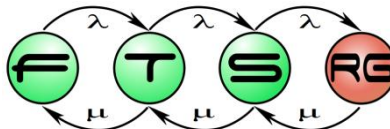
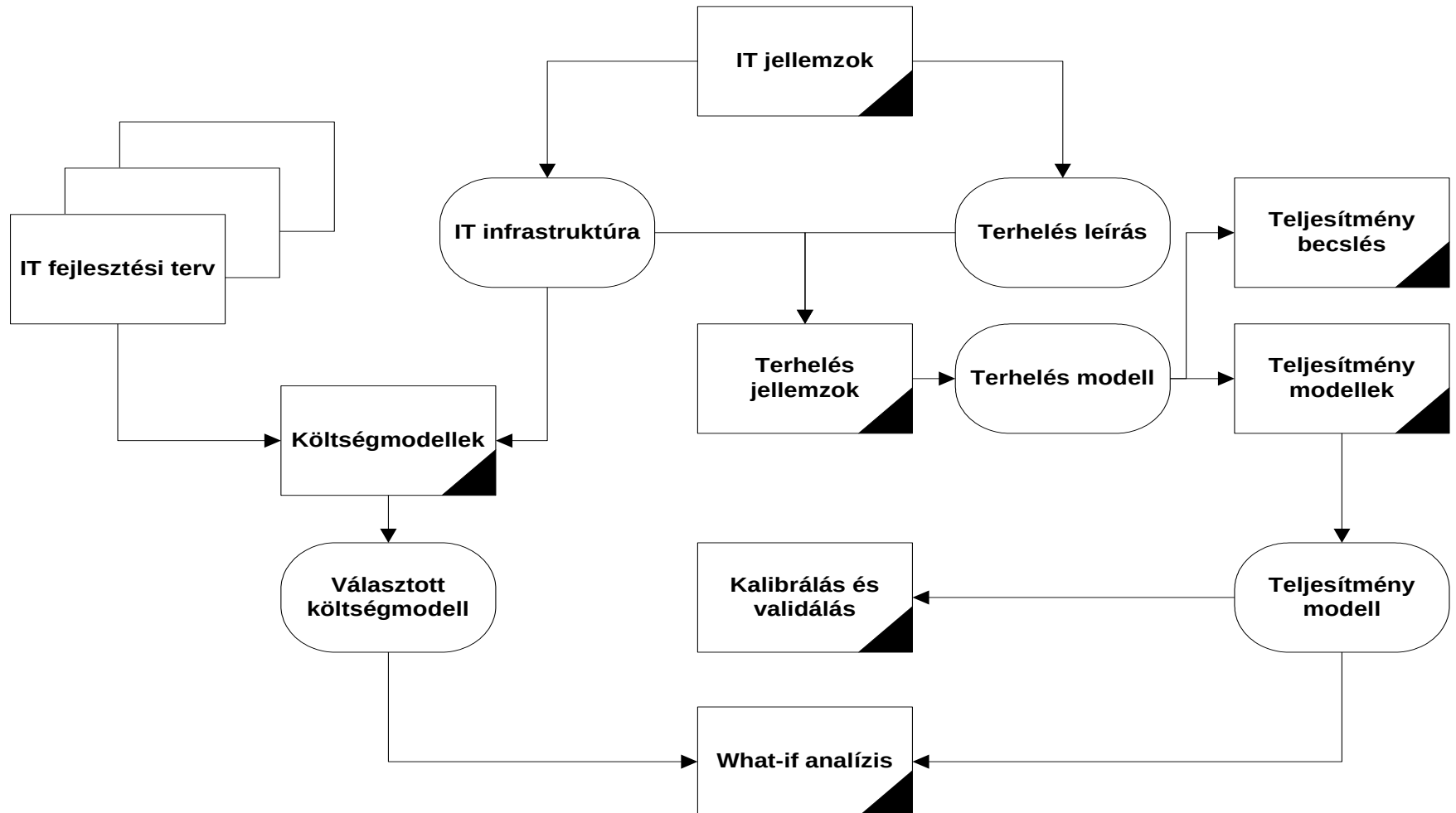


