j+1}^*(s) + \frac{\mu_j}{\lambda_j + \mu_j} \frac{M-1}{M} f_{i-1,j-1}^*(s) + \frac{\mu_j}{\lambda_j + \mu_j} \frac{1}{M} \cdot 1 \Big),$$

mivel ilyenkor az első M igény között oszlik meg a kiszolgálási ráta, és azon belül $1/M$ eséllyel a kitüntetett igényt szolgálja ki a szerver.

Kiszolgálási idő eloszlás általános Markov-sorra

$j = K$, azaz tele buffer esetén nem történhet érkezés, csak kiszolgálás. Ha $i \leq M$, akkor is teljesül

$$f^*_{1,K}(s) = \dots f^*_{M,K}(s),$$

továbbá

$$f^*_{i,K}(s) = \frac{\mu_j}{\mu_j + s}\left(\frac{M-1}{M} f^*_{i-1,M-1}(s) + \frac{1}{M} \cdot 1\right),$$

Kiszolgálási idő eloszlás általános Markov-sorra

$j = K$, azaz tele buffer esetén nem történhet érkezés, csak kiszolgálás. Ha $i \leq M$, akkor is teljesül

$$f^*_{1,K}(s) = \dots f^*_{M,K}(s),$$

továbbá

$$f^*_{i,K}(s) = \frac{\mu_j}{\mu_j + s}\left(\frac{M-1}{M} f^*_{i-1,M-1}(s) + \frac{1}{M} \cdot 1\right),$$

Hasonlóan tele buffer, csak $i > M$ esetén a kitüntetett igény biztosan nem kap kiszolgálást, így

$$f^*_{i,K}(s) = \frac{\mu_j}{\mu_j + s} f^*_{i-1,M-1}(s).$$

Kiszolgálási idő eloszlás általános Markov-sorra

Az előbbiek megadnak az ismeretlen $f_{i,j}^*(s)$ függvényekre egy lineáris egyenletrendszert, amiben az együtthatók ismert számok és függvények, így az egyenletrendszert meg tudjuk oldani.

Az $f_{i,j}^*(s)$ függvények racionális törtfüggvények lesznek, azaz $\frac{p(s)}{q(s)}$ alakúak, ahol $p(s)$ és $q(s)$ polinomok. Ezekre mindig teljesül $\deg(p) < \deg(q)$, mert $f_{i,j}^*(s)$ egy sűrűségfüggvény Laplace-transzformáltja.

Kiszolgálási idő eloszlás általános Markov-sorra

Az előbbiek megadnak az ismeretlen $f_{i,j}^*(s)$ függvényekre egy lineáris egyenletrendszert, amiben az együtthatók ismert számok és függvények, így az egyenletrendszert meg tudjuk oldani.

Az $f_{i,j}^*(s)$ függvények racionális törtfüggvények lesznek, azaz $\frac{p(s)}{q(s)}$ alakúak, ahol $p(s)$ és $q(s)$ polinomok. Ezekre mindig teljesül $\deg(p) < \deg(q)$, mert $f_{i,j}^*(s)$ egy sűrűségfüggvény Laplace-transzformáltja.

Az $f_{i,j}^*(s)$-ek ismeretében megadható egy véletlen igény kiszolgálási idejének Laplace-transzformáltja, $f^*(s)$ is, ugyanolyan súlyozással, mint korábban:

$$f^*(s) = \frac{x_0 \lambda_0}{\lambda_e} f_{1,1}^*(s) + \frac{x_1 \lambda_1}{\lambda_e} f_{2,2}^*(s) + \cdots + \frac{x_{K-1} \lambda_{K-1}}{\lambda_e} f_{K,K}^*(s).$$

ILT racionális törtfüggvényekre

Végül annyi van hátra, hogy $f^*(s)$-et inverz Laplace transzformáljuk (ILT), és megkapjuk a kiszolgálási idő eloszlásának az $f(t)$ sűrűségfüggvényét. Ha az $F(t)$ eloszlásfüggvényre van szükségünk, akkor annyi a teendő, hogy $f^*(s)/s$-et inverz Laplace transzformáljuk.

