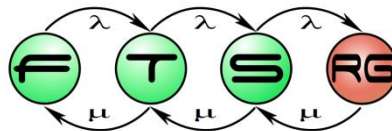


# Teljesítménymodellezés

**Budapest University of Technology and Economics  
Fault Tolerant Systems Research Group**



# Egy ismerős példa...

Budapesti Műszaki és Gazdaságtudományi Egyetem



Hallgatói BME\_W2K8H\_02(168)

Nyelv:    

Azonosító:

Jelszó:

Bejelentkezés >



Build: 430 (2014.11.11.) P20150304

Támogatott böngészők:

Microsoft Internet Explorer 9.0+ ; Mozilla Firefox ; Google Chrome

## Friss hírek

[KTH hallgatói hírlevél 2014/15/1 félévzárásról 1. – vizsgajelentkezés, vizsgaidőszak](#)

(2014.11.27.  
8:56:21)


Kedves Hallgató!

2014. december 1-jén 18 órakor indul a vizsgajelentkezés. Milyen egyéb határidőkre kell figyelnie a félév végén? Melyek az ebben az időszakban aktuális Neptun kérvények? Mit

## Letölthető dokumentumok

 Tantárgyfelvétel: eddig a legsimábban.pdf  
(2015.02.07. 14:28:14)

 Tantárgyfelvétel: nem és nem.pdf  
(2014.08.28. 21:19:50)

 Vizsgajelentkezés: szokványosan.pdf  
(2014.05.06. 21:14:04)

# Egy ismerős példa...

Kiszolgálóhiba történt az alkalmazásban: „/hallgato”.

## Futásidejű hiba

**Leírás:** Alkalmazáshiba történt a kiszolgálón. Az alkalmazás jelenlegi egyéni hibakezelési beállításai (biztonsági okok miatt) nem engedik meg az adatainak távoli megjelenítését. A böngészőben azonban megtekinthetők.

**Részletek:** Ha távoli gépeken is meg szeretné jeleníteni a hibaüzenet részletes adatait, hozzon létre a konfigurációs fájlban elhelyezett „web.config” címke „mode” attribútumát állítsa „Off” értékre.

```
<!-- Web.Config konfigurációs fájl -->
```

```
<configuration>  
  <system.web>  
    <customErrors mode="Off"/>  
  </system.web>  
</configuration>
```

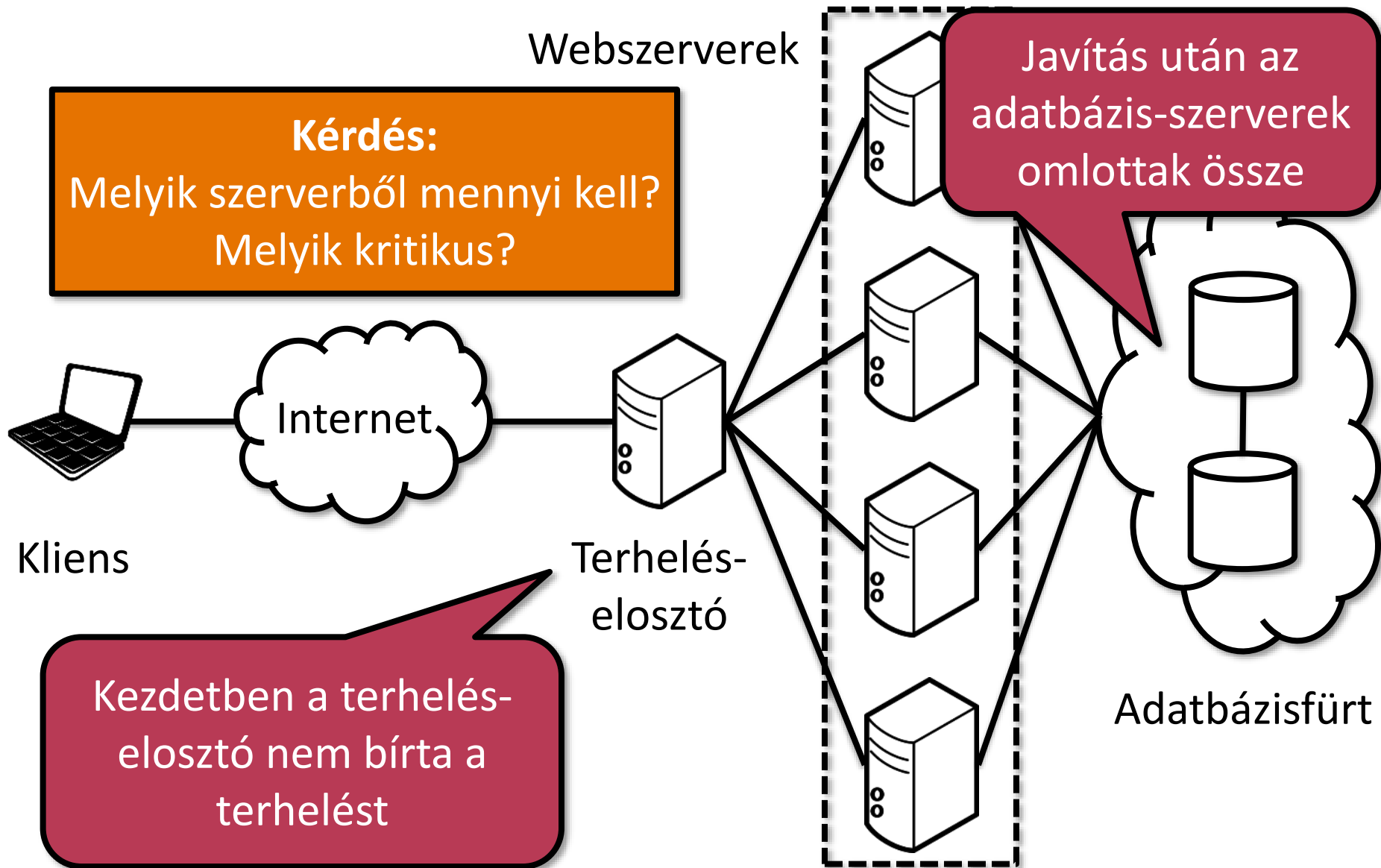
**Megjegyzések:**

```
<!-- Web.Config  
<configuration>  
  <system.web>  
    <customErr  
  </system.web>  
</configuration>
```

**Motiváció:  
Már tervezési időben  
készüljünk a terhelésre!**



# A Neptun vázlatos felépítése



# Terhelésmodellezés

- Emlékeztető: nemfunkcionális követelmények
  - Teljesítmény, átbecsátóképesség, stb.
  - Hogyan ellenőrizzük őket a rendszer megépítése nélkül?
- **Terhelésmodellezés:**
  - Az eddigi modellek kiegészítése időzítéssel, erőforrásokkal, kapacitáskorlátokkal
  - Célja:
    - A rendszer teljesítményének értékelése a tervezési fázisban
    - Kritikus részek azonosítása
    - Skálázás, méretezés