Teljesítménymodellezés

Gönczy László
gonczy@mit.bme.hu



Erőforrás szintű kapacitástervezés lépései

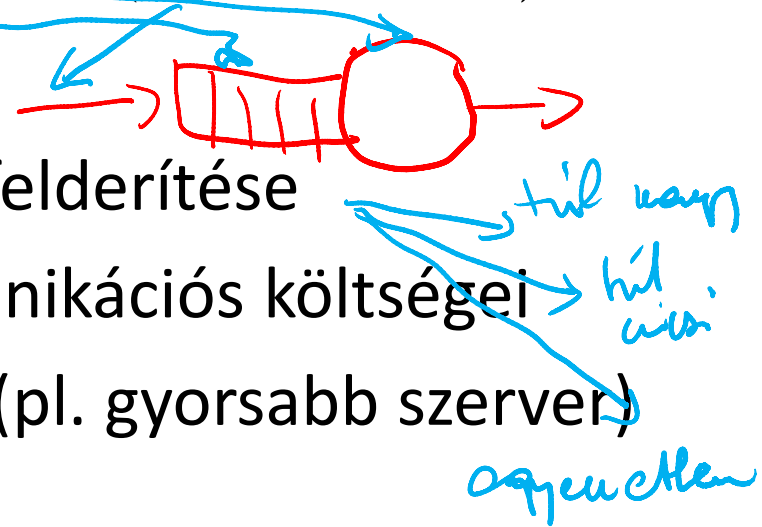


Teljesítménymodellek

$$T_{\text{Kiszolgálás}} = T_{\text{Várakozás}} + T_{\text{TénylegesKiszolgálás}} (+ T_{\text{HálózatiKésleltetés}})$$

■ Modellezés célja

- erőforrás foglalási problémák felderítése
- elosztott alkalmazások kommunikációs költségei
- rendszer változásának hatásai (pl. gyorsabb szerver)
- **előrejelzés** támogatása



■ Ökölszabályok: teljesítménymodell elfogadható, ha

- az erőforrások kihasználtságát 10%
- az áteresztőképességet 10%
- a válaszidőt 20% hibával becsli

Modellek fajtái

■ Analitikus modell

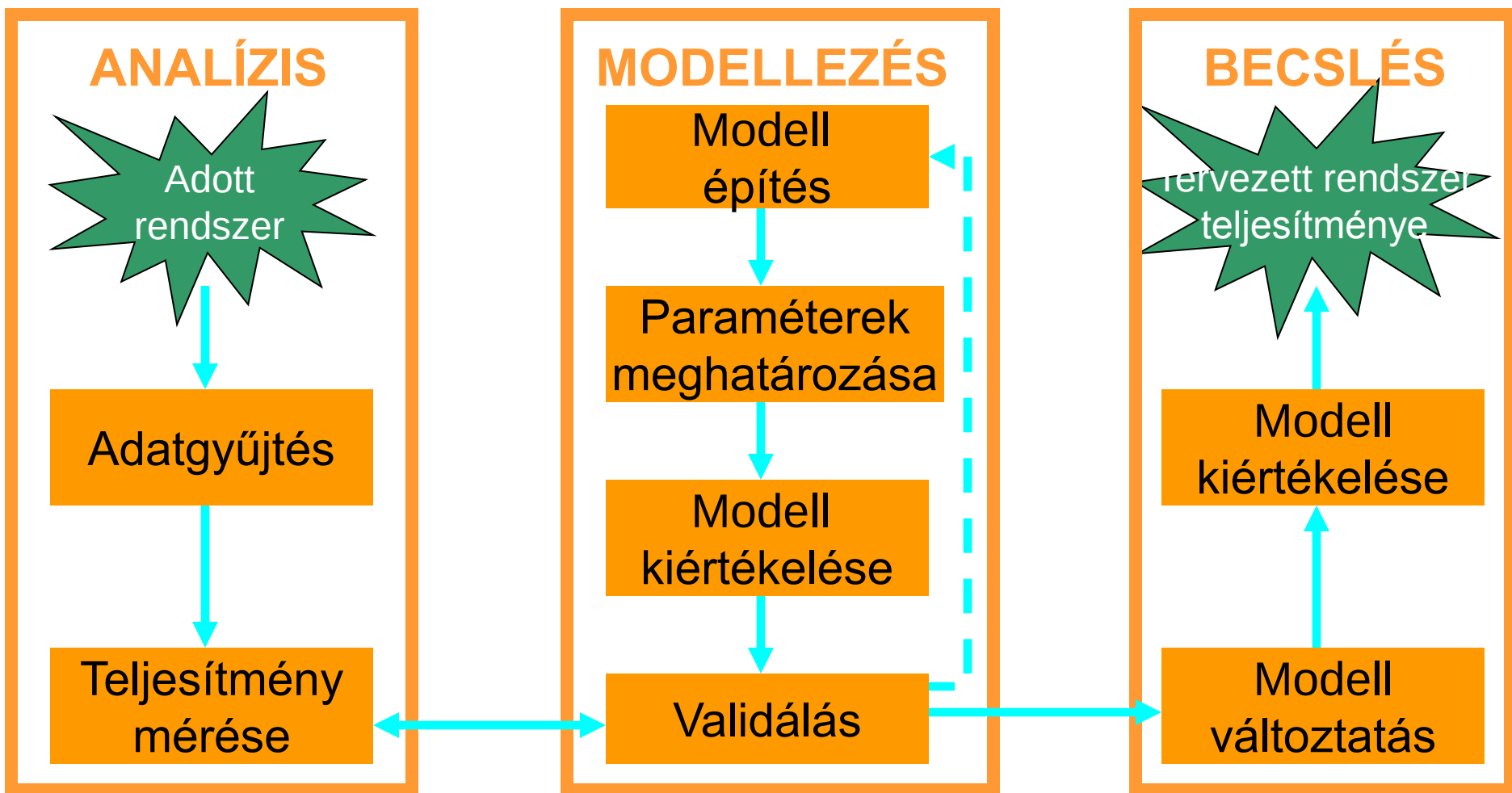
- a rendszert egyenletekkel írja le, pl.