ILT racionális törtfüggvényekre

Végül annyi van hátra, hogy $f^*(s)$-et inverz Laplace transzformáljuk (ILT), és megkapjuk a kiszolgálási idő eloszlásának az $f(t)$ sűrűségfüggvényét. Ha az $F(t)$ eloszlásfüggvényre van szükségünk, akkor annyi a teendő, hogy $f^*(s)/s$-et inverz Laplace transzformáljuk.

$f^*(s)$ racionális törtfüggvény, és a racionális törtfüggvényeknek az inverz Laplace-transzformáltja analitikusan kiszámítható parciális törtekre bontással (vagy a python analitikus ILT számoló függvénye is elvégzi nekünk).

ILT racionális törtfüggvényekre

Végül annyi van hátra, hogy $f^*(s)$-et inverz Laplace transzformáljuk (ILT), és megkapjuk a kiszolgálási idő eloszlásának az $f(t)$ sűrűségfüggvényét. Ha az $F(t)$ eloszlásfüggvényre van szükségünk, akkor annyi a teendő, hogy $f^*(s)/s$-et inverz Laplace transzformáljuk.

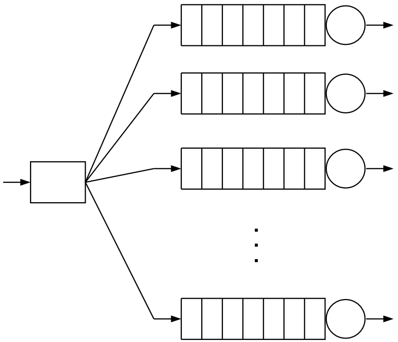
$f^*(s)$ racionális törtfüggvény, és a racionális törtfüggvényeknek az inverz Laplace-transzformáltja analitikusan kiszámítható parciális törtekre bontással (vagy a python analitikus ILT számoló függvénye is elvégzi nekünk).

Még egy apróság: ugyan $f(t)$ elvileg több információt tartalmaz, mint a várható érték (és $f(t)$-ből ki is tudnánk számítani integrálással), de ha csak a várható értékre van szükségünk, akkor közvetlenül a várható értéket kiszámítani általában gyorsabb és numerikusan stabilabb.

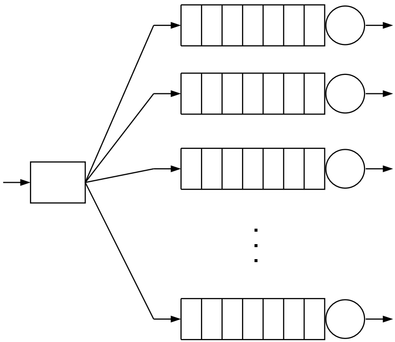
Terheléselosztásos rendszer

A következőben a kiszolgálási idő eloszlását számítjuk ki egy terheléselosztásos rendszerben.

Emlékeztető. Egy olyan rendszert tekintünk, ami N szerverből és egy diszpécserből áll.



Terheléselosztásos rendszer



A bejövő igények mind a diszpécserhez érkeznek Poisson-folyamat szerint, λN rátával. A diszpécsernek nincs sora, azonnal továbbítja a beérkező igényeket a szerverek felé JSQ (Join-Shortest-Queue) szerint, azaz a bejövő igényt a legrövidebb sorhoz továbbítja.

Terheléselosztásos rendszer

A szerverek egyformák és véges K nagyságú bufferrel rendelkeznek. Nem FIFO, hanem LPS kiszolgálást követnek; az egyszerűség kedvéért feltesszük, hogy $M = K$, azaz minden egyes szerver a bufferben lévő összes igénynek nyújt kiszolgálást (Processor Sharing, PS).

Terheléselosztásos rendszer

A szerverek egyformák és véges K nagyságú bufferrel rendelkeznek. Nem FIFO, hanem LPS kiszolgálást követnek; az egyszerűség kedvéért feltesszük, hogy $M = K$, azaz minden egyes szerver a bufferben lévő összes igénynek nyújt kiszolgálást (Processor Sharing, PS).

A kiszolgálási ráta függ a bufferben lévő igények számától; i igény esetén a kiszolgálási ráta μ_i.

