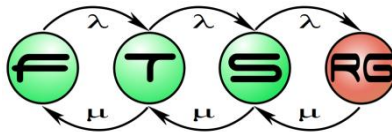
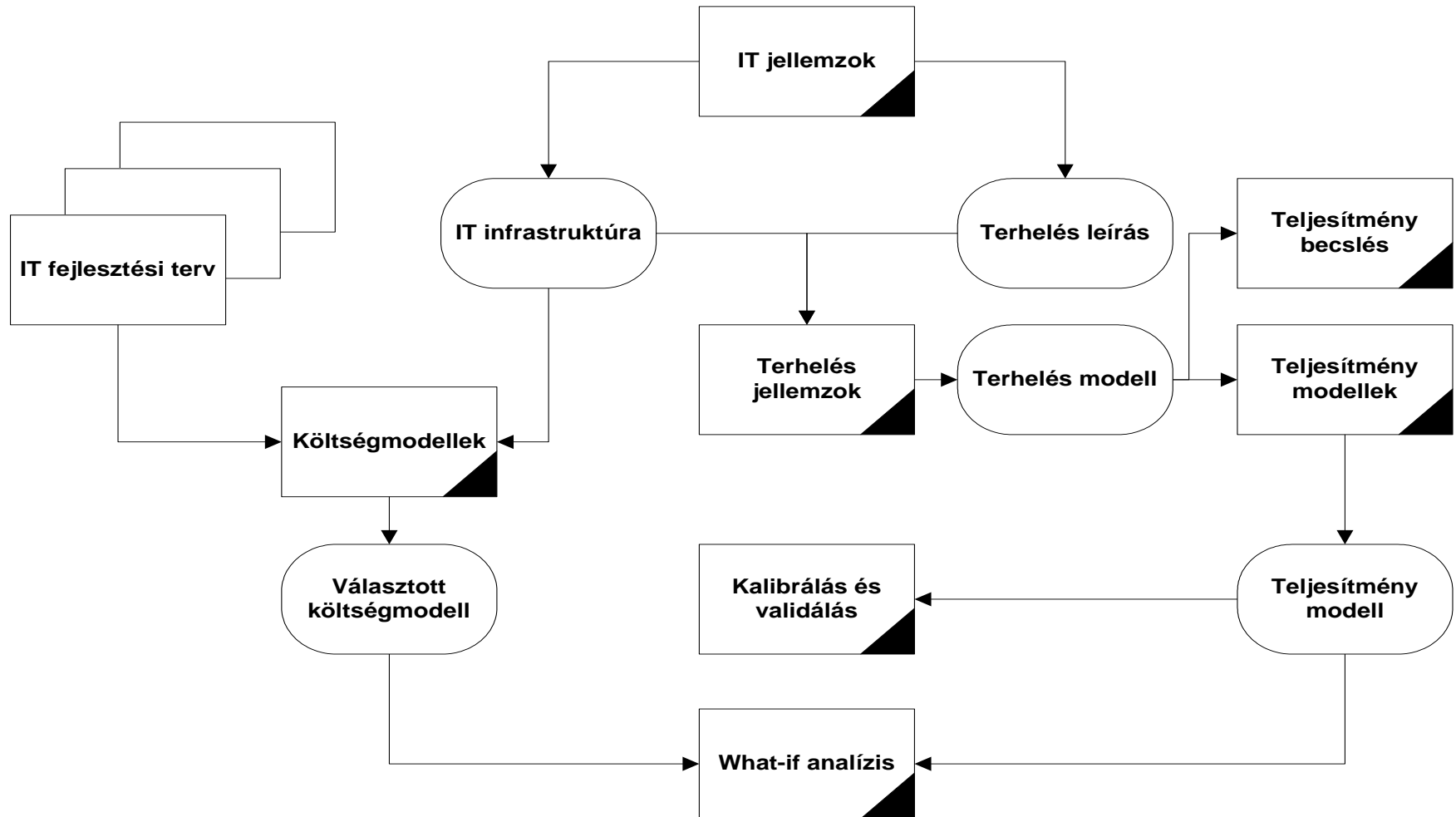