$$RT_{\min} = RTT + kérés_{\min} + Feldolgozásidő + válasz_{\min}$$

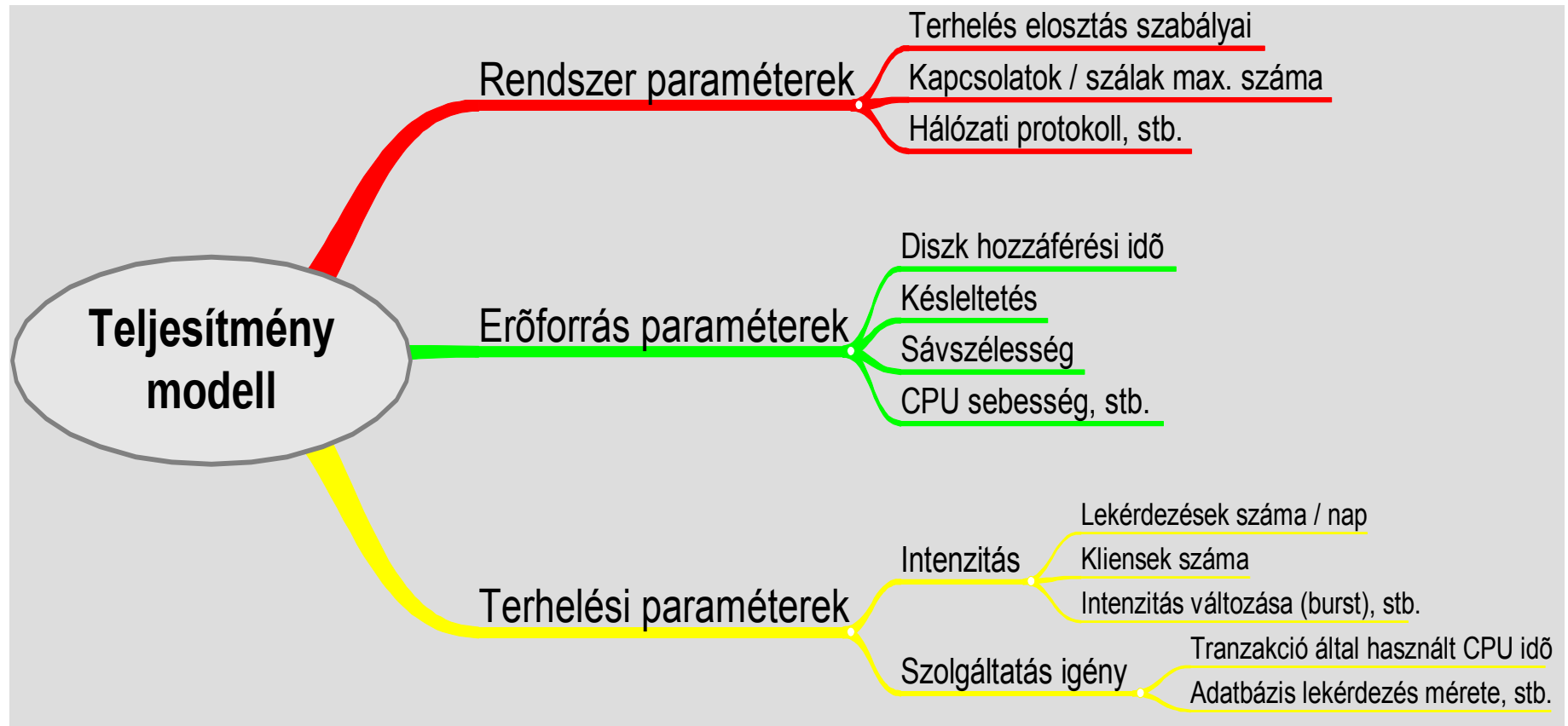
■ Szimulációs modell

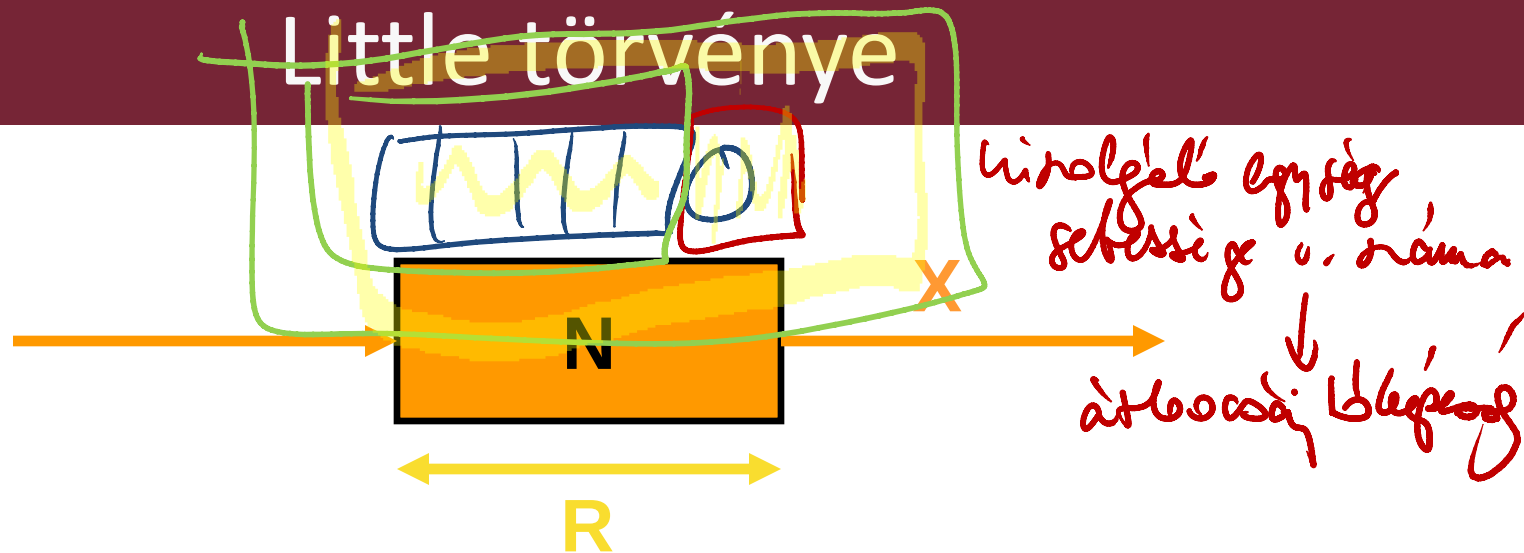
- szimulációt futtat
- az előfordulásnak megfelelő tranzakció gyakorisággal
- előny: általános vizsgálat 😊
- hátrány: drága, nehéz kifejleszteni ☹️

Modellezési/becslési paradigma



Teljesítmény modell paramétere





- N: a rendszerben lévő kérések átlagos száma
- X: áteresztőképesség
- R: a rendszerben töltött átlagos idő

↑ kérés
↓ kérés
↑ kérés
↓ kérés

$$N = X \times R$$

1 embert
szolgálati

1 kérés/perc
(beleértve a két mozgásidőt)

→ 1 perc

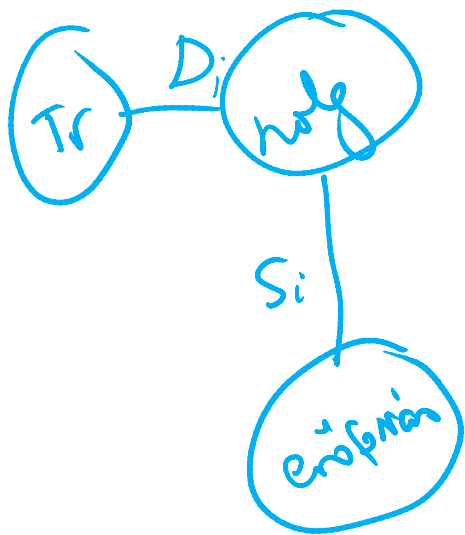
Szolgáltatás igény és idő

- Az i . erőforrásra:

D_i : egy tranzakció átlagos szolgáltatásigénye

