

# Teljesítménymodellezés

Hibatűrő Rendszerek Kutatócsoport

2017

## Tartalomjegyzék

1. Alapfogalmak	1	3.3. Erőforrások kihasználtsága . . . . .	3
2. Rendszerszintű tulajdonságok és a Little-törvény	2	3.4. Az átbocsátóképesség és a szűk keresztmetszet . . . . .	4
3. Erőforrások tulajdonságai	2	3.5. A szolgáltatásigény törvénye . . . . .	5
3.1. Rendszerek és alrendszereik kapcsolata . . . . .	2	4. Átlagos mértékek számítása mérési eredményekből	6
3.2. Felhasználói kérések szolgáltatásigénye . . . . .	3	Irodalomjegyzék	7

## Bevezetés

### 1. Alapfogalmak

Teljesítménymodellezéskor egy *rendszert* vizsgálunk, amely *felhasználói kérések* kiszolgálásához illetve feldolgozásához különböző (véges) *erőforrásokat* használ. Vizsgálatunk fókuszában elsősorban az egyes tranzakciók feldolgozási ideje (*válaszidő*), az egységnyi idő alatt feldolgozott tranzakciók száma (*átbocsátás*), illetve az erőforrások *kihasználtsága*, mindez a rendszer *egyensúlyi állapotában*, tehát *átlagos* értékeket mérve.

Egy rendszert sokszor alrendszerek együtteseként modellezünk (ilyen alrendszernek tekinthetők az erőforrások is), ilyenkor az egyes fogalmak több szinten is megjelenhetnek. A továbbiakban a teljes rendszer felé érkező felhasználói kéréseket *tranzakcióknak* fogjuk hívni (darabszámának mértékegysége tr), az ezek feldolgozása során a rendszer által az alrendszereknek továbbított feladatrészeket pedig *kéréseknek* (darabszámának mértékegysége k). Fontos megjegyezni, hogy valójában ugyanarról a fogalomról van szó, de *más rendszerekre nézve*.<sup>1</sup>

Általában is fontos, hogy mindig pontosan definiáljuk az éppen vizsgált rendszert. A továbbiakban bemutatásra kerülő képletek szempontjából is fontos a *rendszer határa*. Egy rendszeren belül lehet várakozási sor, illetve feldolgozó egység, utóbbi állhat több alrendszerből is.<sup>2</sup> Ha egy rendszerben nincs átlapolódás, akkor minden pillanatban (tehát átlagosan is) legfeljebb egy tranzakció lehet a rendszerben. Ha van várakozási sor, vagy több feldolgozó egység is van, akkor definíció szerint van átlapolódás is.

<sup>1</sup>Ezért sem tesz különbséget a fogalmak között a tárgyhoz kiadott diáor.

<sup>2</sup>A rendszer részének tekinthetjük még a hálózati kapcsolatot, és bármi mást, ami késleltetést okozhat, de ettől ebben a segédletben most eltekintünk.

## 2. Rendszerszintű tulajdonságok és a Little-törvény

**Definíció.** Fogalmak (lásd 1. ábra):

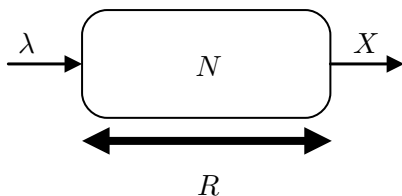
- *Érkezési ráta* (jele:  $\lambda$ ): A vizsgált rendszer határához egységnyi idő alatt *érkező* felhasználói kérések átlagos száma.<sup>a</sup> Mértékegysége: db/s.
- *Átbocsátás* (jele:  $X$ , mint „throughput”): A vizsgált rendszert egységnyi idő alatt *elhagyó* feldolgozott felhasználói kérések átlagos száma. Mértékegysége: db/s.
- *Válaszidő* (jele:  $R$ , mint „round-trip time”): A felhasználói kérések által a rendszer határain belül töltött átlagos idő. Mértékegysége: s.
- *Rendszerben lévő kérések átlagos száma* (jele:  $N$ ): Nevezhetnénk az átlapolódás mértékének is. Mértékegység: db.

<sup>a</sup>A felhasználói kérés itt lehet tranzakció vagy kérés is, attól függ, honnan nézzük. Ennek megfelelően a darab, mint mértékegység is specializálанд az adott esetnek megfelelően.