## Teljesítménymodellezés



# 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

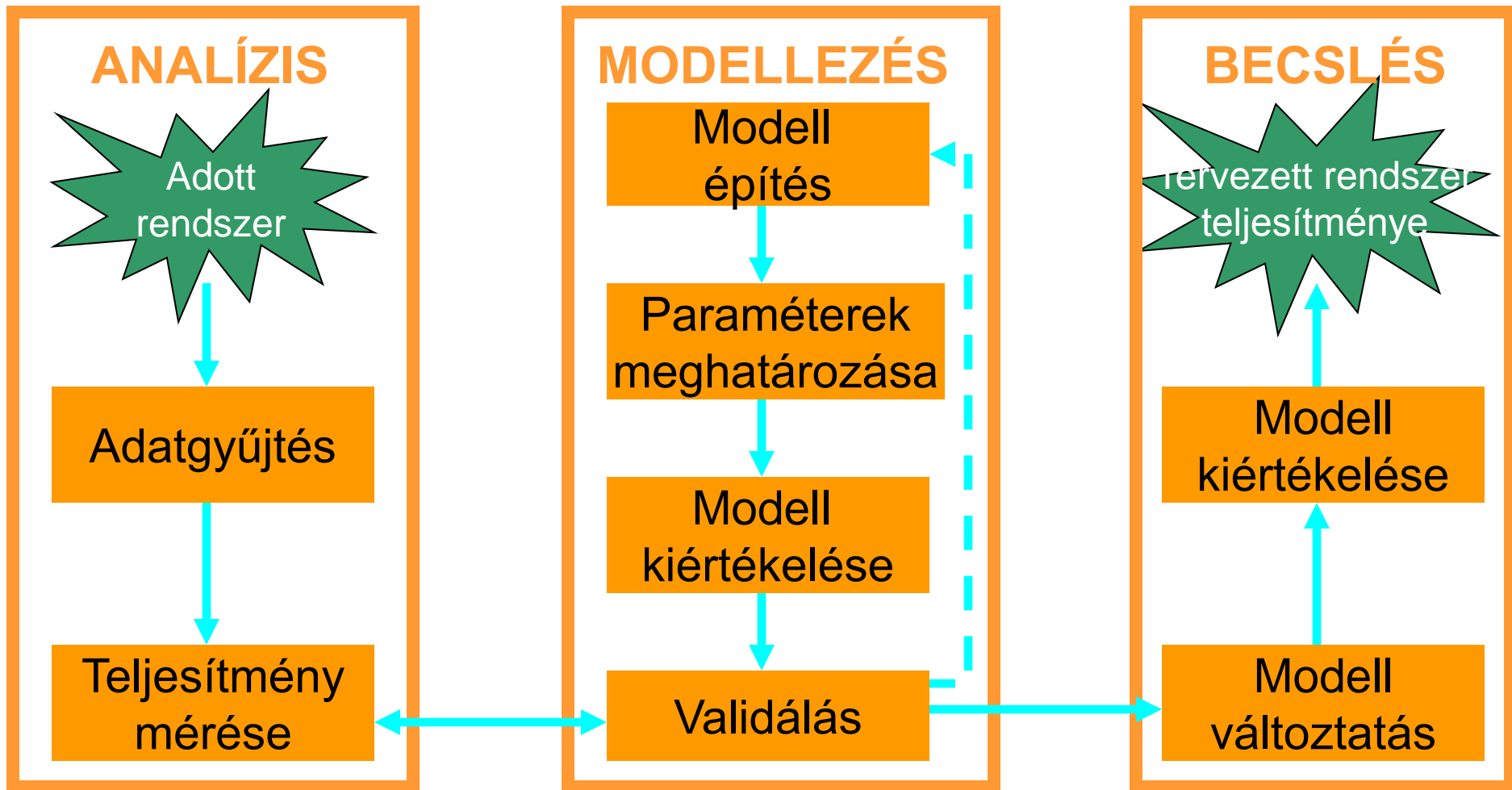
- a rendszert egyenletekkel írja le, pl.

$$RT_{\min} = RTT + \text{kérés}_{\min} + \text{Feldolgozásidő} + \text{válasz}_{\min}$$