Alapfogalmak

Terhelési diagram

Erőforrásmodellezés

# TARTALOM

Alapfogalmak

Terhelési diagram

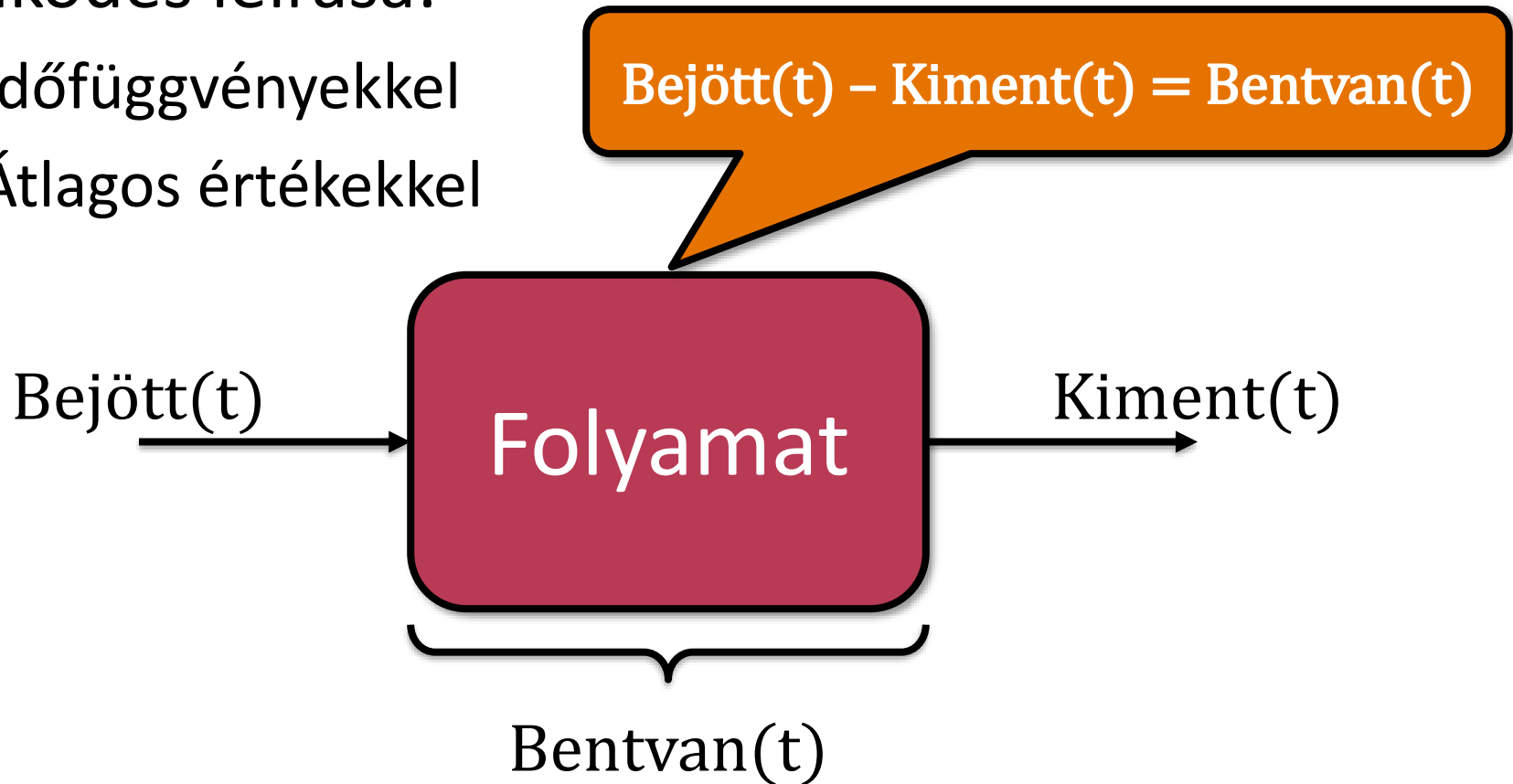
Erőforrásmodellezés

# ALAPFOGALMAK

Érkezési ráta, átbocsátás

# Alapmodell

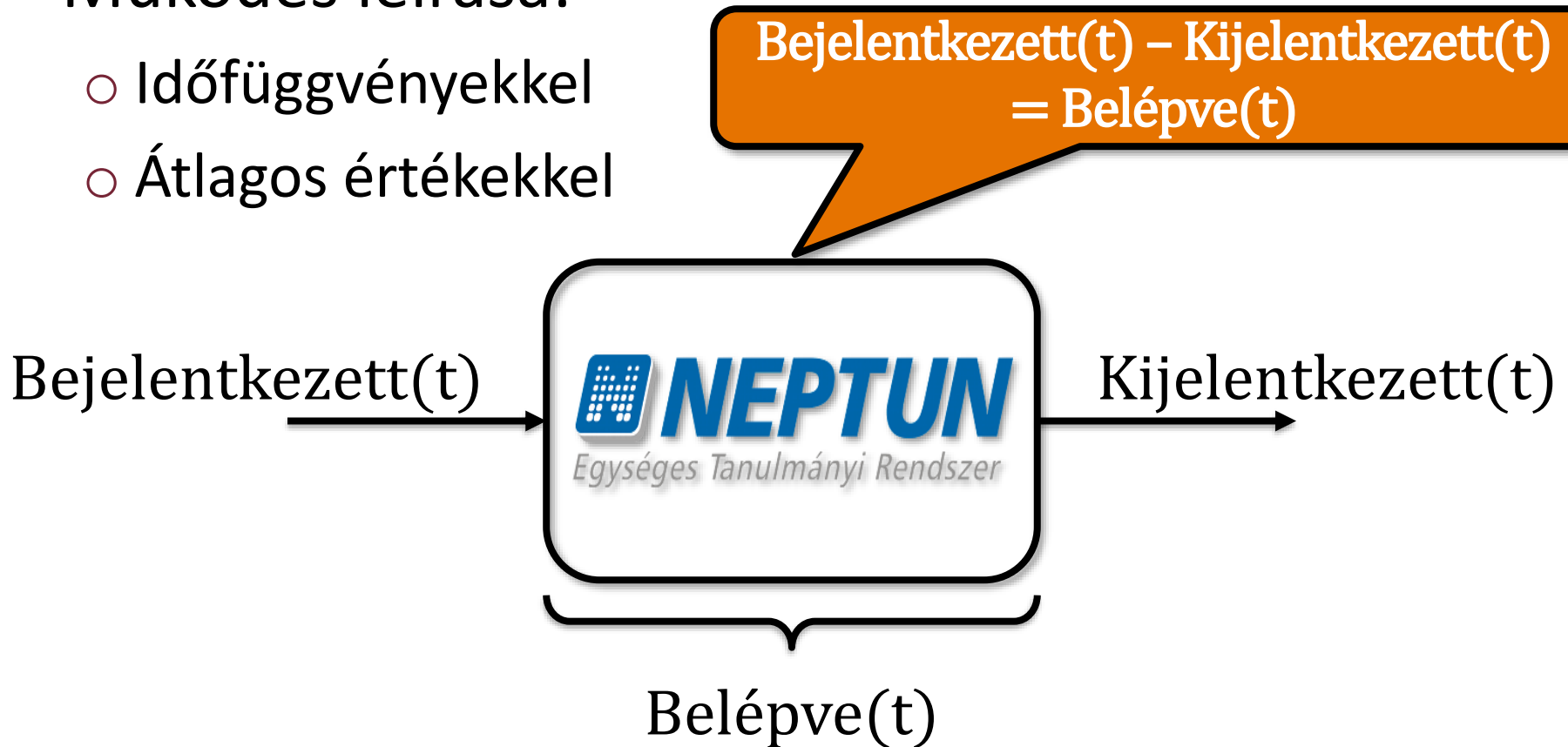
- Folyamatmodell végrehajtása sok kérésre
  - Vizsgálat tárgya: időfüggő viselkedés
- Működés leírása:
  - Időfüggvényekkel
  - Átlagos értékekkel





# Alapmodell

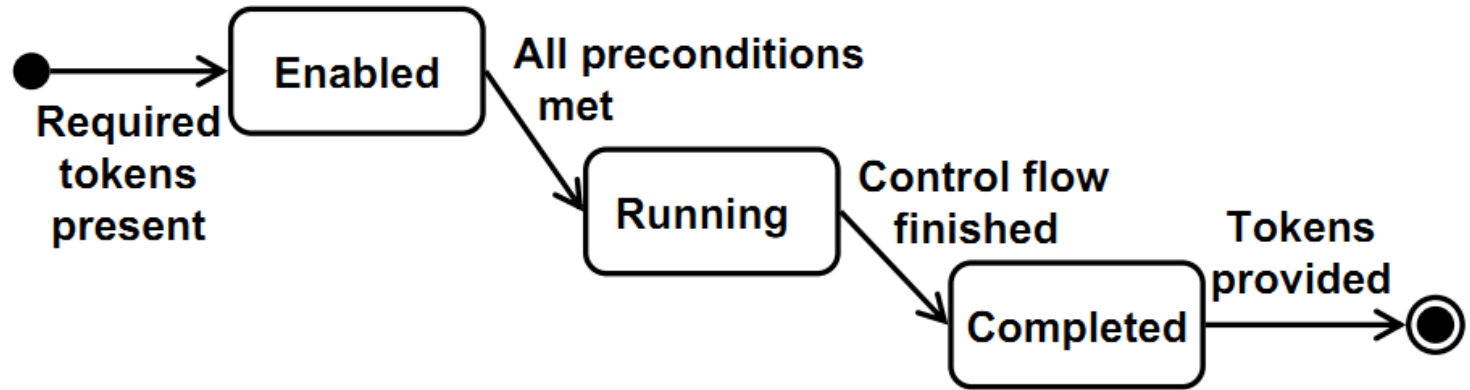
- Folyamatmodell végrehajtása sok kérésre
  - Vizsgálat tárgya: időfüggő viselkedés
- Működés leírása:
  - Időfüggvényekkel
  - Átlagos értékekkel



# Emlékeztető: végrehajtási állapotok

- Folyamat/tevékenység végrehajtásának állapota:

## State Machine



- Bejött(t): „*Enabled*” állapotban lévő tokenek
- Bentvan(t): „*Running*” állapotban lévő tokenek
- Kiment(t): „*Completed*” állapotban lévő tokenek

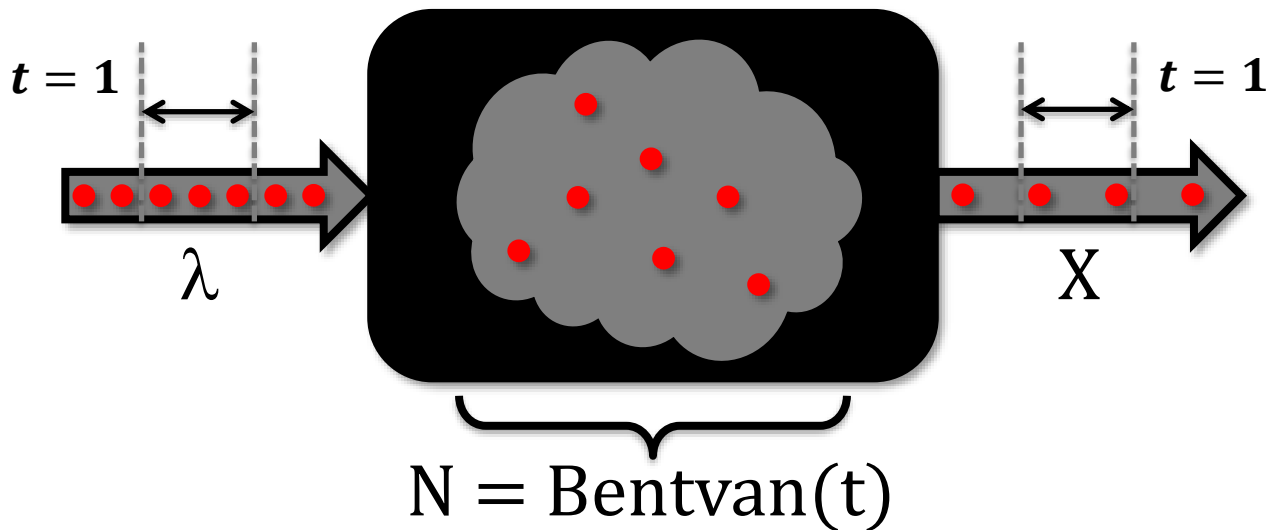
# Definíció: ráta, átbocsátás

**Érkezési ráta:** egységnyi idő alatt **érkező** kérések

$$\lambda = \frac{Bejött(t)}{t} \quad [\lambda] = \frac{1}{s}$$

**Átbocsátás:** egységnyi idő alatt **feldolgozott** kérések

$$X = \frac{Kiment(t)}{t} \quad [X] = \frac{1}{s}$$

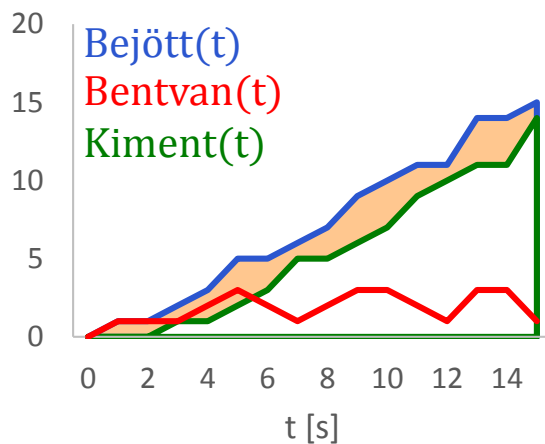


# Definíció: egyensúlyi állapot

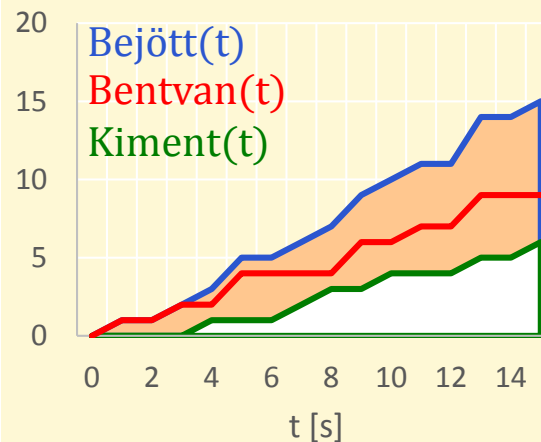
- **Egyensúlyi állapot:  $Bentvan(t)$  közel állandó**
  - Átlagos értékek csak ilyenkor használhatók!
  - Ilyenkor:

$$\lambda = X$$

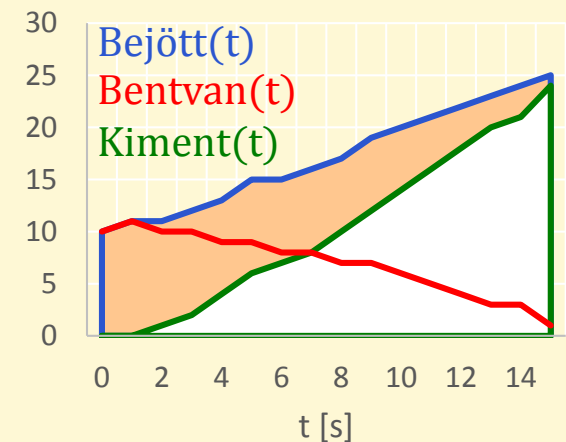
$$\lambda = X$$



$$\lambda > X$$



$$\lambda < X$$



# Definíció: egyensúlyi állapot

- **Egyensúlyi állapot:** *Bentvan(t)* közel állandó
  - Átlagos értékek csak ilyenkor használhatók!
  - Ilyenkor:

$$\lambda = X$$

**Egyensúlyi állapotban:**  
Percenként ugyanannyian lépnek be, mint ki



$$N = \text{Belépve}(t)$$

# Korlátos kapacitás – DDoS

- Idáig N akár végtelen is lehetett
- Mi van, ha véges?

## Denial of Service Attack

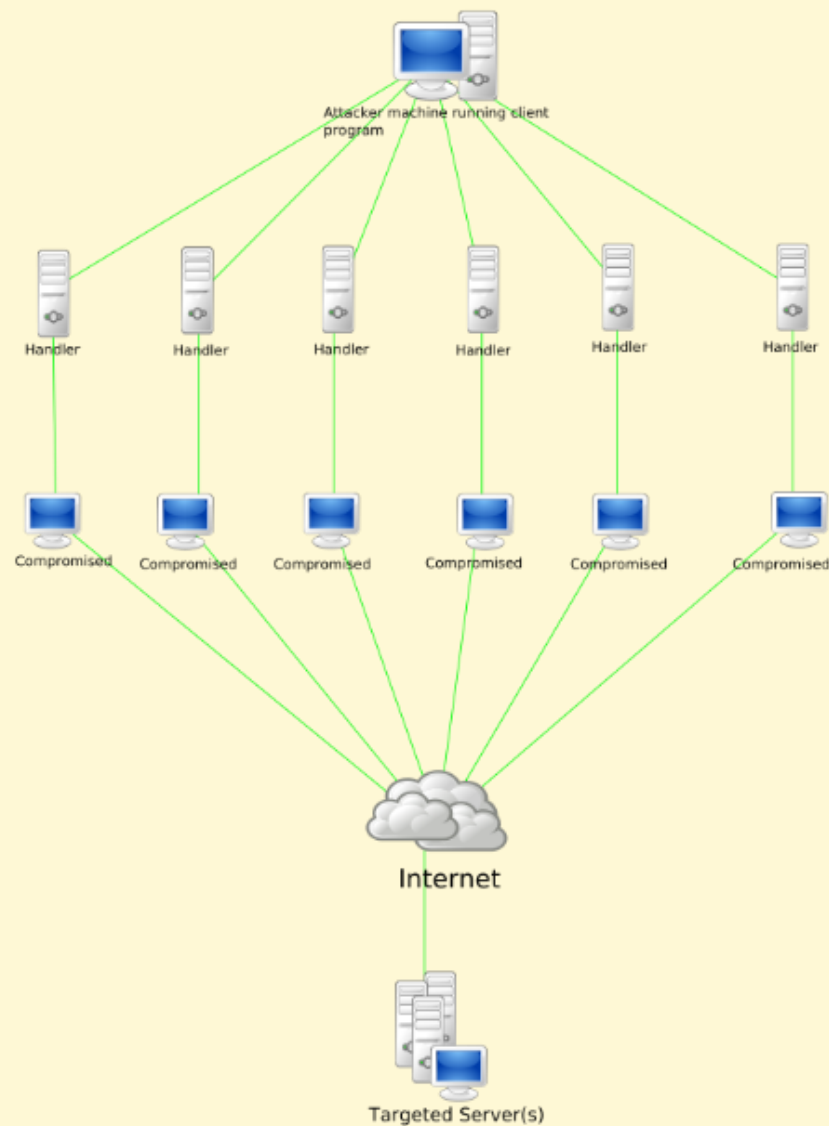
Thursday, August 6, 2009 | By Biz Stone (@biz) 08/06/2009 - 15:00

Tweet

On this otherwise happy Thursday morning, Twitter is the target of a [denial of service attack](#). Attacks such as this are malicious efforts orchestrated to disrupt and make unavailable services such as online banks, credit card payment gateways, and in this case, Twitter for intended customers or users. We are defending against this attack now and will continue to update our [status blog](#) as we continue to defend and later investigate.

# Distributed Denial of Service (DDoS)

- Tömeges kéresgenerálás  
→ rendszer lefoglalása
- A leterhelt rendszert könnyebb támadni
- Teljes szolgáltatások leállíthatóak
- Az Anonymous kedvelt módszere