V_i : a tranzakció átlagos erőforrás használata

S_i : egy használat átlagos erőforrás igénye



$$D_i = V_i \times S_i$$

$$S_i = \frac{D_i}{V_i}$$

Könnyebben
meghatározható

Kihasználtság törvénye

■ Kihasználtság törvénye (Utilization Law)

$$U_i = B_i/T = B_i/(C_0/X_i) = (B_i/C_0) \times X_i = S_i \times X_i$$

U_i : az i . erőforrás kihasználtsága

B_i : foglaltsági ideje a monitorozás alatt

T : mérési idő

C_0 : tranzakciók száma

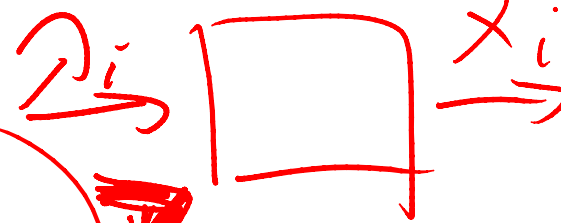
S_i : átlagos kiszolgálási idő

X_i : átlagos áteresztőképesség

λ_i : érkezési ráta

Egyensúly : $\lambda_i = X_i$

*de
idő*



$$U_i = X_i \times S_i = \lambda_i \times S_i$$

*csak egyenlőre
all*

További törvények

Forced Flow törvény:

ü.