Terheléselosztásos rendszer

A szerverek egyformák és véges K nagyságú bufferrel rendelkeznek. Nem FIFO, hanem LPS kiszolgálást követnek; az egyszerűség kedvéért feltesszük, hogy $M = K$, azaz minden egyes szerver a bufferben lévő összes igénynek nyújt kiszolgálást (Processor Sharing, PS).

A kiszolgálási ráta függ a bufferben lévő igények számától; i igény esetén a kiszolgálási ráta μ_i.

A μ_i értékekről általában feltesszük, hogy

$$\mu_1 \leq \mu_2 \leq \cdots \leq \mu_K, \qquad \mu_1 \geq \frac{\mu_2}{2} \geq \cdots \geq \frac{\mu_k}{k}.$$

Terheléselosztásos rendszer

A szerverek egyformák és véges K nagyságú bufferrel rendelkeznek. Nem FIFO, hanem LPS kiszolgálást követnek; az egyszerűség kedvéért feltesszük, hogy $M = K$, azaz minden egyes szerver a bufferben lévő összes igénynek nyújt kiszolgálást (Processor Sharing, PS).

A kiszolgálási ráta függ a bufferben lévő igények számától; i igény esetén a kiszolgálási ráta μ_i.

A μ_i értékekről általában feltesszük, hogy

$$\mu_1 \leq \mu_2 \leq \cdots \leq \mu_K, \qquad \mu_1 \geq \frac{\mu_2}{2} \geq \cdots \geq \frac{\mu_k}{k}.$$

Azt is feltesszük, hogy

$$\lambda < \mu_K.$$

Terheléselosztásos rendszer

A rendszert a mean-field stacionárius tartományban vizsgáljuk (azaz a $t \to \infty$ és $N \to \infty$ limeszben): egyrészt azt feltételezzük, hogy a rendszer olyan hosszú ideje fut, hogy már beállt stacionárius állapotra, másrészt a szerverek száma olyan nagy, hogy gyakorlatilag végtelennek tekinthető.

Terheléselosztásos rendszer

A rendszert a mean-field stacionárius tartományban vizsgáljuk (azaz a $t \to \infty$ és $N \to \infty$ limeszben): egyrészt azt feltételezzük, hogy a rendszer olyan hosszú ideje fut, hogy már beállt stacionárius állapotra, másrészt a szerverek száma olyan nagy, hogy gyakorlatilag végtelennek tekinthető.

Szeretnénk ebben a rendszerben kiszámítani egy igény kiszolgálási idejének eloszlását.

Stacionárius eloszlás

Egy ilyen rendszer mean-field stacionárius eloszlása

$$v_{\mathrm{st}} = (x_0 \; x_1 \; \dots \; x_K),$$

ahol x_k az olyan szerverek számának aránya az összeshez képest, amelyben k igény van a bufferben.

Stacionárius eloszlás

Egy ilyen rendszer mean-field stacionárius eloszlása

$$v_{\mathrm{st}} = (x_0 \; x_1 \; \dots \; x_K),$$

ahol x_k az olyan szerverek számának aránya az összeshez képest, amelyben k igény van a bufferben.

JSQ miatt a stacionárius esetben csak kétféle sorhossz lehet a rendszerben, i_0 és $i_0 + 1$, ahol $i_0 + 1$ a legkisebb olyan index, amire $\lambda < \mu_{i_0+1}$.

Stacionárius eloszlás

Egy ilyen rendszer mean-field stacionárius eloszlása

$$v_{st} = (x_0\, x_1\, \ldots\, x_K),$$

ahol x_k az olyan szerverek számának aránya az összeshez képest, amelyben k igény van a bufferben.

JSQ miatt a stacionárius esetben csak kétféle sorhossz lehet a rendszerben, i_0 és $i_0 + 1$, ahol $i_0 + 1$ a legkisebb olyan index, amire $\lambda < \mu_{i_0+1}$.

Ha lenne i_0-nál rövidebb sor, akkor a teljes λ rátájú beérkező terhelés azokra irányulna JSQ miatt, és ott a kiszolgálási ráta nem elég nagy ahhoz, hogy tartsa a lépést az érkezésekkel, ezért ezek a sorok feltöltődnének i_0-ig.