Alapfogalmak

Terhelési diagram

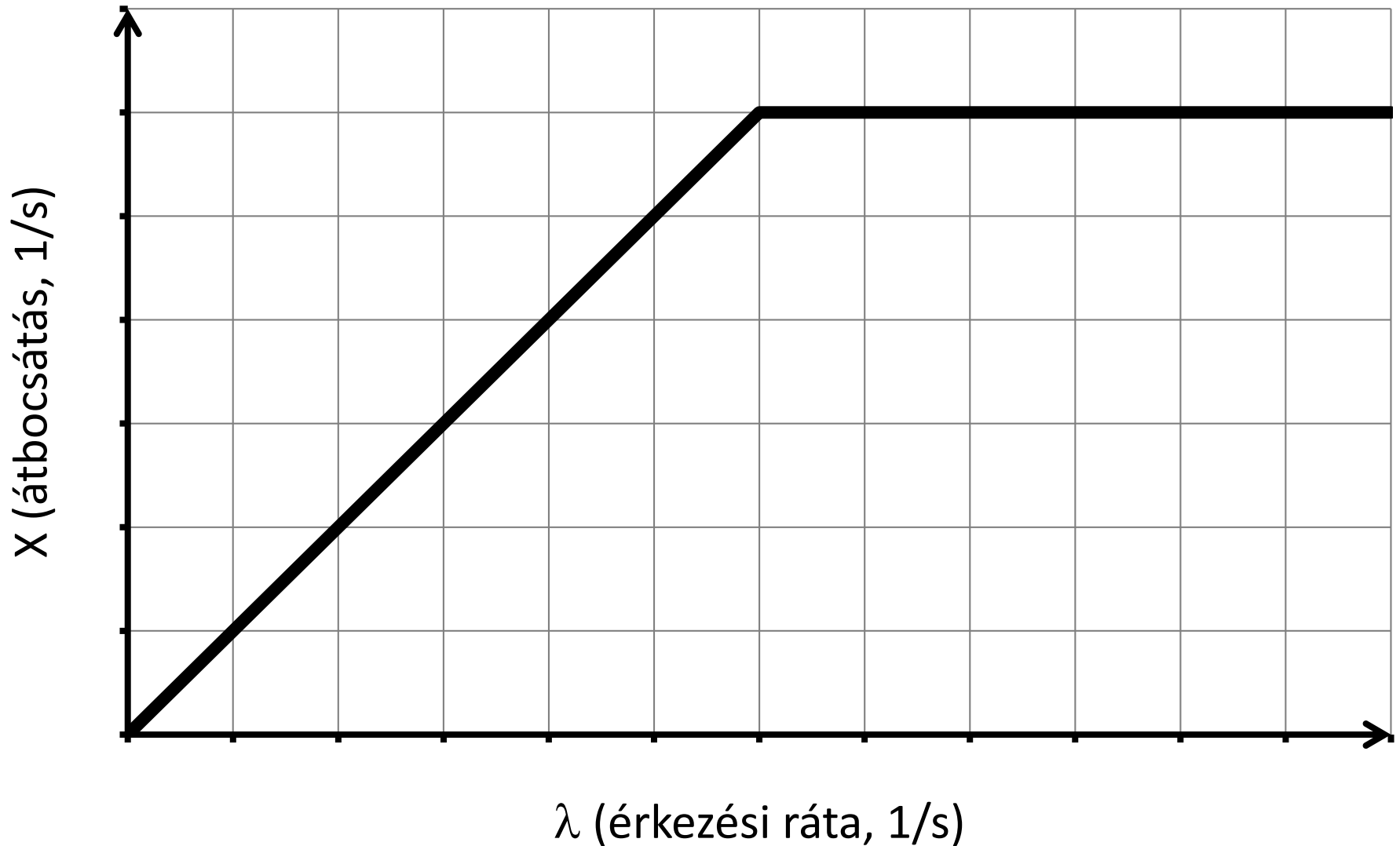
Erőforrásmodellezés

# TERHELÉSI DIAGRAM

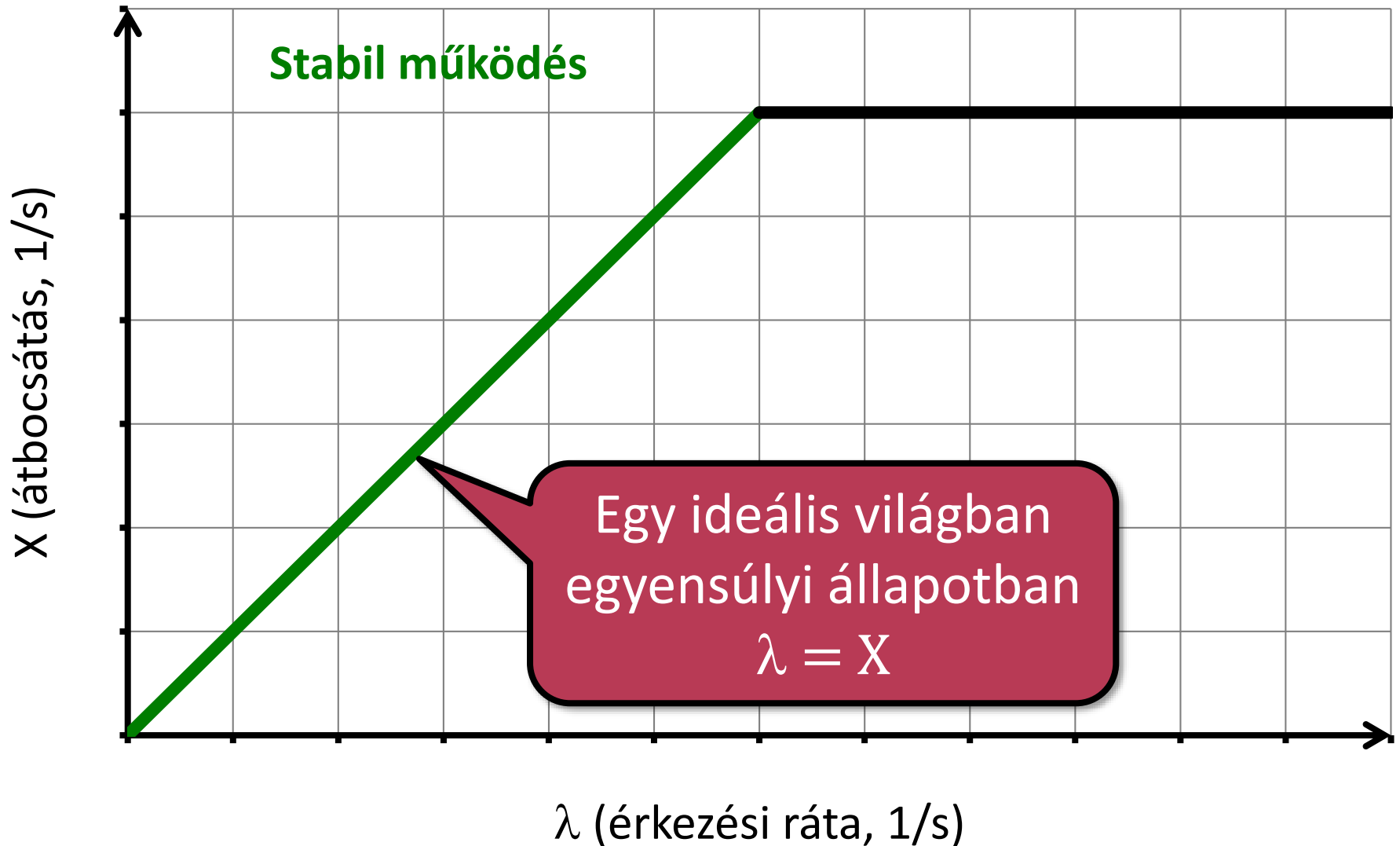
Az érkezési ráta és az átbocsátás kapcsolata,  
átbocsátóképesség