X_i : az i . erőforrás áteresztőképessége $X_i = V_i \times X_0$

V_i : „látogatások” átlagos száma

X_0 : tranzakciók átlagos száma

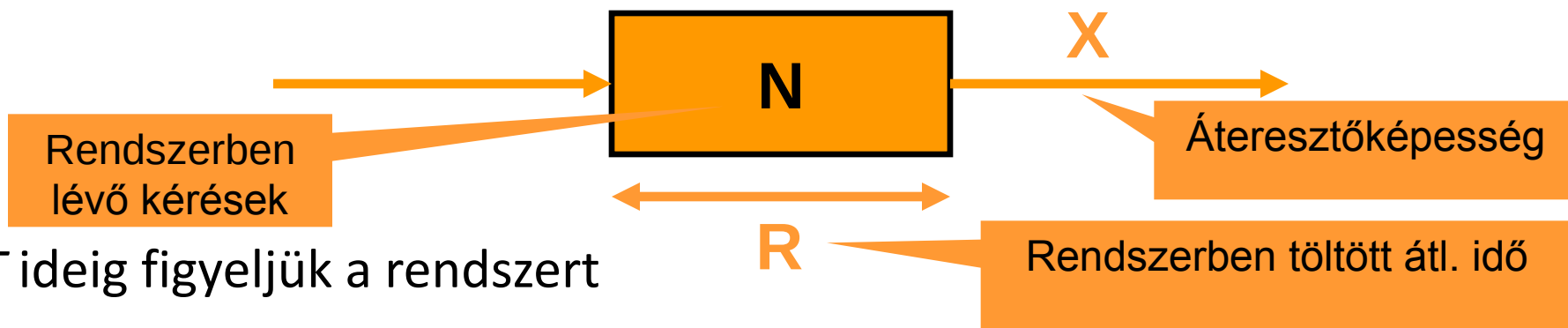
Szolgáltatás igény törvénye (Service Demand):

$$D_i = V_i \times S_i = (X_i/X_0) \times (U_i/X_i) = U_i/X_0$$

Forced Flow

Kihasználtság tv.

Little törvénye



T ideig figyeljük a rendszert

k rendszerben lévő kérések az intervallumban,

f_k az idő amíg k darab kérés van

a rendszerben

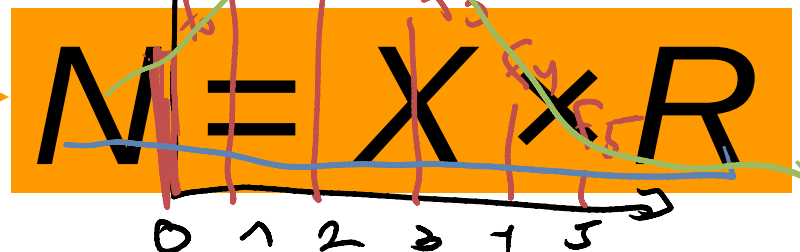
r_k a rendszerben töltött idő összege

C_0 : ennyi kérés hagyta el a rendszert

$$N = \sum_k k \times f_k = \sum_k k \times \frac{r_k}{T}$$

$$N = \frac{C_0}{T} \times \frac{\sum_k k \times r_k}{C_0}$$

R



A Little törvény alkalmazása

■ Levelezési szolgáltatást nyújtó portál

2.000.000 regisztrált felhasználó, 30%-uk küld levelet csúcsterhelésnél (1 óra hosszú). Egy email feldolgozása 5.0 másodperc, 3.5 levelet küld egy felhasználó. Egy levél mérete átlagosan 7120 bájt.

Mekkora spool fájl szükséges?

$$\begin{aligned} \text{LevelekÁtlagosSzám} &= \frac{\text{Áterestőképesség} \times \text{Válaszidő}}{3,600} \\ &= \frac{(2.000.000 \times 0.30 \times 3.5 \times 5.0)}{3,600} = 2916.67 \end{aligned}$$

Handwritten notes: "ahh felh." with arrows pointing to the user count and the 3.5 multiplier.

Átlagos levélméret alapján

$$\begin{aligned} \text{SpoolFileMéret} &= \text{LevelekÁtlagosSzám} \times \text{LevelekÁtlagosMérete} \\ &= 2916.67 \times 7120 \text{ bájt} = 20.767 \text{ Mb} \end{aligned}$$

A Little törvény alkalmazása 2.

■ Webes bróker, 3 rétegű architektúra

(Web szerver, Alkalmazás szerver, Adatbázis szerver)

- 1.1 millió felhasználó, 20000 használja a rendszert egyszerre (csúcsterhelés)
- A rendszer 3.6 millió kérést dolgoz fel óránként
- Minden kérés átlagosan 1.4 tranzakciót generál az adatbázis szerveren, ami a vállalati mainframe-en található
- A mainframe 11500 kérést dolgoz fel másodpercenként

■ Mennyi az átlagos válaszidő?