Stacionárius eloszlás

Egy ilyen rendszer mean-field stacionárius eloszlása

$$v_{\mathrm{st}} = (x_0\, x_1\, \ldots\, x_K),$$

ahol x_k az olyan szerverek számának aránya az összeshez képest, amelyben k igény van a bufferben.

JSQ miatt a stacionárius esetben csak kétféle sorhossz lehet a rendszerben, i_0 és i_0+1, ahol i_0+1 a legkisebb olyan index, amire $\lambda < \mu_{i_0+1}$.

Ha lenne i_0-nál rövidebb sor, akkor a teljes λ rátájú beérkező terhelés azokra irányulna JSQ miatt, és ott a kiszolgálási ráta nem elég nagy ahhoz, hogy tartsa a lépést az érkezésekkel, ezért ezek a sorok feltöltődnének i_0-ig.

Hasonlóan ha lenne i_0+1-nél hosszabb sor, azokra nem irányul beérkező terhelés, tehát azoknak a hossza visszaesik i_0+1-ig.

Stacionárius eloszlás

x_{i_0} és x_{i+1} értékét a

$$\lambda = x_{i_0}\mu_{i_0} + x_{i_0+1}\mu_{i_0+1},$$
$$1 = x_{i_0} + x_{i_0+1}$$

egyenletek alapján lehet meghatározni.

Stacionárius eloszlás

x_{i_0} és x_{i+1} értékét a

$$\lambda = x_{i_0}\mu_{i_0} + x_{i_0+1}\mu_{i_0+1},$$
$$1 = x_{i_0} + x_{i_0+1}$$

egyenletek alapján lehet meghatározni.

Példa. Legyen a buffer méret $K = 5$, a kiszolgálási ráták

k	1	2	3	4	5
μ_k	1.0	1.1	1.2	1.3	1.4

és $\lambda = 1.25$.

Ekkor

Stacionárius eloszlás

x_{i_0} és x_{i+1} értékét a

$$\lambda = x_{i_0}\mu_{i_0} + x_{i_0+1}\mu_{i_0+1},$$
$$1 = x_{i_0} + x_{i_0+1}$$

egyenletek alapján lehet meghatározni.

Példa. Legyen a buffer méret $K = 5$, a kiszolgálási ráták

k	1	2	3	4	5
μ_k	1.0	1.1	1.2	1.3	1.4

és $\lambda = 1.25$.

Ekkor

- $i_0 = 3$, és

Stacionárius eloszlás

x_{i_0} és x_{i+1} értékét a

$$\lambda = x_{i_0}\mu_{i_0} + x_{i_0+1}\mu_{i_0+1}\,,$$
$$1 = x_{i_0} + x_{i_0+1}$$

egyenletek alapján lehet meghatározni.

Példa. Legyen a buffer méret $K = 5$, a kiszolgálási ráták

k	1	2	3	4	5
μ_k	1.0	1.1	1.2	1.3	1.4

és $\lambda = 1.25$.

Ekkor

- $i_0 = 3$, és
- $v_{st} = (0\ 0\ 0\ 0.5\ 0.5\ 0)$.

Stacionárius viselkedés

Tehát a mean-field stacionárius eloszlásban csak kétféle sorhossz szerepel, i_0 és $i_0 + 1$.

Stacionárius viselkedés

Tehát a mean-field stacionárius eloszlásban csak kétféle sorhossz szerepel, i_0 és $i_0 + 1$.

Szeretnénk a rendszer mean-field stacionárius viselkedését megérteni.

Stacionárius viselkedés

Tehát a mean-field stacionárius eloszlásban csak kétféle sorhossz szerepel, i_0 és $i_0 + 1$.

Szeretnénk a rendszer mean-field stacionárius viselkedését megérteni.

Mi a helyzet, ha érkezés történik egy i_0 hosszú sorba?

Stacionárius viselkedés

Tehát a mean-field stacionárius eloszlásban csak kétféle sorhossz szerepel, i_0 és $i_0 + 1$.

Szeretnénk a rendszer mean-field stacionárius viselkedését megérteni.