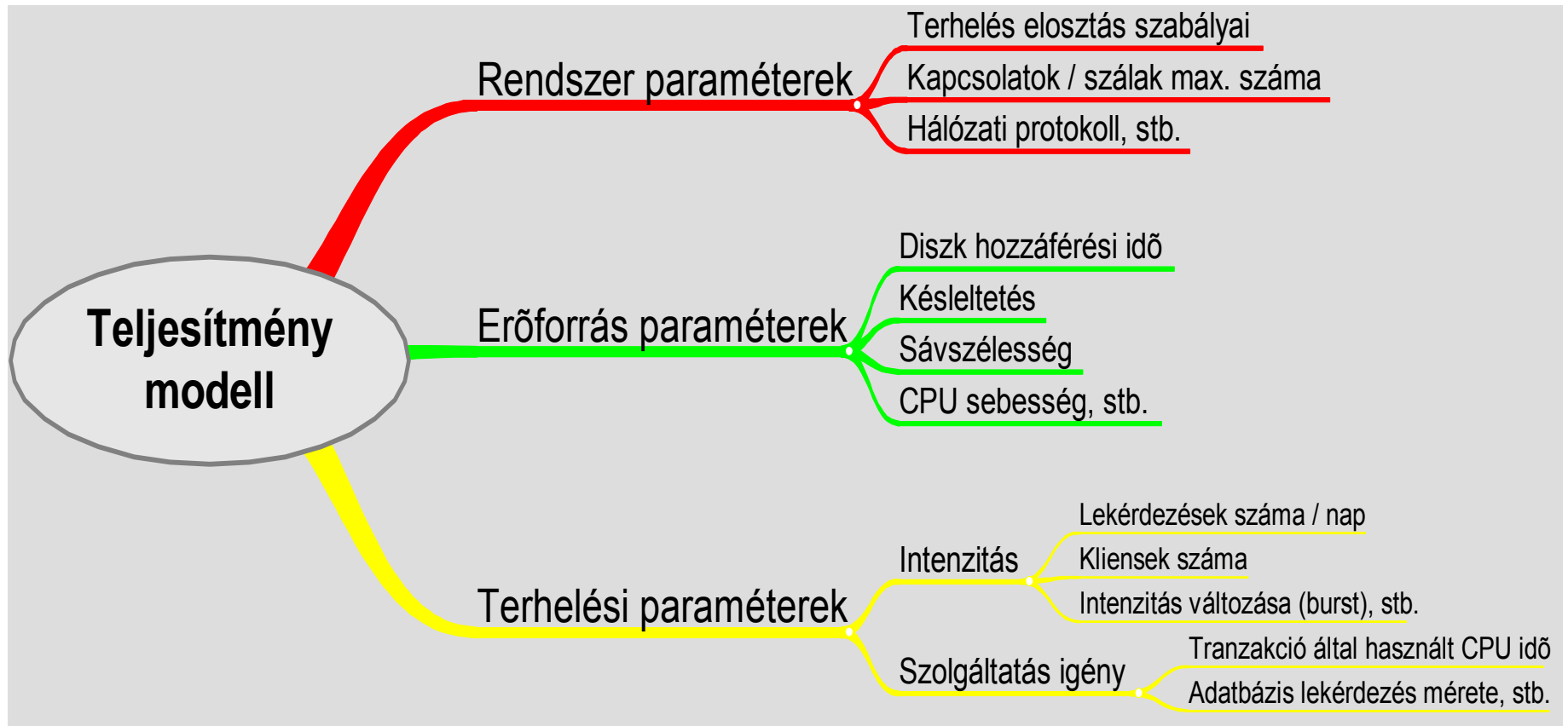
## ■ Szimulációs modell

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

# Modellezési/becslési paradigma



# Teljesítmény modell paramétere



# 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

$$D_i = V_i \times S_i$$

Könnyebben meghatározható

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

# Hálózati egyenletek

Egy üzenethez tartozó csomagok száma:

$$Csomagszám = \left\lceil \frac{\text{Üzenetméret}}{\text{MaxSzegmensMéret}} \right\rceil$$