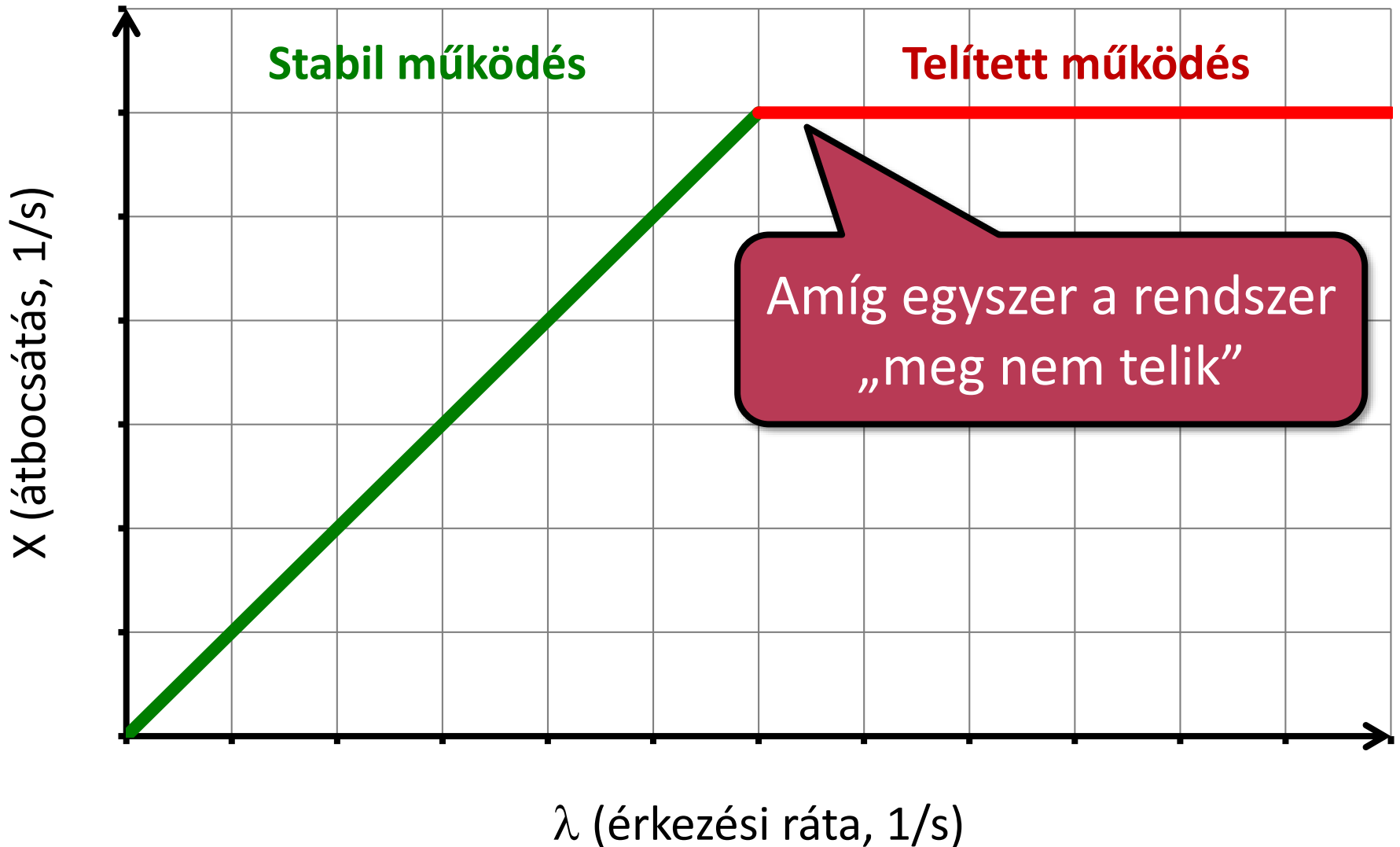
# Terhelési diagram



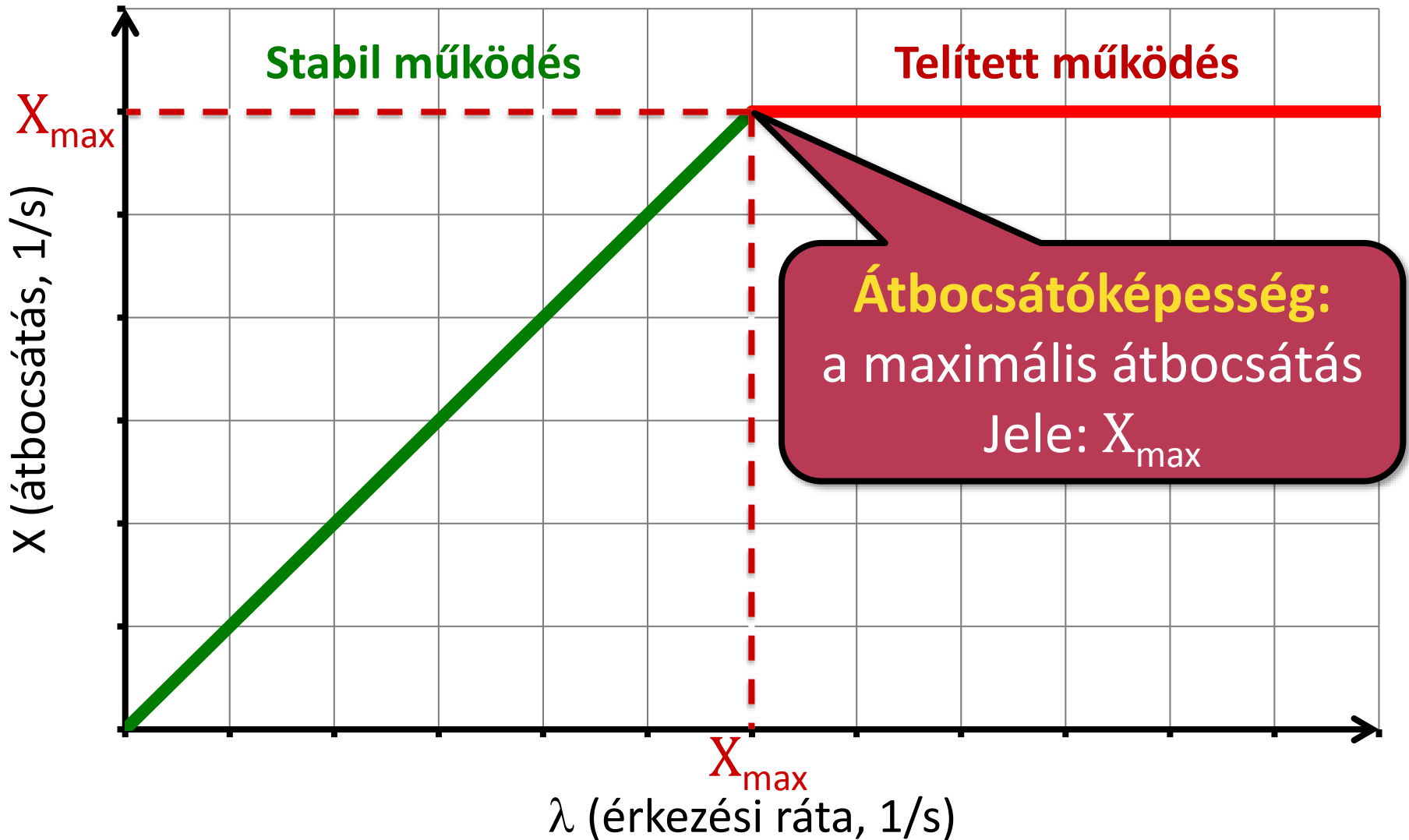
# Terhelési diagram



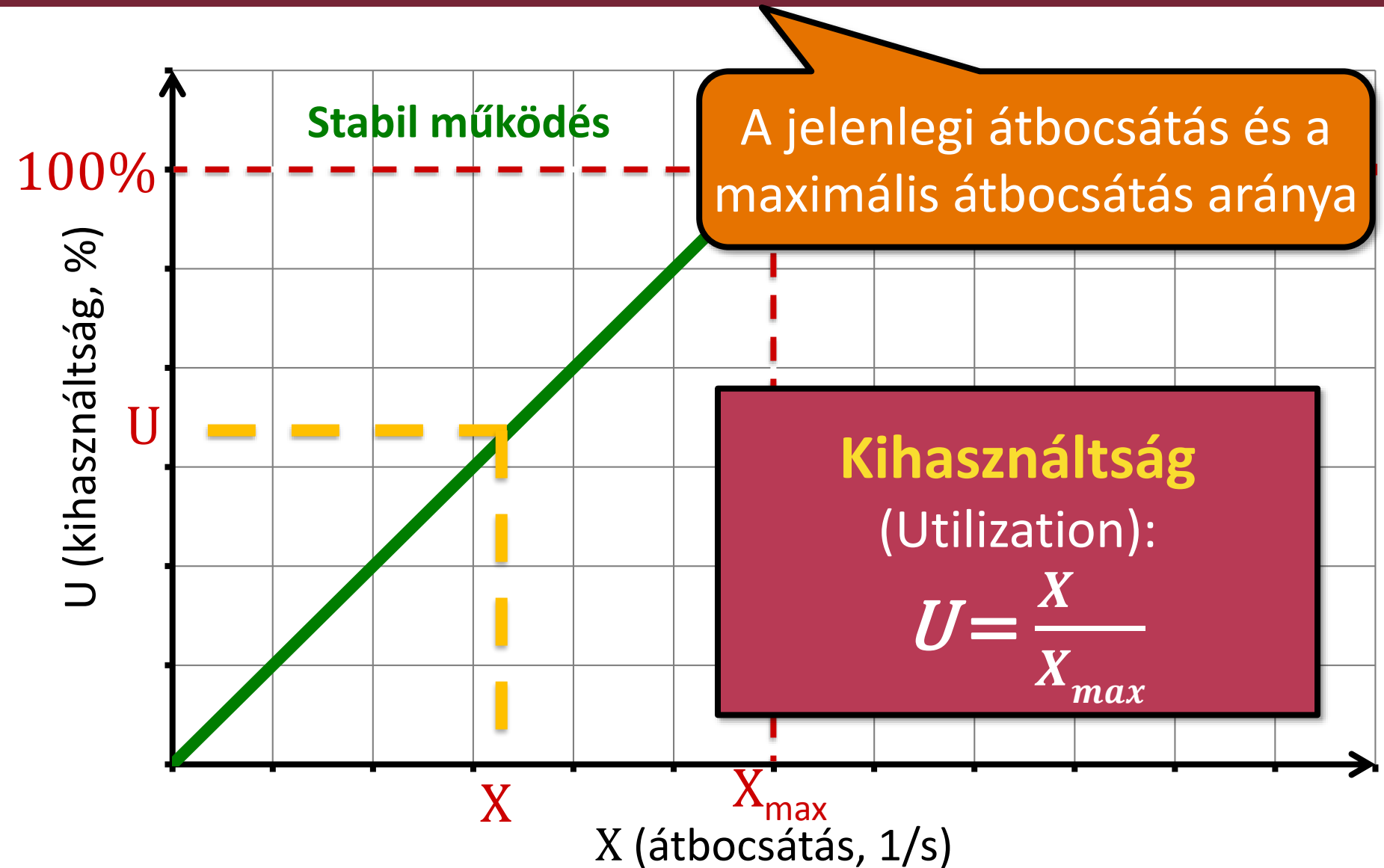
# Terhelési diagram



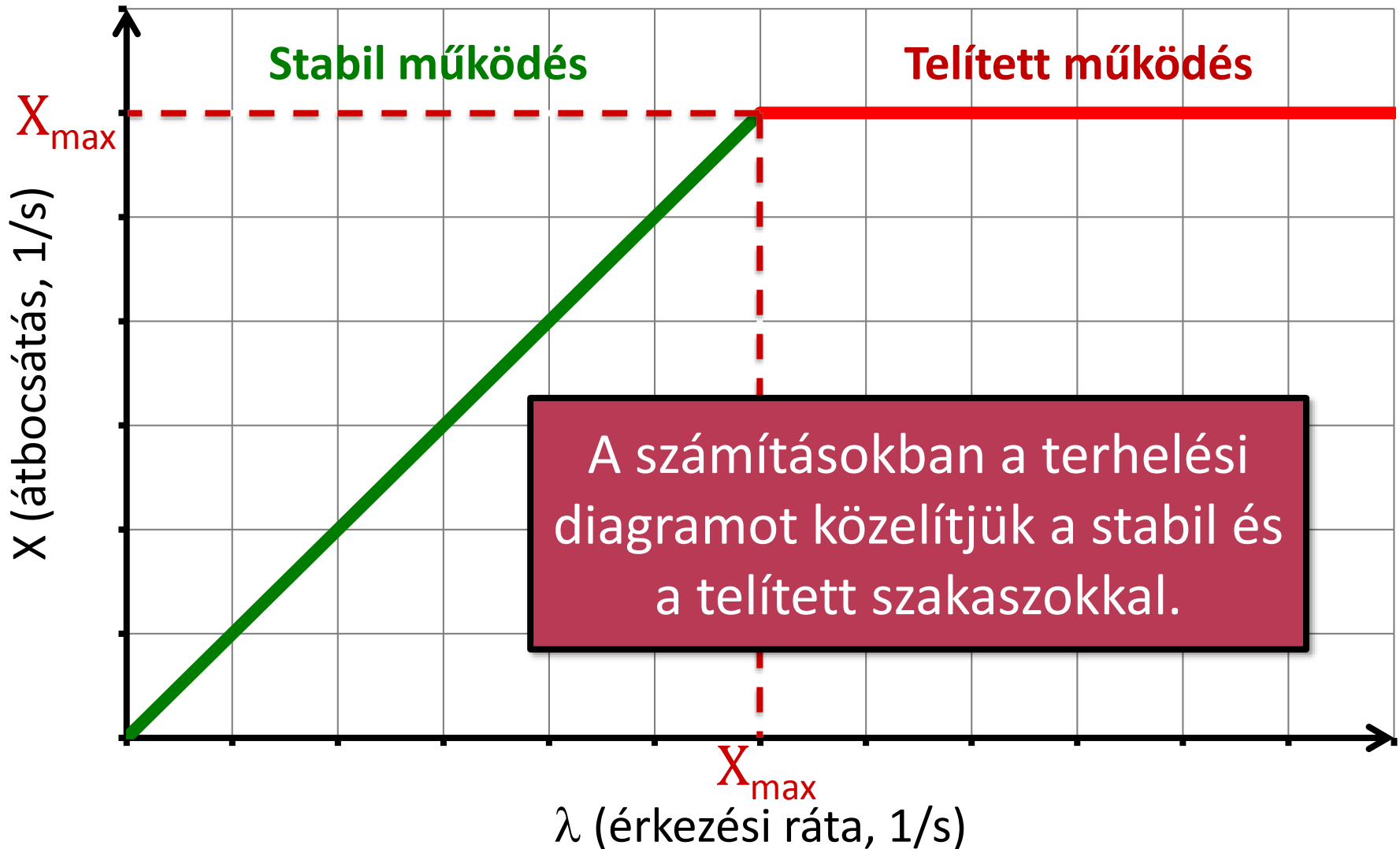
# Átbocsátóképesség



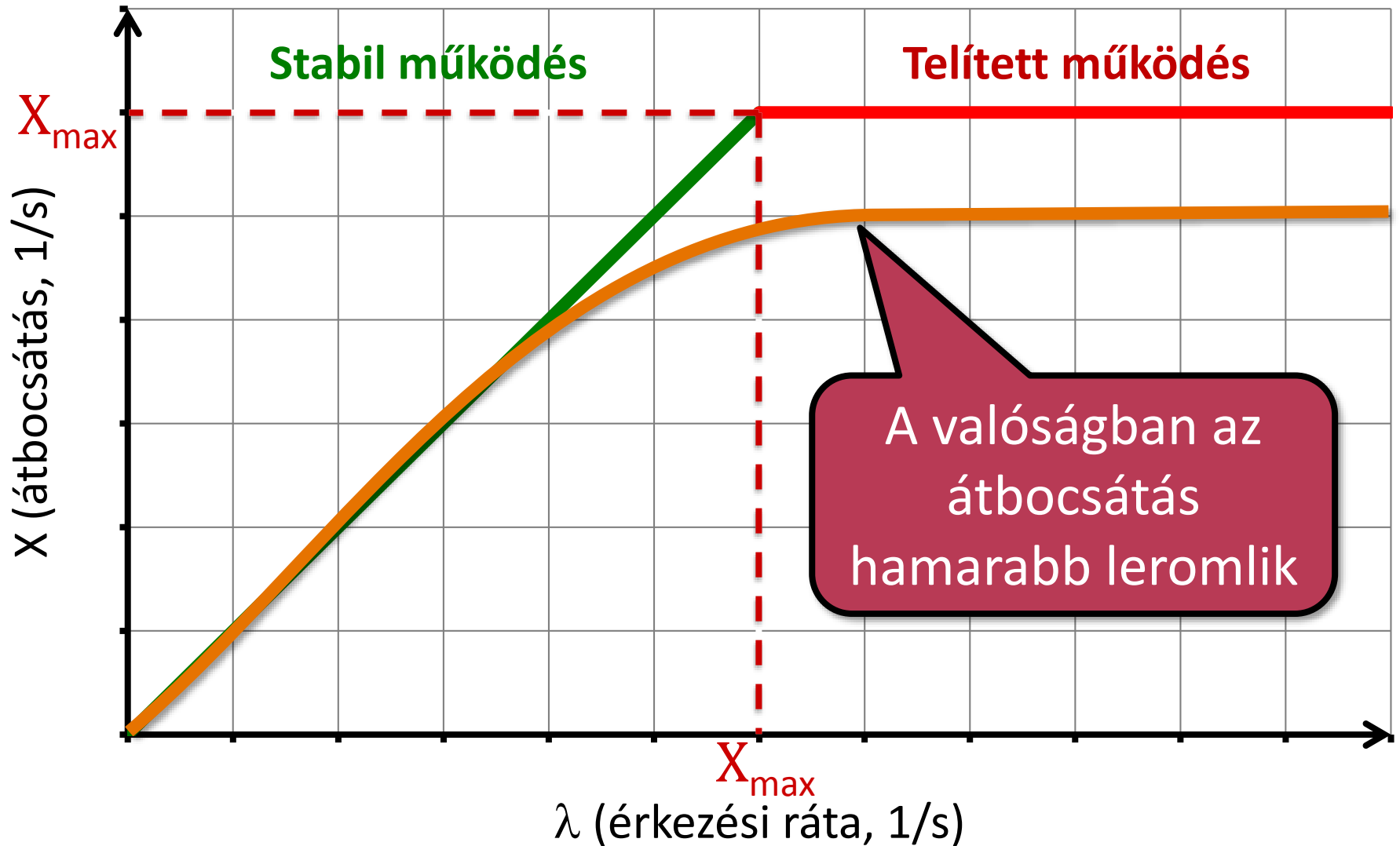
# Kihasználtság



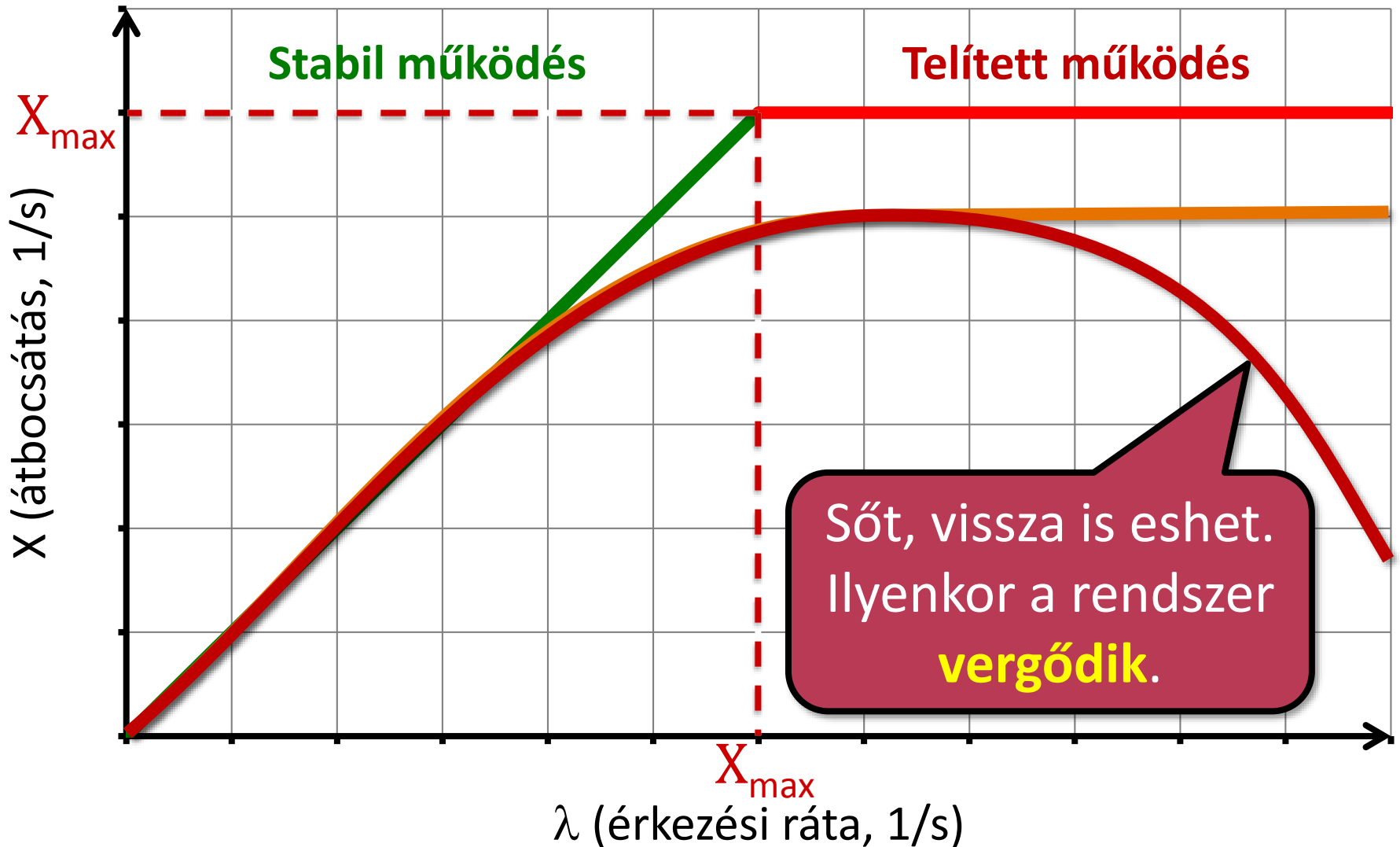
# Közelítő terhelési függvény



# Valós terhelési diagram



# Valós terhelési diagram





# Definíciók

**Átbocsátókéesség:** a maximális átbocsátás.

○ Jele:  $X_{\max}$

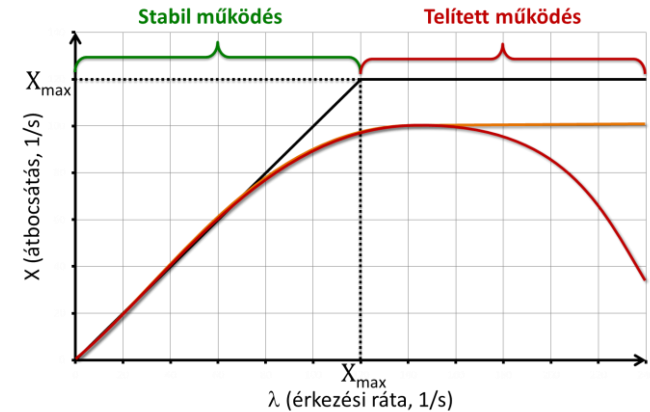
**Kihasználtság (Utilization):** a jelenlegi átbocsátás és a maximális átbocsátás aránya.

○  $U = \frac{X}{X_{\max}}$

**Vergődés:** a rendszer átbocsátása visszaesik a telített működés során.

# A modellben elhanyagolt hatások

- Taszkváltások számításigénye
  - Takarítás az előző után
  - Előkészítés a következőhöz
- Erőforrásváltás számításigénye
- Többszörös telítés
  - Egyszerre több erőforrás (pl. szerver) is telített
  - Pl. ha az M7-esen baleset van, bedugul a 70-es út is



# Elriasztott kérések esete

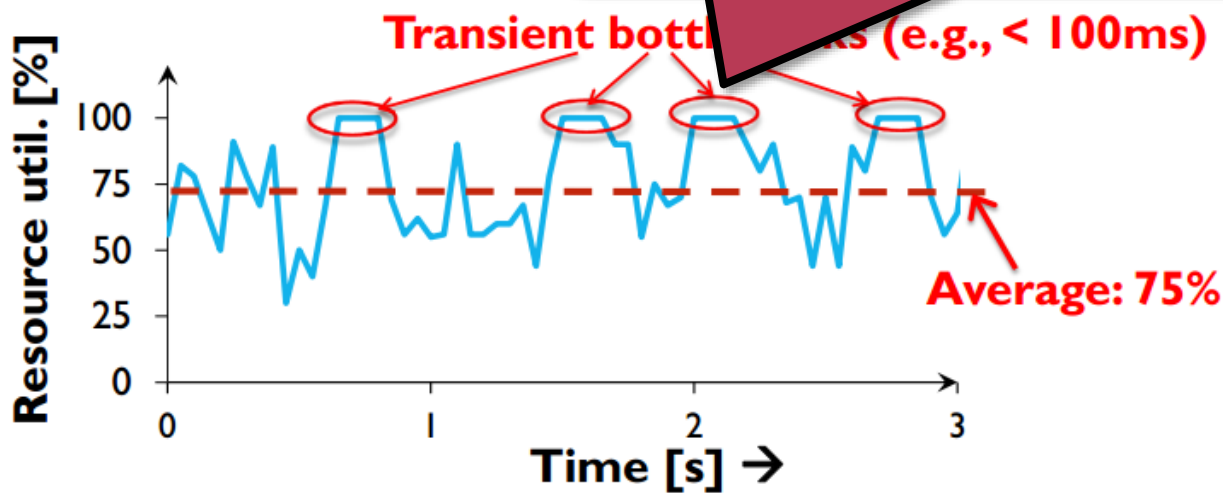
- Az igények csak a fanatikus klienseknél függetlenek a sor hosszától....



# A terhelésingadozás hatása

## ■ Átlagos értékek vs. Valós terhelés

Telített rendszerben a válaszidő **2-3 nagyságrenddel** nagyobb is lehet!