Mi a helyzet, ha érkezés történik egy i_0 hosszú sorba? Semmi különös, a sorban onnantól kezdve $i_0 + 1$ igény van.

Stacionárius viselkedés

Tehát a mean-field stacionárius eloszlásban csak kétféle sorhossz szerepel, i_0 és $i_0 + 1$.

Szeretnénk a rendszer mean-field stacionárius viselkedését megérteni.

Mi a helyzet, ha érkezés történik egy i_0 hosszú sorba? Semmi különös, a sorban onnantól kezdve $i_0 + 1$ igény van.

Mi a helyzet, ha kiszolgálás történik egy $i_0 + 1$ hosszú sorban?

Stacionárius viselkedés

Tehát a mean-field stacionárius eloszlásban csak kétféle sorhossz szerepel, i_0 és $i_0 + 1$.

Szeretnénk a rendszer mean-field stacionárius viselkedését megérteni.

Mi a helyzet, ha érkezés történik egy i_0 hosszú sorba? Semmi különös, a sorban onnantól kezdve $i_0 + 1$ igény van.

Mi a helyzet, ha kiszolgálás történik egy $i_0 + 1$ hosszú sorban? Itt sincs semmi különös, a kiszolgált igény távozik, és a sorban onnantól kezdve i_0 igény van.

Stacionárius viselkedés

Mi a helyzet, ha kiszolgálás történik egy i_0 hosszú sorban?

Stacionárius viselkedés

Mi a helyzet, ha kiszolgálás történik egy i_0 hosszú sorban? Ez már érdekesebb; a kiszolgált igény távozik, a sorban $i_0 - 1$ igény lesz. De az összes többi sor ennél hosszabb, emiatt a JSQ egyedül erre a szerverre irányítja a teljes λN beérkezési rátát. Minél nagyobb N, annál hamarabb „visszatölti" a sort i_0 hosszúságra; a mean-field limeszben ez azonnal megtörténik, tehát a sort 0 ideig látjuk $i_0 - 1$ hosszúságúnak.

Stacionárius viselkedés

Mi a helyzet, ha kiszolgálás történik egy i_0 hosszú sorban? Ez már érdekesebb; a kiszolgált igény távozik, a sorban $i_0 - 1$ igény lesz. De az összes többi sor ennél hosszabb, emiatt a JSQ egyedül erre a szerverre irányítja a teljes λN beérkezési rátát. Minél nagyobb N, annál hamarabb „visszatölti" a sort i_0 hosszúságra; a mean-field limeszben ez azonnal megtörténik, tehát a sort 0 ideig látjuk $i_0 - 1$ hosszúságúnak.

Másképp fogalmazva: a beérkező igények egy része arra "fordítódik", hogy az i_0 hosszú sorokban a kiszolgálást ellensúlyozza. Ezt upkeep-nek nevezzük. Az i_0 hosszú sorokban a teljes kiszolgálási ráta $x_{i_0}\mu_{i_0}N$, ennek megfelelően

$$\lambda_u := x_{i_0}\mu_{i_0}$$

jelöli az upkeep rátáját.

Stacionárius viselkedés

A maradék

$$\lambda_d = \lambda - \lambda_u$$

ráta az i_0 hosszú sorokba érkező igények rátája (d mint “dynamic dispatch”); ezeknél az igényeknél a diszpécser által látott legrövidebb sorhossz i_0, és azok közül választ találomra.

Stacionárius viselkedés

A maradék

$$\lambda_d = \lambda - \lambda_u$$

ráta az i_0 hosszú sorokba érkező igények rátája (d mint “dynamic dispatch”); ezeknél az igényeknél a diszpécser által látott legrövidebb sorhossz i_0, és azok közül választ találomra.

Ez egyúttal azt is jelenti, hogy egy beérkező igény λ_d/λ valószínűséggel kerül i_0 hosszú sorba, és λ_u/λ valószínűséggel $i_0 + 1$ hosszú sorba.

Stacionárius viselkedés

A maradék

$$\lambda_d = \lambda - \lambda_u$$

ráta az i_0 hosszú sorokba érkező igények rátája (d mint "dynamic dispatch"); ezeknél az igényeknél a diszpécser által látott legrövidebb sorhossz i_0, és azok közül választ találomra.