Összes protokoll overhead (n. hálózaton):

$$Overhead_n = Csomagszám \times (TCPOvhd + IPOvhd + FrameOvhd_n)$$

ahol figyelembe vesszük a TCP, IP és a konkrét hálózati protokollból eredő overhead-et



# Hálózati egyenletek 2.

Üzenet továbbítási ideje az n. hálózatban

$$Szolgáltatásldő_n = \frac{(\text{Üzenetméret} + \text{Overhead}_n) \times 8}{10^6 \times \text{Sávszélesség}}$$

ahol a méret byte, a sávszélesség Mbps  
mértékegységű

Az n. hálózat kihasználtsága (utilization)

$$U_n = \sum_{\text{üzenetek}} \lambda_j \times Szolgáltatásldő_n^j$$

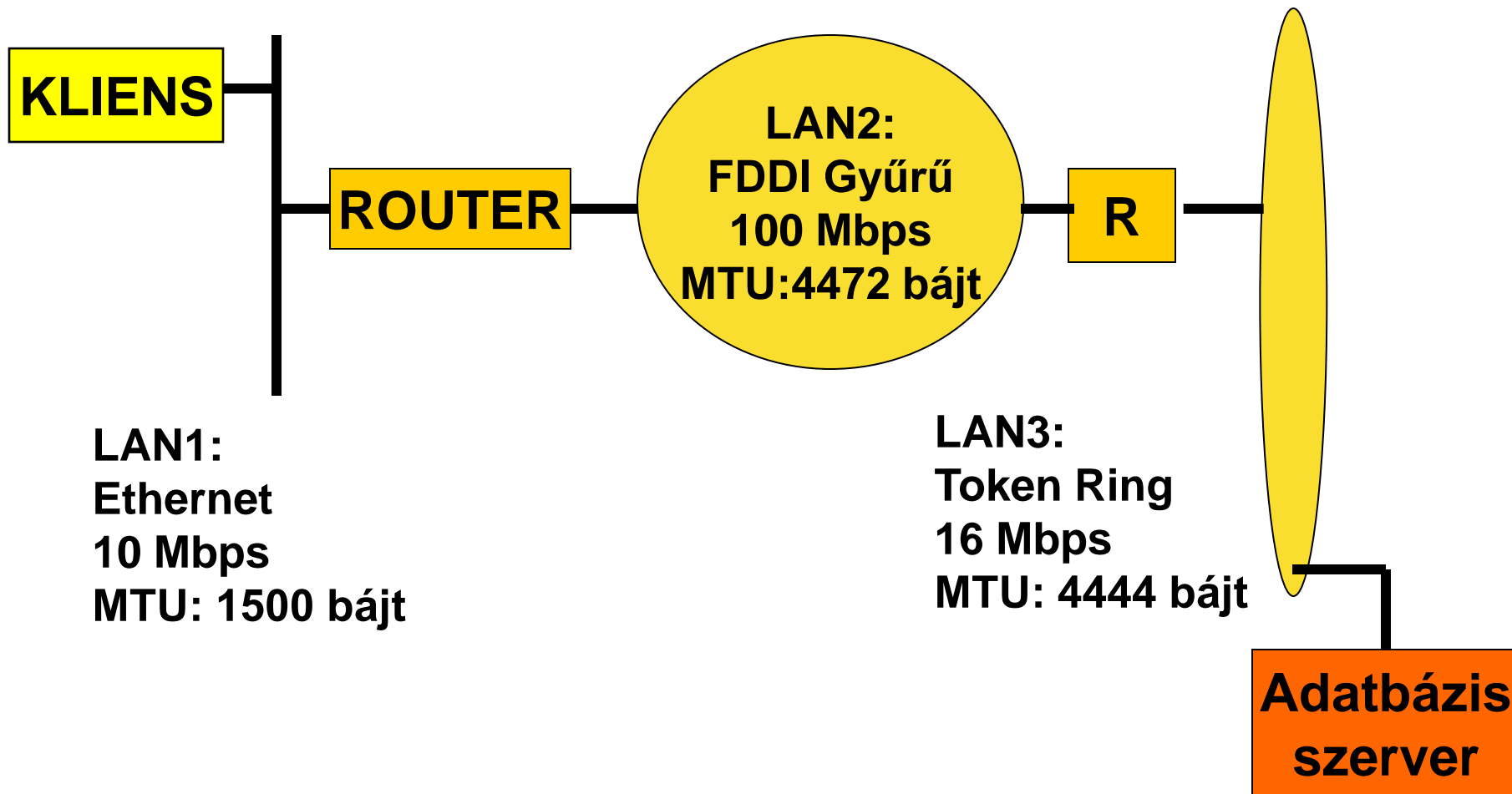
# Hálózati protokollok

Protokoll	PDU elnevezés	Max. PDU (bájt)	Overhead (bájt)	Max. Adatterület
TCP	segment	65,515	20	65,495
UDP	datagram	65,515	8	65,507
IPv4	datagram	65,535	20	65,515
IPv6	datagram	65,535	40	65,495
ATM	cell	53	5	48
Ethernet	frame	1,518	18	1,500
IEEE 802.3	frame	1,518	21	1,497
IEEE 802.5 Token Ring	frame	4,472	28	4,444
FDDI (RFC 1390)	frame	4,500	28	4,472

PDU = Protocol Data Unit

MTU=Maximum Transmission Unit

# Hálózati környezet példa



# Hálózati teljesítmény példa

Az előző hálózatban a kliens küld egy 300 bájtos kérést:

## LAN1:

*Overhead = TCPOvhd + IPOvhd + EthOvhd* (1 csomag)

$$Ovhd1 = 20 + 20 + 18 = 58$$

$$Szolgldő1 = \frac{358 \times 8}{10,000,000} = 0.000286sec$$

## LAN2:

$$Szolgldő2 = \frac{368 \times 8}{100,000,000} = 0.00002944sec$$

## LAN3:

$$Szolgldő3 = \frac{368 \times 8}{16,000,000} = 0.000184sec$$