■ A válaszidő mekkora részét tölti a válasz a mainframe-en?

A Little törvény alkalmazása 3.

Az egész rendszer „fekete doboz”:

$$\begin{aligned}\text{ÁtlagosVálaszidő} &= \frac{\text{FelhasználókÁtlagosSzama}}{\text{RendszerÁteresztőképesség}} \\ &= \frac{20,000}{\frac{3,600,000}{3,600}} = 20 \text{ sec}\end{aligned}$$

kérdés *mp*

A mainframe a „fekete doboz”:

$$\begin{aligned}\text{ÁtlagosVálaszidő} &= \frac{\text{TranzakciókÁtlagosSzama}}{\text{MainframeÁteresztőképesség}} \\ &= \frac{20,000 \times 1.4}{11,500} = 2.43 \text{ sec}\end{aligned}$$

transzakció

A kettő aránya:

$$\frac{\text{MainframeHasználat}}{\text{RendszerHasználat}} = \frac{1.4 \times 2.43 \text{ sec}}{20.0 \text{ sec}} = 17\%$$

felh.

Könyvesbolt példa

■ Bemenő adatok:

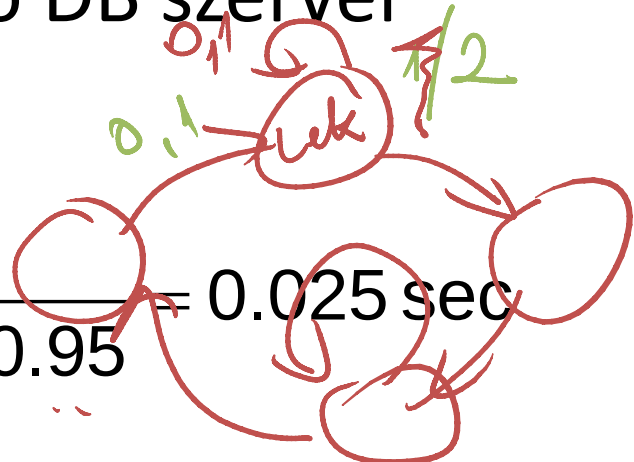
**Mérhető
értékek**

Adat	Érték
Összes session száma	35000
Web szerver foglaltság összesen (sec)	1200
DB szerver foglaltság összesen (sec)	2100
Mérési intervallum (sec)	3600
Keresések száma/Session	2,50
Web szerver kérés/Keresés	1,95
DB szerver kérés/Keresés	0,95

**Modellből kapjuk
(CBMG, CSID)**

■ Mennyi egy keresés tranzakció DB szerver szolgáltatási ideje?

$$S_i = \frac{\sum D_i}{\sum V_i} = \frac{2100}{35000 \times 2.5 \times 0.95} = 0.025 \text{ sec}$$



Könyvesbolt példa 2.

■ Web szerver és Adatbázis szerver

A Web szerver TCP fölött kommunikál, a kérés mérete 400 bájt, a válasz 9150 bájt. A hálózat 100BASE-T Ethernet.

Milyen a hálózat kihasználtsága?

$$X_0 = \frac{\text{KeresésekSzáma}}{\text{MérésIntervallum}} = \frac{35,000 \times 2.5}{3,600 \text{ sec}} = 24.3 \frac{1}{\text{sec}}$$

$$\text{CsomagokSzáma}_{\text{kérés}} = \left\lceil \frac{\text{ÜzenetMéret}_{\text{kérés}}}{\text{MaxSzegmensMéret}} \right\rceil = \left\lceil \frac{400}{1460} \right\rceil = 1$$

$$\text{CsomagokSzáma}_{\text{kválasz}} = \left\lceil \frac{\text{ÜzenetMéret}_{\text{válasz}}}{\text{MaxSzegmensMéret}} \right\rceil = \left\lceil \frac{9150}{1460} \right\rceil = 7$$

Könyvesbolt példa 3.

$$\begin{aligned}
 Szolgld\ddot{o}_{k\acute{e}r\acute{e}s} &= \frac{(\ddot{U}zenetM\acute{e}ret_{k\acute{e}r\acute{e}s} + Ovhd_{k\acute{e}r\acute{e}s}) \times 8}{10^6 \times S\acute{a}vsz\acute{e}less\acute{e}g} = \frac{(400 + 20 + 20 + 18) \times 8}{10^6 \times 100} \\
 &= 0.366 \text{ msec}
 \end{aligned}$$

netto Grútió
↑

$$Szolgld\ddot{o}_{v\acute{a}lasz} = \frac{[6 \times (1460 + 58) + 1 \times (390 + 58)] \times 8}{10^6 \times 100} = 7.64 \text{ msec}$$

$$\begin{aligned}
 U_n &= \sum_{\ddot{u}zenetek} \lambda_j \times Szolg\acute{a}ltat\acute{a}sld\ddot{o}_n^j \\
 &= 24.3 \times (0.366 + 7.64) 10^{-3} = 19.5\%
 \end{aligned}$$

Könyvesbolt példa 4.