Ez egyúttal azt is jelenti, hogy egy beérkező igény λ_d/λ valószínűséggel kerül i_0 hosszú sorba, és λ_u/λ valószínűséggel $i_0 + 1$ hosszú sorba.

Az upkeep a mean-field stacionárius eloszlásban közvetlenül nem látszik, de ettől még a megfelelő dinamikus egyensúlyban szerepet játszik.

Stacionárius viselkedés

A maradék

$$\lambda_d = \lambda - \lambda_u$$

ráta az i_0 hosszú sorokba érkező igények rátája (d mint "dynamic dispatch"); ezeknél az igényeknél a diszpécser által látott legrövidebb sorhossz i_0, és azok közül választ találomra.

Ez egyúttal azt is jelenti, hogy egy beérkező igény λ_d/λ valószínűséggel kerül i_0 hosszú sorba, és λ_u/λ valószínűséggel $i_0 + 1$ hosszú sorba.

Az upkeep a mean-field stacionárius eloszlásban közvetlenül nem látszik, de ettől még a megfelelő dinamikus egyensúlyban szerepet játszik.

Egy kitüntetett sor szemszögéből a mean-field stacionárius eloszlás egyfajta konstans környezetet jelent: olyan sok sor van, hogy egy-egy sorban történő kiszolgálás vagy érkezés nem változtatja meg a sorhosszak arányát. Másrészt az összes többi sor között az i_0 illetve $i_0 + 1$ hosszú sorok aránya dinamikus egyensúlyban van.

Kiszolgálási idő eloszlása terheléselosztásos rendszerben

Jelölje X_i egy olyan igény kiszolgálási idejét, amelyik i hosszú sorban van. A PS miatt nem szükséges nyilvántartani, hogy az igény hányadik helyen van a sorban, elég az összes igények számát.

Sőt, JSQ miatt csak kétfajta sorhossz lesz, emiatt nekünk lényegében elég lesz $i \in \{i_0, i_0 + 1\}$-re vizsgálni X_i-t.

Kiszolgálási idő eloszlása terheléselosztásos rendszerben

Jelölje X_i egy olyan igény kiszolgálási idejét, amelyik i hosszú sorban van. A PS miatt nem szükséges nyilvántartani, hogy az igény hányadik helyen van a sorban, elég az összes igények számát.

Sőt, JSQ miatt csak kétfajta sorhossz lesz, emiatt nekünk lényegében elég lesz $i \in \{i_0, i_0 + 1\}$-re vizsgálni X_i-t.

Legyen X_i Laplace-transzformáltja $f_i^*(s)$.

Kiszolgálási idő eloszlása terheléselosztásos rendszerben

Jelölje X_i egy olyan igény kiszolgálási idejét, amelyik i hosszú sorban van. A PS miatt nem szükséges nyilvántartani, hogy az igény hányadik helyen van a sorban, elég az összes igények számát.

Sőt, JSQ miatt csak kétfajta sorhossz lesz, emiatt nekünk lényegében elég lesz $i \in \{i_0, i_0 + 1\}$-re vizsgálni X_i-t.

Legyen X_i Laplace-transzformáltja $f_i^*(s)$.

A Markov-sorokkal analóg módon szeretnénk felírni egy lineáris egyenletrendszert az $f_i^*(s)$-ekre. Ugyanúgy az alapján, hogy egy kitüntetett sorban mi a következő esemény (érkezés vagy kiszolgálás).

Kiszolgálási idő eloszlása terheléselosztásos rendszerben

$$
\begin{aligned}
f^*_{i_0-1}(s) &= f^*_{i_0}(s) \\
f^*_{i_0}(s) &= \frac{\lambda_d + \mu_{i_0}}{\lambda_d + \mu_{i_0} + s} \times \Big(\frac{\lambda_d}{\lambda_d + \mu_{i_0}} f^*_{i_0+1}(s) + \\
&\qquad \frac{\mu_{i_0}}{\lambda_d + \mu_{i_0}} \frac{i_0 - 1}{i_0} f^*_{i_0-1}(s) + \frac{\mu_{i_0}}{\lambda_d + \mu_{i_0}} \frac{1}{i_0} \cdot 1 \Big) \\
f^*_{i_0+1}(s) &= \frac{\mu_{i_0}}{\mu_{i_0} + s} \Big(\frac{i_0}{i_0 + 1} f^*_{i_0}(s) + \frac{1}{i_0 + 1} \cdot 1 \Big)
\end{aligned}
$$