# Hálózati teljesítmény példa 2.

- A szerver válaszol: 10000 bájt

szegmentálás: 1460 bájt az egység

*EthernetMéret* – (*TCPOvhd* + *IPOvhd*) = 1500 – 40

7 csomag : 10000 = (6\*1460 + 1240)

**LAN3:**

$$\frac{(6 \times 1,528 + 1,308) \times 8}{16,000,000} = 0.00524 \text{ sec}$$

**LAN2:**

$$\frac{[6 \times (1,460 + 20 + 20 + 28) + (1,240 + 20 + 20 + 28)] \times 8}{100,000,000} = 0.000838 \text{ sec}$$

**LAN1:**

$$\frac{[6 \times (1,460 + 20 + 20 + 18) + (1,240 + 20 + 20 + 18)] \times 8}{10,000,000} = 0.00832 \text{ sec}$$

# Hálózati teljesítmény példa 3.

- A kliens átlag percenként 3-szor lekérdezi az adatbázis szerveret

Lekérdezés: 400 bájt, Válasz: 80% 8092 bájt, 20% 100,000 bájt

Milyen a hálózati kihasználtság?

Csomagok száma:

Kérdés:

Rövid válasz:  $\lceil 4 / 1,460 \rceil = 1$

Hosszú válasz:  $\lceil 8,092 / 1,460 \rceil = 6$

$$\lceil 100,000 / 1,460 \rceil = 69$$

# Hálózati teljesítmény példa 4.

		<b>Kérés</b>	<b>Rövid válasz</b>	<b>Hosszú válasz</b>
<b>LAN1</b>	Csomagok száma	1	6	69
	Overhead (bájt)	58	348	4002
	Átviteli idő (ms)	0.366	6.75	83.2
<b>LAN2</b>	Csomagok száma	1	6	69
	Overhead (bájt)	68	408	4692
	Átviteli idő (ms)	0.0374	0.68	8.38
<b>LAN3</b>	Csomagok száma	1	6	69
	Overhead (bájt)	68	408	4692
	Átviteli idő (ms)	0.234	4.25	52.3

$$0.8 \times (\text{KérésÁtvldő} + \text{RövidVálaszÁtvldő}) + 0.2 \times (\text{KérésÁtvldő} + \text{HosszúVálaszÁtvldő})$$