■ Bemenő adatok:

**Mérhető
értékek**

Adat	Érték
Összes session száma	35000
Web szerver foglaltság összesen (sec)	1200
DB szerver foglaltság összesen (sec)	2100
Mérési intervallum (sec)	3600
Keresések száma/Session	2,50
Web szerver kérés/Keresés	1,95
DB szerver kérés/Keresés	0,95
Átlagos Web szerver szolg. igény (D_{WS}) msec	13,71
Átlagos DB szerver szolg. igény (D_{DB}) msec	24,00
Átlagos Web szerver szolg. idő (S_{WS}) msec	7,03
Átlagos DB szerver szolg. idő (S_{DB}) msec	25,26

Kiszámolható

**Modellből kapjuk
(CBMG, CSID)**

■ Változhat

- vásárlók száma
- vásárlók viselkedése
- rendszer paraméterek (pl. fájlok átlagos mérete)

Könyvesbolt példa 5.

Milyen lesz a szerverek kihasználtsága, ha megduplázódik a felhasználók száma (csúcsterhelés)?

- megduplázódik a Session-ök száma
- tfh. nem változik a felhasználók viselkedése

$$U_{WS} = X_0 \times D_{WS} = \frac{2 \times 35,000 \times 2.5}{3600} \times 0.01371 = 66.67\%$$

$$U_{DB} = X_0 \times D_{DB} = 48.61 \times 0.024 = 116.7\%$$

Fejleszteni kell az adatbáziszerveret!!!

Könyvesbolt példa 6.

Milyen lesz a hálózat kihasználtsága, ha a képek mérete megduplázódik?

Az egyszerűség kedvéért tfh. a keresés válaszának mérete is megduplázódik. Ha minden más változatlan (válasz 18300 bájt):

$$Szolgldő_{válasz} = \frac{[12 \times (1460 + 58) + 1 \times (780 + 58)] \times 8}{10^6 \times 100} = 15.24 \text{ msec}$$

$$\begin{aligned} U_n &= \sum_{\text{üzenetek}} \lambda_j \times Szolgáltatásldő_n^j \\ &= 24.3 \times (0.366 + 15.24) 10^{-3} = 36\% \end{aligned}$$

Könyvesbolt példa 7.

Milyen lesz a teljesítmény, ha megváltozik a felhasználók viselkedése és átlagosan 50%-kal többet keresnek az oldalon?

Átbocsátást és kihasználtságot számolunk.

$$X_0 = \frac{35,000 \times 3.75}{3,600} = 36,46 \frac{1}{\text{sec}}$$

$$U_{WS} = X_0 \times D_{WS} = 36.45 \frac{1}{\text{sec}} \times 0.01371 \text{ sec} = 49.97\%$$

$$U_{DB} = X_0 \times D_{DB} = 36.45 \frac{1}{\text{sec}} \times 0.024 \text{ sec} = 87.48\%$$

Kritikus!

Szűk keresztmetszet, skálázhatóság

- Szűk keresztmetszet:
 - azok az erőforrások, melyek korlátozzák a teljesítményt
 - analízis maximális terhelésnél (pesszimista érték)
- Nem mindig kell pontos szám
 - Aszimptotikus Határérték Analízis
(Asymptotic Bound Analysis)
- Fogalmak
 - K : erőforrások száma
 - D_i : szolgáltatásigény az i . erőforráson
 - $D_{\max}$: a legnagyobb szolg. igény
 - $D_{\min}$: a legkiseb szolg. igény
 - $\lambda_{\max}$: érkezési ráta

Nyílt modellek

- Nincs explicit korlát a rendszerben lévő kérésekre
- Ha az egyensúly teljesül:
- 3 rétegű architektúra példa

EZ MÁR

$$\lambda_{\max} \leq \frac{1}{D_{\max}}$$

NEVE
ZH

Egy tipikus E-Business szolgáltatás adatai			
Réteg	Szerverek száma	Szolg. átl. igénybevétele	Átl. szolg. idő
Web szerver	5	1,8	110 msec
Alkalmazás szerver	3	2,5	230 msec
DB szerver	2	2,3	180 msec



- Mi a szűk keresztmetszet?
- Mennyi a max. áteresztőképesség?