Kiszolgálási idő eloszlása terheléselosztásos rendszerben

$$
\begin{aligned}
f^*_{i_0-1}(s) &= f^*_{i_0}(s) \\
f^*_{i_0}(s) &= \frac{\lambda_d + \mu_{i_0}}{\lambda_d + \mu_{i_0} + s} \times \Big(\frac{\lambda_d}{\lambda_d + \mu_{i_0}} f^*_{i_0+1}(s) + \\
&\qquad \frac{\mu_{i_0}}{\lambda_d + \mu_{i_0}} \frac{i_0 - 1}{i_0} f^*_{i_0-1}(s) + \frac{\mu_{i_0}}{\lambda_d + \mu_{i_0}} \frac{1}{i_0} \cdot 1 \Big) \\
f^*_{i_0+1}(s) &= \frac{\mu_{i_0}}{\mu_{i_0} + s} \Big(\frac{i_0}{i_0 + 1} f^*_{i_0}(s) + \frac{1}{i_0 + 1} \cdot 1 \Big)
\end{aligned}
$$

Az első egyenlet: egy $i_0 - 1$ hosszú sor azonnal visszatöltődik i_0 hosszúságúra.

Kiszolgálási idő eloszlása terheléselosztásos rendszerben

$$
\begin{aligned}
f^*_{i_0-1}(s) &= f^*_{i_0}(s)\\
f^*_{i_0}(s) &= \frac{\lambda_d+\mu_{i_0}}{\lambda_d+\mu_{i_0}+s}\times\Big(\frac{\lambda_d}{\lambda_d+\mu_{i_0}}f^*_{i_0+1}(s)+\\
&\qquad \frac{\mu_{i_0}}{\lambda_d+\mu_{i_0}}\frac{i_0-1}{i_0}f^*_{i_0-1}(s)+\frac{\mu_{i_0}}{\lambda_d+\mu_{i_0}}\frac{1}{i_0}\cdot 1\Big)\\
f^*_{i_0+1}(s) &= \frac{\mu_{i_0}}{\mu_{i_0}+s}\Big(\frac{i_0}{i_0+1}f^*_{i_0}(s)+\frac{1}{i_0+1}\cdot 1\Big)
\end{aligned}
$$

Az első egyenlet: egy i_0-1 hosszú sor azonnal visszatöltődik i_0 hosszúságúra.

Második egyenlet: egy i_0 hosszú sorban $\mathrm{EXP}(\lambda_d+\mu_{i_0})$ idő után történik egy érkezés vagy kiszolgálás; ha kiszolgálás, akkor az is számít, hogy az i_0 igény közül a kitüntetettet távozott (ennek valószínűsége $1/i_0$) vagy nem.

Kiszolgálási idő eloszlása terheléselosztásos rendszerben

$$
\begin{aligned}
f^*_{i_0-1}(s) &= f^*_{i_0}(s)\\
f^*_{i_0}(s) &= \frac{\lambda_d+\mu_{i_0}}{\lambda_d+\mu_{i_0}+s}\times\Bigg(\frac{\lambda_d}{\lambda_d+\mu_{i_0}}f^*_{i_0+1}(s)+\\
&\qquad \frac{\mu_{i_0}}{\lambda_d+\mu_{i_0}}\frac{i_0-1}{i_0}f^*_{i_0-1}(s)+\frac{\mu_{i_0}}{\lambda_d+\mu_{i_0}}\frac{1}{i_0}\cdot 1\Bigg)\\
f^*_{i_0+1}(s) &= \frac{\mu_{i_0}}{\mu_{i_0}+s}\left(\frac{i_0}{i_0+1}f^*_{i_0}(s)+\frac{1}{i_0+1}\cdot 1\right)
\end{aligned}
$$

Az első egyenlet: egy i_0-1 hosszú sor azonnal visszatöltődik i_0 hosszúságúra.