# Hálózati teljesítmény példa 5.

	Átlagos átviteli idő (s)		
	LAN1	LAN2	LAN3
	0.0224	0.00223	0.0141
Kliensek száma	Kihhasználtság (%)		
	LAN1	LAN2	LAN3
40	4.5	0.4	2.8
80	9	0.9	5.6
120	13.4	1.4	8.5
160	17.9	1.8	11.3
200	22.4	2.3	14.1
240	26.9	2.7	16.9
280	31.4	3.2	19.7

Hasonlóan kiszámítható, ha megváltozik a kérések intenzitása.



# 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

$\lambda_i$ : érkezési ráta

*Egyensúly* :  $\lambda_i = X_i$

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

# További törvények

Forced Flow törvény:

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

$V_i$ : „látogatások” átlagos száma

$X_0$ : tranzakciók átlagos száma

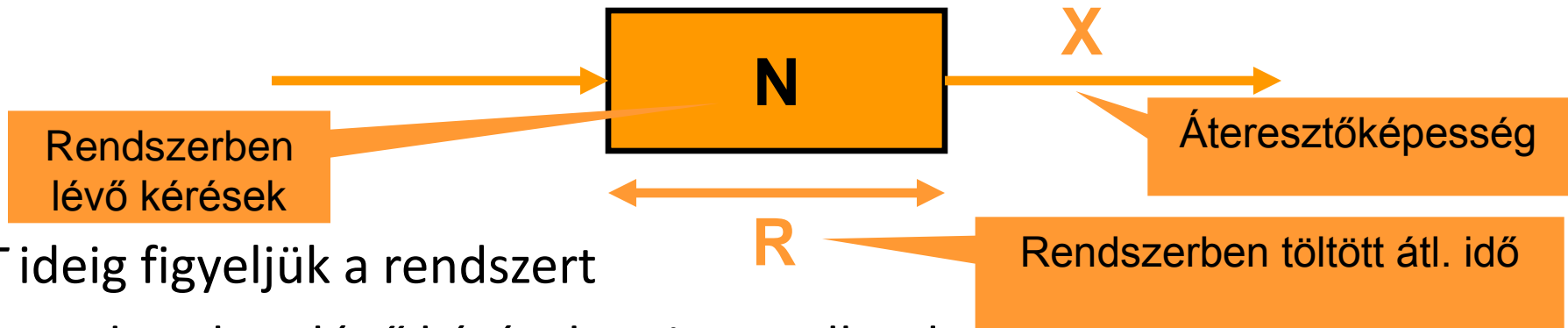
Szolgáltatás igény törvénye (Service Demand):

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

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

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

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

$R$

$$N = X \times R$$

# 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?

$$\begin{aligned} \text{LevelekÁtlagosSzama} &= \text{Áteresztőképesség} \times \text{Válaszidő} \\ &= \frac{(2,000,000 \times 0.30 \times 3.5 \times 5.0)}{3,600} = 2916.67 \end{aligned}$$

Átlagos levélméret alapján

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

# 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”:

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

A mainframe a „fekete doboz”:

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

A kettő aránya:

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

# 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.

$$\begin{aligned} \text{Szolgldő}_{\text{kérés}} &= \frac{(\text{ÜzenetMéret}_{\text{kérés}} + \text{Ovhd}_{\text{kérés}}) \times 8}{10^6 \times \text{Sávszélesség}} = \frac{(400 + 20 + 20 + 18) \times 8}{10^6 \times 100} \\ &= 0.366 \text{ msec} \end{aligned}$$

$$\text{Szolgldő}_{\text{válasz}} = \frac{[6 \times (1460 + 58) + 1 \times (390 + 58)] \times 8}{10^6 \times 100} = 7.64 \text{ msec}$$

$$\begin{aligned} U_n &= \sum_{\text{üzenetek}} \lambda_j \times \text{Szolgáltatásldő}_n^j \\ &= 24.3 \times (0.366 + 7.64) 10^{-3} = 19.5\% \end{aligned}$$