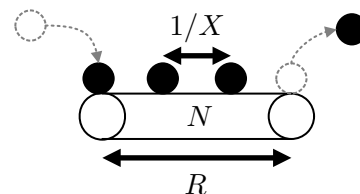
Azt mondjuk, hogy egy rendszer *egyensúlyi állapotban* van, ha  $\lambda = X$ , vagyis egységnyi idő alatt ugyanannyi új felhasználói kérés érkezik a rendszerbe, mint amennyit ezalatt az idő alatt feldolgozott. Egyensúlyi állapotban igaz a *Little-törvény*:

$$N = X \cdot R$$

Szavakkal, a rendszerben tartózkodó kérések átlagos száma megegyezik az átbocsátás és az átlagos rendszerben töltött idő szorzatával. A rendszert például egy futószalagként (2. ábra) elképzelve ez azt jelenti, hogy ha a szalagon  $R$  ideig tart végighaladni, de  $1/X = 1/\lambda$  időnként ráteszünk egy-egy újabb elemet, akkor  $R$  idő múlva az első elem levételének pillanatában  $R/(1/X) = R \cdot X$  elem lesz a szalagon, vagyis a rendszer határain belül.



1. ábra. Rendszerszintű tulajdonságok.



2. ábra. A Little-törvény szemléltetése.

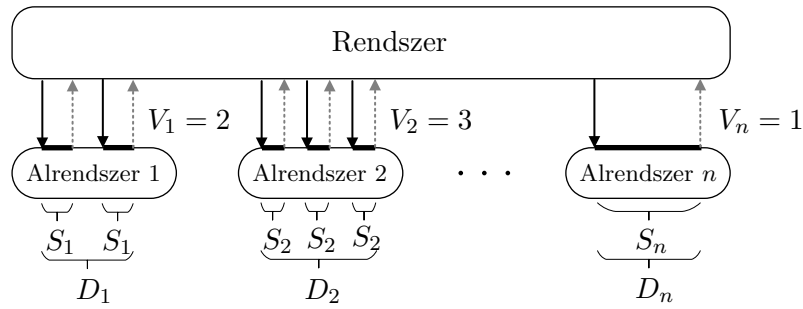
## 3. Erőforrások tulajdonságai

A rendszer tulajdonságai elsősorban a belső szerkezetétől, az alrendszerektől és főképp az erőforrásoktól függ. A rendszerszintű teljesítményjellemzőket ezek tulajdonságaiból kell levezetni. A továbbiakban nullás index jelöli a rendszerszintű tulajdonságokat, míg az  $i$ . alrendszer (erőforrás) tulajdonságait  $i$ -vel indexeljük.

### 3.1. Rendszerek és alrendszereik kapcsolata

Az egyes alrendszerek és erőforrások teljesítményjellemzőit a következő mértéket felhasználva tudjuk átváltani a rendszer jellemzőire, illetve fordítva:

**Definíció.** *Látogatások átlagos száma* (jele:  $V_i$ , mint „visits”): Megadja, hogy egy tranzakció átlagosan hány kérést generál az  $i$ . alrendszer (erőforrás) felé. Mértékegysége: k/tr (kérés/tranzakció).



3. ábra. Rendszer és alrendszereinek kapcsolata.

### 3.2. Felhasználói kérések szolgáltatásigénye

Egy tranzakció „terhelését” a rendszerre nézve a *szolgáltatásigény* fogalmával ragadjuk meg. A szolgáltatásigény az az átlagos időtartam, amíg a tranzakció feldolgozása közben a rendszer egy adott erőforrást használ (az egyes kérések során összesen), tehát minden erőforráshoz külön érték tartozik. Az alábbiakban feltételezzük, hogy az alrendszerek erőforrások.<sup>3</sup>

**Definíció.** *Szolgáltatásigény* (jele:  $D_i$ , mint „service demand”): Megadja, hogy egy *tranzakció* átlagosan mennyi ideig használja az adott erőforrást (alrendszert). Mértékegysége:  $\frac{s}{tr}$ .

**Definíció.** *Erőforrásigény* (jele:  $S_i$ , mint „resource demand”): Megadja, hogy egy *kérés* átlagosan mennyi ideig használja az adott erőforrást (alrendszert). Mértékegysége:  $\frac{s}{k}$ .