# A terhelésingadozás hatása

- **Átlagos értékek vs. Valós terhelés**

**Nem érdemes 40-60%-nál nagyobb kihasználtságra tervezni!**

# Neptun tárgyfelvétel



Max. egyidejű  
felhasználó

Max. egyidejű felhasználó  
optimális működés mellett

Az elmúlt időkből előfordultak a következők:

Dátum	Jelleg	Max. felhasználó	Op. felhasználó
2010.11.29 18:00	vizsga	7303	4623
2010.12.22 06:00	tárgy (EO)	831	831
2011.01.10 18:00	tárgy	12062	4837
2011.01.12 18:00	tárgy (EP)	1765	1765
2011.01.31 16:00	tárgy	1519	1519
2011.05.02 18:00	vizsga	2761	2761
2011.06.07 18:00	tárgy	6095	6095
2011.11.28 18:00	vizsga	4897	4897
2012.01.16 18:00	tárgy	8120	5328
2012.01.30 16:00	tárgy	1703	1703
2012.05.02 18:00	vizsga	2603	2603

# Neptun tárgyfelvétel



≈ Érkezési ráta ( $\lambda$ )

≈ Átbocsátókéesség ( $X_{\max}$ )

Az elmúlt időkbén előfordultak a következők:

Dátum	Jelleg	Max. felhasználó	Op. felhasználó
2010.11.29 18:00	vizsga	7303	4623
2010.12.22 06:00	tárgy (EO)	831	831
2011.01.10 18:00	tárgy	12062	4837
2011.01.12 18:00	tárgy (EP)	1765	1765
2011.01.31 16:00	tárgy	1519	1519
2011.05.02 18:00	vizsga	2761	2761
2011.06.07 18:00	tárgy	6095	6095
2011.11.28 18:00	vizsga	4897	4897
2012.01.16 18:00	tárgy	8120	5328
2012.01.30 16:00	tárgy	1703	1703
2012.05.02 18:00	vizsga	2603	2603

Mikor volt telített (túlterhelt) a Neptun?