# 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 adatbáziszszerveret!!!

# 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 bájtt):

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

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

# 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:

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

- 3 rétegű architektúra példa

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

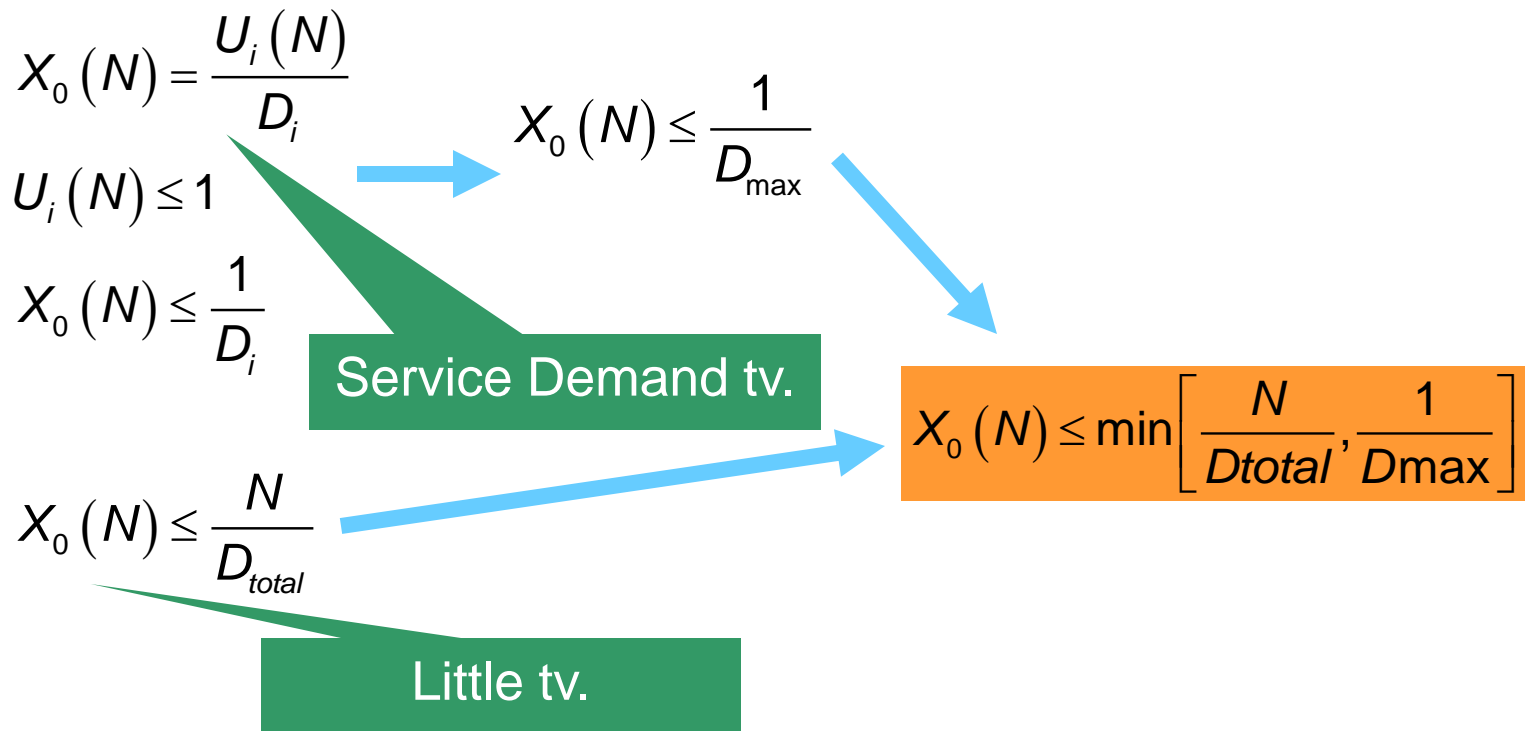
$$X_0 \leq \frac{X_{DB}}{V_{DB}} = \frac{4.83}{2.3} = 2.1 \frac{\text{szolgáltatás}}{\text{sec}}$$

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:  $N = 20$

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

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

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

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