Látható, hogy a két fogalom gyakorlatilag ugyanazt takarja, de az egyik a rendszer szintjén, a másik pedig az erőforrás (alrendszer) szintjén.<sup>4</sup> A két mennyiség közötti kapcsolatot a látogatások átlagos számának ( $V_i$ ) segítségével a következő képlet adja meg:

$$D_i = V_i \cdot S_i$$

Látható, hogy itt sem történik semmi meglepő – ha egy kérés átlagosan  $S_i$  ideig foglalja az erőforrást, egy tranzakció pedig átlagosan  $V_i$  kérést generál, akkor a tranzakció átlagosan  $D_i$  ideig fogja használni az erőforrást, tehát az erőforrásra vonatkozó szolgáltatásigénye  $D_i$ . Ez az összefüggés látható a 3. ábrán is.

### 3.3. Erőforrások kihasználtsága

Véges készletű erőforrások esetén a teljesítmény szempontjából fontos tulajdonság az erőforrások átlagos *kihasználtsága* (jele:  $U$ , mint **utilization**), ugyanis ez mutatja meg, hogy a globális teljesítménykorlátoktól nagyjából milyen távol működik a rendszer.

Vegyük észre, hogy az erőforrás, mint alrendszer önmagában is egy rendszert alkot, ezért a 2. szakaszban leírtak itt is alkalmazhatók. A felhasználói kérés ekkor a tranzakció által generált kérés, a kérés által a rendszerben töltött átlagos idő ( $R$ ) pedig az erőforrásigénynek ( $S_i$ ) felel meg. Az erőforrás, mint alrendszer átbecsátását az ún. *Forced Flow törvény* segítségével számíthatjuk ki a teljes rendszer átbecsátásából:<sup>5</sup>

$$X_i = V_i \cdot X_0$$

Ezzel felírva a Little-törvényt, az alábbi képletet kapjuk:

<sup>3</sup>Ez bizonyos szempontból csak formaság, annyi a jelentősége, hogy erőforrás szint alatt nem foglalkozunk a további alrendszerekkel, itt húzzuk meg vizsgálódásaink határát.

<sup>4</sup>Az elnevezés is csak azért különbözik, hogy külön tudjunk hivatkozni a két értékre, az „Erőforrásigény” név pedig kihasználja, hogy most csak az erőforrásokat tekintjük alrendszereknek.

<sup>5</sup>Figyelem! A mértékegység az alrendszerek átbecsátása esetén k/s, a teljes rendszer esetén pedig tr/s!

$$N_i = S_i \cdot X_i$$

Az  $N$  érték tehát szokás szerint megadja, hogy átlagosan hány kérés tartózkodik a rendszerben, ez esetben az erőforráson belül. Az egyes erőforrásokból több példány is rendelkezésre állhat, ezeken belül feltételezzük, hogy nincs átlapolódás. Az erőforrás tehát egyszerre legfeljebb  $n_i$  kérést képes kiszolgálni, ahol  $n_i$  az  $i$ . erőforrásból elérhető példányok száma. Ha az  $i$ . erőforrásban, mint rendszerben átlagosan  $N_i$  kérés tartózkodik, és ez kevesebb, mint a maximális  $n_i$ , akkor a rendszer nem használja ki az erőforrást. Gyakorlatilag ezzel definiáltuk is a kihasználtság fogalmát, amire a *kihasználtság törvénye* ad képletet:

$$U_i = \frac{N_i}{n_i} = \frac{S_i \cdot X_i}{n_i}$$

Az is látható, hogy két darabszámot osztunk el egymással, tehát az eredmény dimenzió nélküli, százalékban kifejezhető arányszám. Az  $n = 1$  speciális esetben  $N$  értéke közvetlenül megadja a kihasználtság értékét:

$$U_i = S_i \cdot X_i$$

Ilyenkor a kihasználtság úgy is értelmezhető, mint az egységnyi időnek azon hányada, amelyben átlagosan az erőforrás munkát végez. Ez az értelmezés bizonyos szempontból analóg a fizikából ismert hatásfok fogalmával, az erőforrás a vizsgált 1s időben  $X_i$  alkalommal  $S_i$  ideig hasznos munkát végzett, ez összesen  $S_i \cdot X_i$  idő hasznos munkát jelent, ami tehát  $\frac{S_i \cdot X_i}{1} = U$  hatásfokot, itt *kihasználtságot* jelent. Több erőforrás példány esetén is hasonló a helyzet, de ilyenkor az egységnyi időt mindegyik példányhoz fel kell számolni.