Nyílt modellek 2.

$$D_{WS} = \left(\frac{V_{WS}}{N_{WS}} \right) \times S_{WS} = \left(\frac{1.8}{5} \right) \times 0.110 = 0.0396 \text{ sec}$$

$$D_{AS} = \left(\frac{V_{AS}}{N_{AS}} \right) \times S_{AS} = 0.192 \text{ sec}$$

$$D_{DB} = \left(\frac{V_{DB}}{N_{DB}} \right) \times S_{DB} = 0.207 \text{ sec}$$

Az adatbázis szerver a szűk keresztmetszet

$$\lambda \leq \frac{1}{0.207 \text{ sec}} = 4.83 \frac{1}{\text{sec}}$$

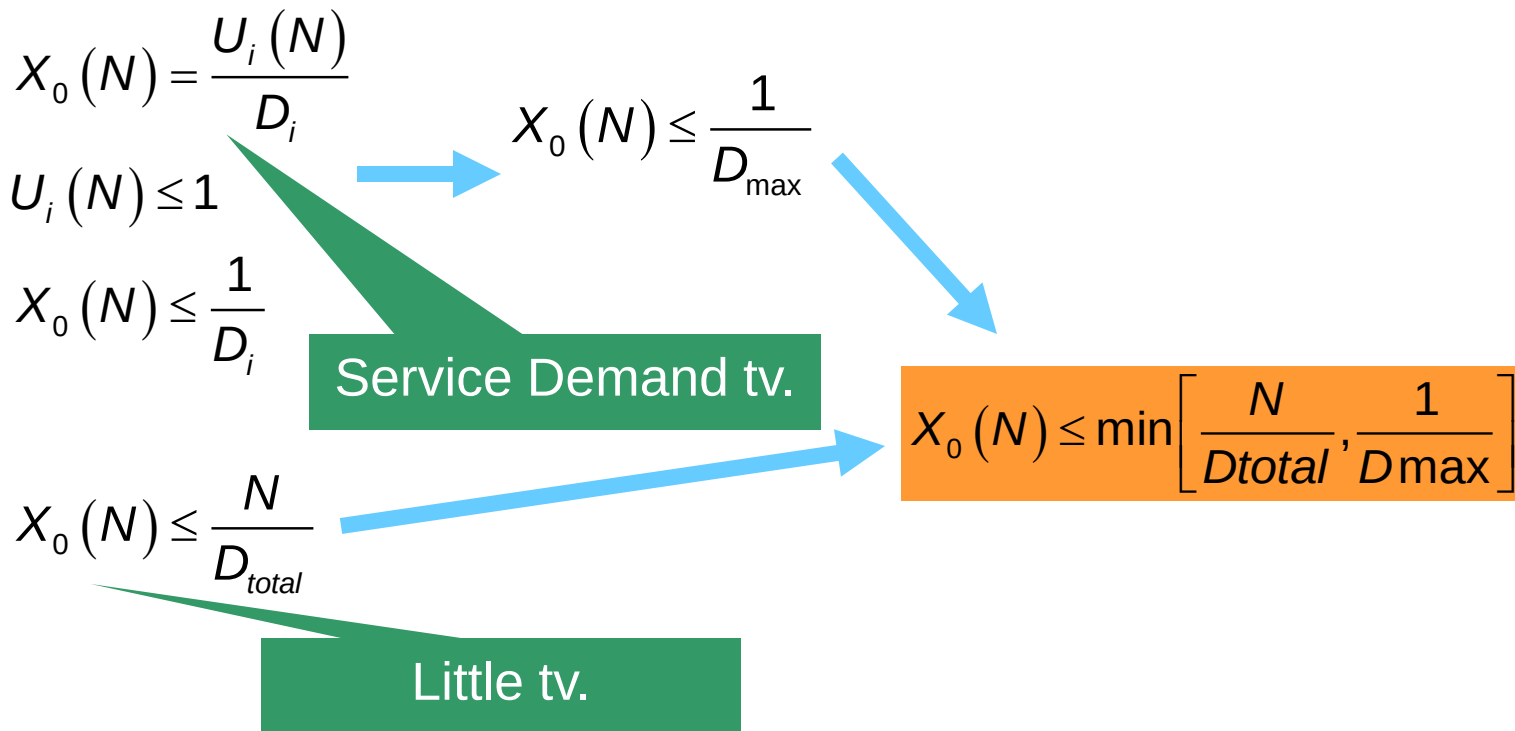
Tranzakciók max. gyakorisága

$$X_0 \leq \frac{X_{DB}}{V_{DB}} = \frac{4.83}{2.3} = 2.1 \frac{\text{szolgáltatás}}{\text{sec}}$$

E-Business szolgáltatások max. gyakorisága (Forced Flow)

Zárt modellek

- Felső korlát a kérések számára vonatkozóan (N)
- Ideális eset: nincs kérés várakozási sorban



Zárt modellek 2.

■ Példa: az előbbi 3 rétegű architektúra

Egyszerre max. 20 kliensnek nyújt szolgáltatást: $N = 20$

$$D_{total} = D_{WS} + D_{AS} + D_{DB} = 0.0396 \text{ sec} + 0.192 \text{ sec} + 0.207 \text{ sec} = 0.4386 \text{ sec}$$

$$X_0(20) \leq \min \left[\frac{20}{0.4386 \text{ sec}}, \frac{1}{0.207 \text{ sec}} \right]$$

$$X_0(20) \leq \min \left[45.59 \frac{1}{\text{sec}}, 4.83 \frac{1}{\text{sec}} \right]$$

$$X_0(20) \leq 4.83 \frac{1}{\text{sec}}$$