# Neptun tárgyfelvétel



≈ Érkezési ráta ( $\lambda$ )

≈ Átlagos várakozási idő ( $X_{max}$ )

Az elmúlt időkből előfordultak a következők:

Dátum	Jelleg	...	...
2010.11.29 18:00	vizsga		
2010.12.22 06:00	tárgy felvétel		
2011.01.10 18:00			
2011.01.12 18:00			
2011.01.31 18:00			
2011.02.01 18:00			2761
2011.02.02 18:00			6095
2011.11.10 18:00			4897
2012.01.10 18:00			5328
2012.01.30 18:00		1703	1703
2012.05.02 18:00		2603	2603

**A túlterhelt rendszer teljesítménye drasztikusan leeshet, akár össze is omolhat!**

Mikor volt túlterhelt (túlterhelt) a Neptun?



# Neptun tárgyfelvétel



≈ Érkezési ráta ( $\lambda$ )

≈ Átbocsátóképesség ( $X_{\max}$ )

Az elmúlt időkből előfordultak a következők:

Dátum	Jelleg	Max. felhasználó	Op. felhasználó
2010.11.29 18:00	vizsga	7303	4623
2010.12.22 06:00	tárgy (EO)	831	831
2011.01.10 18:00	tárgy	12062	4837
2011.01.12 18:00	tárgy (EP)	1765	1765
2011.01.31 16:00	tárgy	1519	1519
2011.05.02 18:00	vizsga	2761	2761
2011.06.07 18:00	tárgy	6095	6095
2011.11.28 18:00	vizsga	4897	4897
2012.01.16 18:00	tárgy	8120	5328
2012.01.30 16:00	tárgy	17	
2012.05.02 18:00	vizsga	26	

Jól konfigurált  
szerverek, megfelelő  
terhelésméretezés!

# Neptun tárgyfelvétel



Ugyanez az alapja  
a DDoS elleni  
védekezésnek is

áta ( $\lambda$ )

$\approx$  Áth

ség ( $X_{max}$ )

Az

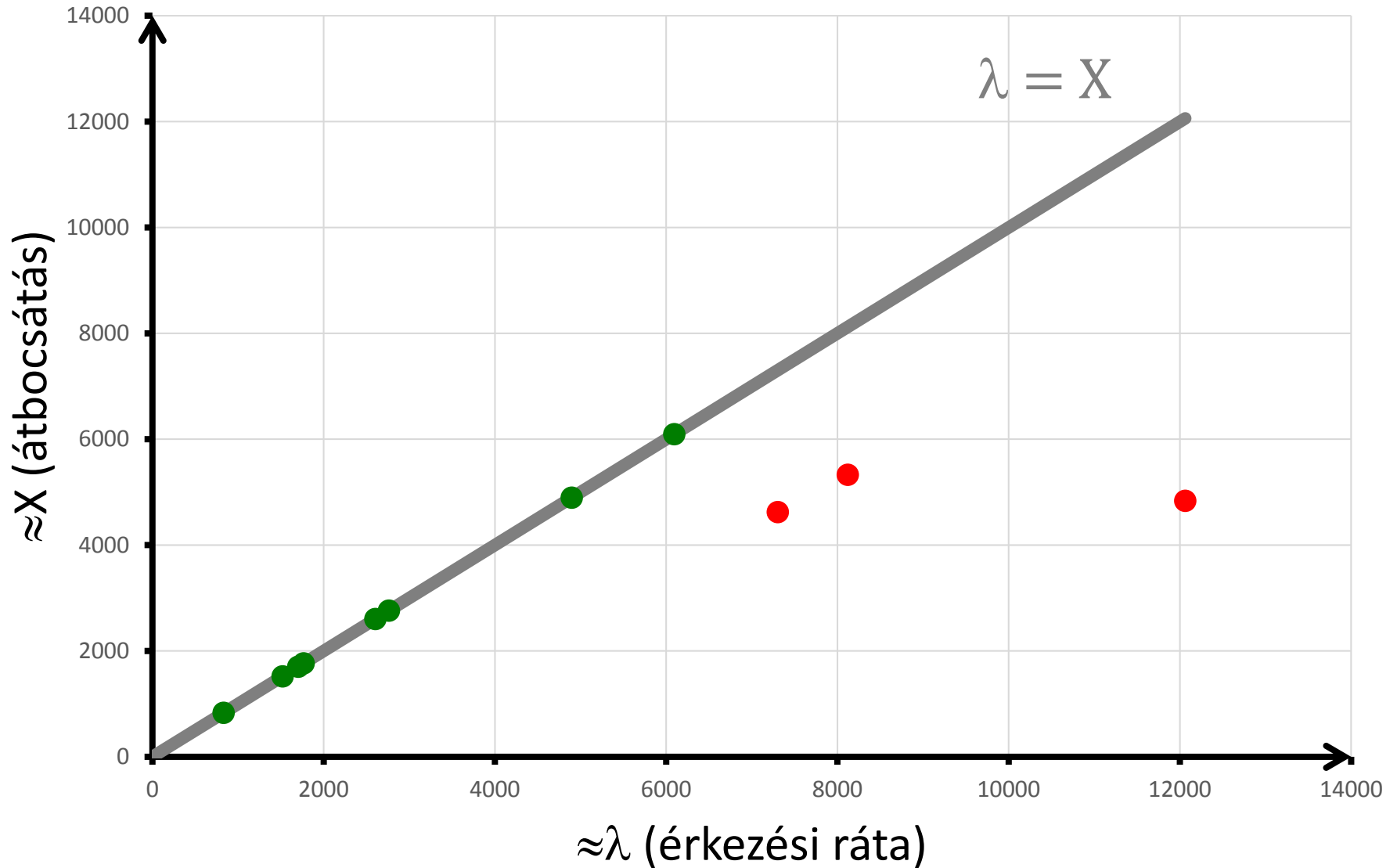
ma

Dátum			Álló
2010.11.29 18:00	vizsga		
2010.12.22 06:00	tárgy /		
2011.01.10 18:00			
2011.01.12 18:00			
2011.01.31 18:00			
2011.05.02 18:00			2761
2011.06.02 18:00			6095
2011.11.28 18:00			4897
2012.01.16 18:00			5328
2012.01.30 16:00		17	
2012.05.02 18:00		26	

**A létszámkorlátos megoldás  
célja egyensúlyi állapotban  
tartani a rendszert**

Jól konfigurált  
szerverek, megfelelő  
terhelésméretezés!

# Előzetes: Adatvizualizáció



Alapfogalmak

Terhelési diagram

Erőforrásmodellezés

# ERŐFORRÁSMODELLEZÉS

Miből származik az átbocsátás korlátja?

# Példa: ZH javítás

- Egy hallgató dolgozatának kijavítása **15 perc**
- **Egy javító óránként** hány dolgozattal végez?

$$X_{max}^{(1)} = \frac{1 \text{ ZH}}{15 \text{ perc}} = 4 \frac{\text{ZH}}{h}$$

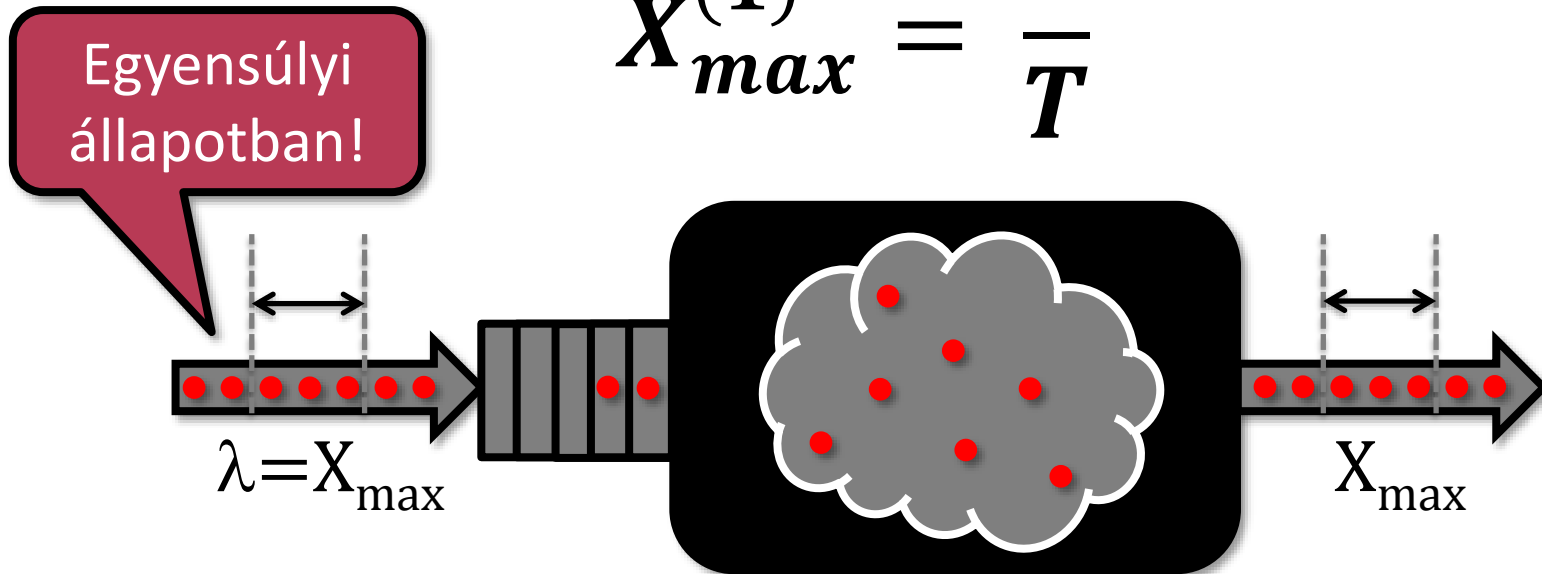
- És **nyolc** javító?

$$X_{max}^{(8)} = 8 \times X_{max}^{(1)} = 8 \times \frac{1 \text{ ZH}}{15 \text{ perc}} = 32 \frac{\text{ZH}}{h}$$

# Kizárólagos erőforrás törvénye

- Ha a kérések közül egyszerre **legfeljebb egy** futhat
  - Pl. egyetlen szerver futtatja őket, azonos változót írnak
  - A többi kérés **sorban áll**
- Ekkor  $T$  az **átlagos végrehajtási idő** mellett:

$$X_{max}^{(1)} = \frac{1}{T}$$



# Kizárólagos erőforrás törvénye

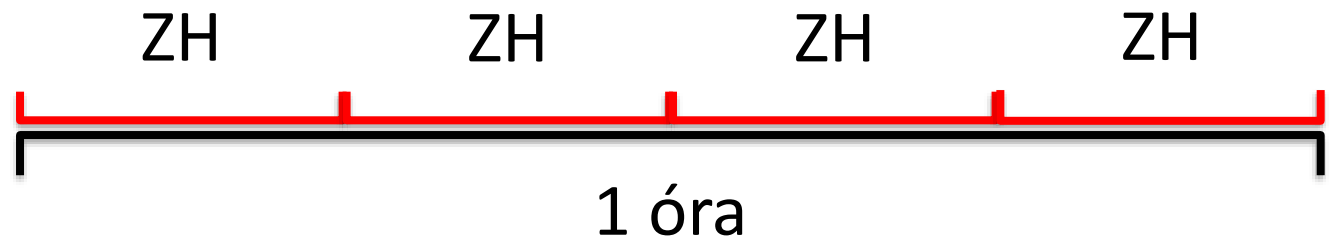
- Ha a kérések közül egyszerre **legfeljebb egy** futhat
  - Pl. egyetlen szerver futtatja őket, azonos változót írnak
  - A többi kérés **sorban áll**
- Ekkor  $T$  az **átlagos végrehajtási** idő mellett:

$$X_{max}^{(1)} = \frac{1}{T}$$

„Hány végrehajtás fér bele egy egységnyi időbe?”