### 3.4. Az átbocsátóképesség és a szűk keresztmetszet

Az imént láttuk, hogy az erőforráskészlet felső határt szab az elvégezhető munka mennyiségének, ezáltal az egységnyi idő alatt kiszolgálható kéréseknek, vagyis az átbocsátásnak. Ezt a felső határt hívjuk *átbocsátóképességnek*.

Meghatározásához az egyes erőforrásokból kell kiindulnunk. Feltételezzük, hogy az erőforrásokat maximálisan kihasználjuk, vagyis  $U_i^{max} = 1$ . A kihasználtság képletéből az átlagos erőforrásigény ( $S_i$ ) ismeretében kiszámolhatjuk az adott erőforrás átbocsátóképességét:

$$X_i^{max} = n_i \cdot \frac{U_i^{max}}{S_i} = n_i \cdot \frac{1}{S_i}$$

Az  $n_i = 1$  esetben ismét egyszerűbben:

$$X_i^{max} = \frac{U_i^{max}}{S_i} = \frac{1}{S_i}$$

Következő lépésként ki kellene számolnunk a teljes rendszer átbocsátását, de míg a teljes rendszerből az erőforrásra következtetni a többi erőforrástól függetlenül is tudtunk a 3.3 szakaszban, visszafelé ez most nem lesz igaz. Az egyes erőforrásokat a teljes rendszer ugyanis különböző mértékben használja, így csak az egyikük (esetleg néhányuk, nagyon ritkán mindegyik) korlátozza ténylegesen a rendszer tényleges átbocsátóképességét. Ezt az erőforrást nevezzük *szűk keresztmetszetnek*.

Ki kell tehát számolnunk az egyes erőforrások átbocsátásából a rendszer átbocsátásának elméleti maximumát, majd az így kapott értékek legkisebbikét kell kiválasztanunk, ez lesz a rendszer átbocsátóképessége, az értékhez tartozó erőforrás(ok) pedig a rendszer szűk keresztmetszete(i):

$$X_0^{max} = \min_i \left( \frac{X_i^{max}}{V_i} \right)$$

Az átbocsátóképességgel felírva a Little-törvényt megkaphatjuk  $N_{max}$ -ot, vagyis az átlapolódás maximális mértékét a rendszerben. Abban a speciális esetben, amikor tudjuk, hogy a rendszer (erőforrás) mindig rendelkezésre áll, és nincs benne átlapolódás, vagyis  $N_{max} = 1$  tr, az átbocsátóképesség (de nem az átbocsátás!) és a rendszer átlagos válaszideje között fordított arányosság áll fenn:<sup>6</sup>

$$X_0^{max} = \frac{1 \text{ tr}}{R}$$

### 3.5. A szolgáltatásigény törvénye

Az eddigiekből levezethető még egy összefüggés, ami sokszor jól használható. A *szolgáltatásigény törvénye* egy adott erőforrásra vonatkozó szolgáltatásigény meghatározását teszi lehetővé az erőforrás kihasználtsága és a rendszer átbocsátóképessége segítségével, egyetlen erőforráspéldány ( $n_i = 1$ ) esetén:

$$D_i = \frac{U_i}{X_0}$$

A levezetés a forced flow törvény és a kihasználtság törvényének  $n_i = 1$  feltétel melletti egyszerűbb alakja szerint:

$$\underbrace{D_i = V_i \cdot S_i}_{\text{Szolgáltatásigény def.}} = \underbrace{\frac{X_i}{X_0}}_{\text{Forced flow tv.}} \cdot \underbrace{\frac{U_i}{X_i}}_{\text{Kihasználtság tv.}} = \frac{U_i}{X_0}$$