Második egyenlet: egy i_0 hosszú sorban EXP$(\lambda_d+\mu_{i_0})$ idő után történik egy érkezés vagy kiszolgálás; ha kiszolgálás, akkor az is számít, hogy az i_0 igény közül a kitüntetettet távozott (ennek valószínűsége $1/i_0$) vagy nem.

Harmadik egyenlet: ugyanaz, mint a második, csak itt nem történhet érkezés.

Kiszolgálási idő eloszlása terheléselosztásos rendszerben

A fenti lineáris egyenletrendszer egyszerűsíthető és megoldható, a megoldásból a következő súlyozással kapjuk egy véletlenszerű igény kiszolgálási idejének Laplace-transzformáltját:

$$f^*(s) = \frac{\lambda_u}{\lambda} f^*_{i_0}(s) + \frac{\lambda_d}{\lambda} f^*_{i_0+1}(s).$$

Kiszolgálási idő eloszlása terheléselosztásos rendszerben

A fenti lineáris egyenletrendszer egyszerűsíthető és megoldható, a megoldásból a következő súlyozással kapjuk egy véletlenszerű igény kiszolgálási idejének Laplace-transzformáltját:

$$f^*(s) = \frac{\lambda_u}{\lambda} f^*_{i_0}(s) + \frac{\lambda_d}{\lambda} f^*_{i_0+1}(s).$$

Ez is racionális törtfüggvény, tudjuk invertálni, hogy megkapjuk $f(t)$-t, egy véletlen igény kiszolgálási idejének sűrűségfüggvényét, vagy ha $f^*(s)/s$-et invertáljuk, akkor az eloszlásfüggvényét.

Kitekintés

A JSQ-n kívül más terheléselosztási elvek is vannak. Most nem nézzük meg általánosan, de az általános analízis a fentiekhez viszonylag hasonló, csak némileg bonyolultabb.

Kitekintés

A JSQ-n kívül más terheléselosztási elvek is vannak. Most nem nézzük meg általánosan, de az általános analízis a fentiekhez viszonylag hasonló, csak némileg bonyolultabb.

Ha a diszpécser függvénynek (definíció az 5. előadáson) szakadása van v_{st}-ben, akkor $\lambda_u > 0$, és lehet, hogy másképp kell kiszámítani, mint JSQ-ra. Ha a diszpécser függvény folytonos v_{st}-ben, akkor $\lambda_u = 0$.

Kitekintés

A JSQ-n kívül más terheléselosztási elvek is vannak. Most nem nézzük meg általánosan, de az általános analízis a fentiekhez viszonylag hasonló, csak némileg bonyolultabb.

Ha a diszpécser függvénynek (definíció az 5. előadáson) szakadása van v_{st}-ben, akkor $\lambda_u > 0$, és lehet, hogy másképp kell kiszámítani, mint JSQ-ra. Ha a diszpécser függvény folytonos v_{st}-ben, akkor $\lambda_u = 0$.

A Little-formula érzéketlen arra, hogy FIFO vagy LPS kiszolgálási elv van. Viszont a kiszolgálási idő teljes eloszlása más lesz FIFO és LPS esetén – csak éppen a várható értékük ugyanannyi a Little-formula miatt.

Kitekintés

A JSQ-n kívül más terheléselosztási elvek is vannak. Most nem nézzük meg általánosan, de az általános analízis a fentiekhez viszonylag hasonló, csak némileg bonyolultabb.

Ha a diszpécser függvénynek (definíció az 5. előadáson) szakadása van v_{st}-ben, akkor $\lambda_u > 0$, és lehet, hogy másképp kell kiszámítani, mint JSQ-ra. Ha a diszpécser függvény folytonos v_{st}-ben, akkor $\lambda_u = 0$.

A Little-formula érzéketlen arra, hogy FIFO vagy LPS kiszolgálási elv van. Viszont a kiszolgálási idő teljes eloszlása más lesz FIFO és LPS esetén – csak éppen a várható értékük ugyanannyi a Little-formula miatt.

A Laplace-transzformáció tulajdonságai közül csak keveset néztünk át, bővebben ajánlom akár a wikipedia szócikkét is.