A szolgáltatásigény törvénye lényegében a kihasználtság törvényének olyan átalakítása, hogy az erőforrás-szintű tulajdonságok helyett rendszerszintű tulajdonságokkal számoljon. Az alábbi levezetés ezt az elvet mutatja be, itt lényegében a  $V_i$  váltószám segítségével mindkét oldalon áttérünk az erőforrás szintjéről a rendszer szintjére:

$$U_i = \frac{S_i \cdot X_i}{n_i}$$

$$S_i = n_i \cdot \frac{U_i}{X_i}$$

$$\underbrace{V_i \cdot S_i}_{D_i} = n_i \cdot \underbrace{V_i \cdot \frac{1}{X_i}}_{\frac{1}{X_0}} \cdot U_i$$

$$D_i = n_i \cdot \frac{U_i}{X_0}$$

Itt is jól látható, hogy a törvény eredeti formája az  $n_i = 1$  esetre érvényes, minden más esetben megjelenik egy  $n_i$ -s szorzó is. A kihasználtság törvényénél látott szemléltető magyarázat ismét alkalmazható: Ha egy tranzakció  $D_i$  ideig használja az adott erőforrást, és egy egységnyi idő alatt  $X_0$  ilyen tranzakció történik, akkor az erőforrás az egységnyi idő  $\frac{D_i \cdot X_0}{1}$  részéig volt foglalt.

<sup>6</sup>Lényegében ugyanez az összefüggés jelenik meg az erőforrás átbocsátóképességének meghatározásakor  $R = S_i \cdot 1k$  választással.

## 4. Átlagos mértékek számítása mérési eredményekből

A fenti értékeket jellemzően mérések vagy szimulációk eredményeképp kapjuk. Ilyen mérések során a rendszert állandósult egyensúlyi állapotban vizsgáljuk, amikor tehát  $\lambda = X_0$ . Ilyenkor az alábbi fogalmak jelennek meg, ezek segítségével tudjuk kiszámolni a rendszer és az erőforrások tulajdonságait:

### Definíció.

- *Mérési idő* (jele:  $T$ , mint „time”). Mértékegysége: s.
- *Tranzakciók száma* (jele:  $C_0$ , mint „count”): A mérési idő alatt elvégzett tranzakciók száma.  $C_i$ -vel jelölhetjük az egyes alrendszerekre vonatkozó értékeket, ha ez szükséges. Mértékegysége: tr (vagy alrendszerek esetén k).
- *Foglaltsági idő* (jele:  $B_i$ , mint „busy time”): Az egyes erőforrások foglaltsági ideje a mért időtartamon belül. Mértékegysége: s.

Ezekből a fogalmakból könnyedén kiszámítható például egy egypéldányos, átlapolódás nélküli erőforrás ( $n_i = 1$ ) átlagos kihasználtsága a mérés ideje alatt, csupán a foglaltsági idő és a mért idő arányát kell kiszámítanunk:

$$U_i = \frac{B_i}{T}$$

A mérés ideje alatt az átlagos átbocsátás (és a rendszer egyensúlya miatt az érkezési ráta is) megkapható a mérési időből és az elvégzett tranzakciók számából:<sup>7</sup>

$$X_0 = \frac{C_0}{T}$$

A tranzakciók átlagos szolgáltatásigénye is megkapható az egyes erőforrásokhoz, ha a foglaltsági időket leosztjuk a tranzakciók számával:

$$D_i = \frac{B_i}{C_0}$$

Érdekességgépp megjegyezzük, hogy a szolgáltatásigény törvénye ( $n_i = 1$  esetre) akár ebből a három összefüggésből is levezethető:

$$U_i = \frac{B_i}{T} = \frac{B_i}{\frac{C_0}{X_0}} = \frac{B_i}{C_0} \cdot X_0 = D_i \cdot X_0$$

<sup>7</sup>Értelemszerűen az egyes alrendszerek vizsgálata esetén az  $i$  index használandó a 0 helyett.

## Hivatkozások

## Tárgymutató

átbocsátás throughput 1, 2

átbocsátóképesség 4

átlag average, mean 1

egyensúlyi állapot 1, 2

érkezési ráta arrival rate 2

erőforrás resource 1

erőforrásigény 3

felhasználói kérés user request 1

foglaltsági idő busy time 6

Forced Flow törvény Forced Flow Law 3

kérés request [ˈrɛkwɛst] 1, 3

kihasználtság 1, 3, 4

kihasználtság törvénye 4

látogatások átlagos száma 2

Little-törvény Little's Law 2

mérési idő 6

rendszer system [ˈsɪstəm] 1

rendszer határa system boundary [ˈbaʊndəri] 1

rendszerben lévő kérések átlagos száma  
average number of requests in the system  
2

szolgáltatásigény 3

szolgáltatásigény törvénye 5

szűk keresztmetszet bottleneck [ˈbɒtlɪnek] 4

tranzakció transaction 1, 3

tranzakciók száma 6

válaszidő service time 1, 2