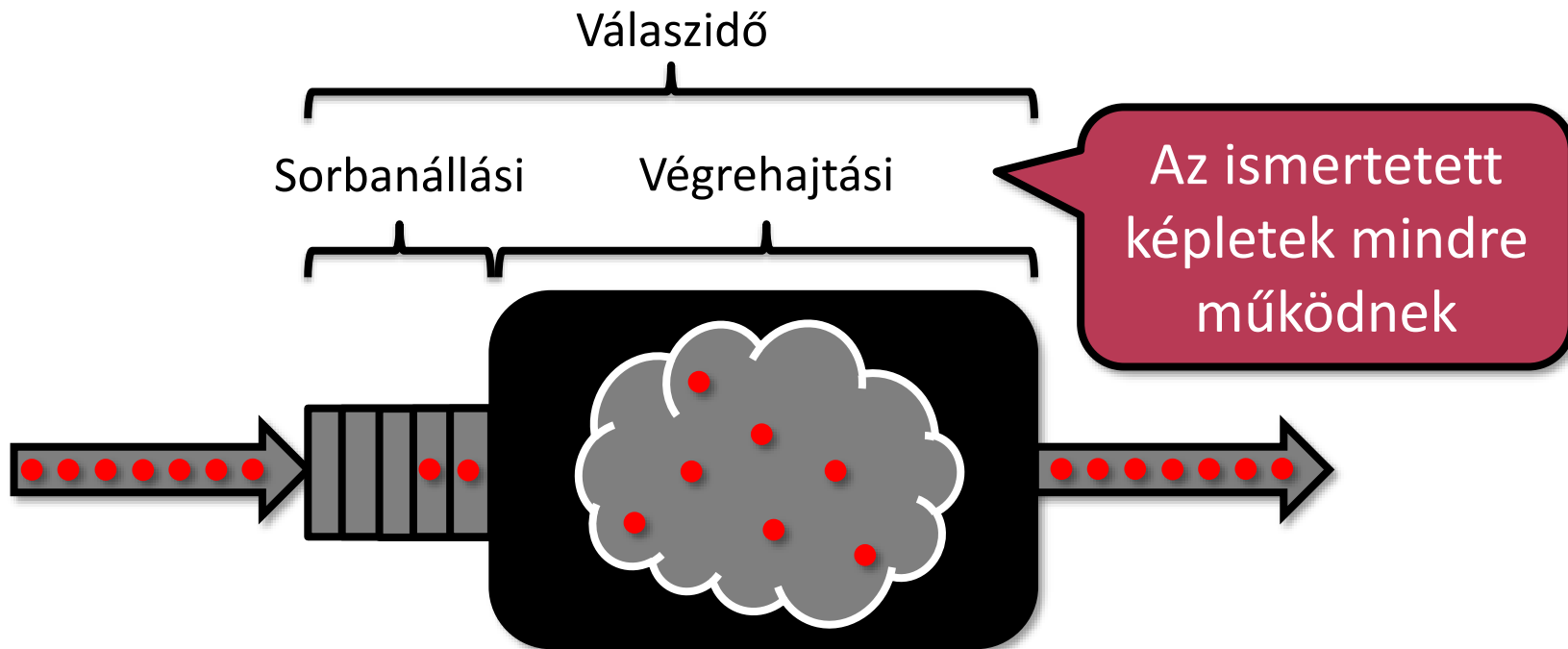
$$T_{ZH} = 15 \text{ perc}$$

$$X_{max} = 4 \frac{1}{h}$$



# Mérhető idők

- **Sorbanállási idő:** várakozás erőforrásra
- **Végrehajtási idő:** kérés feldolgozása
- **Válaszidő:** Sorbanállási + végrehajtási





# Kizárólagos erőforrás kihasználtsága

- Kihasználtság:

„Az átbocsátás és az átbocsátóképeség aránya”

$$X_{max}^{(1)} = \frac{1}{T} \Rightarrow X_{max}^{(1)} \times T = \mathbf{1}$$

# Kizárólagos erőforrás kihasználtsága

- Kihasználtság:

„Az átbocsátás és az átbocsátóképeség aránya”

$$X_{max}^{(1)} = \frac{1}{T} \Rightarrow X_{max}^{(1)} \times T = 1 = U$$

- A kihasználtság képlete tehát:

$$U = X \times T$$

- Intuitíven:

„Az egységnyi idő alatt beérkező  $X$  taszk végrehajtása  $T$  végrehajtási idővel az egységnyi idő hányadrészét teszi ki?”

# Kizárólagos erőforrás kihasználtsága

- Kihasználtság:

„Az átbocsátás és az átbocsátókéesség aránya”

$$X_{max}^{(1)} = \frac{1}{T} \Rightarrow X_{max}^{(1)} \times T = 1 = U$$

- A kihasználtság képlete tehát:

$$U = X \times T$$

- Intuitíven:

$$X = 3 \frac{1}{h}$$

$$T_{ZH} = 15 \text{ perc}$$

$$U = 75\%$$



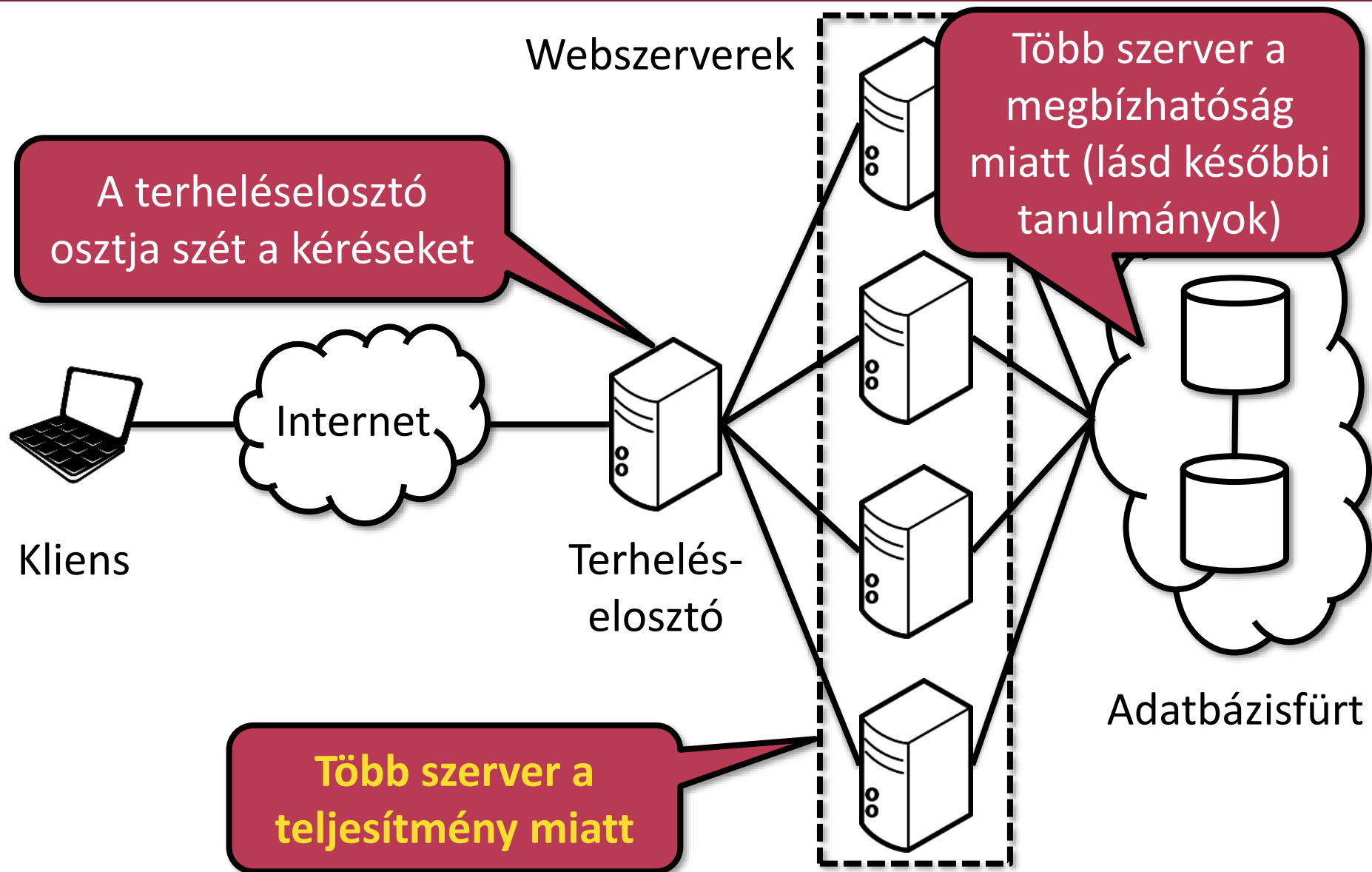
# Kitekintés: skálázás

- **Vertikális skálázás** (Scale-up):
  - A feldolgozóegység **teljesítményét** növeljük
  - Pl. erősebb CPU, több RAM
  - Egyszerű és nagyszerű 😊
  - Technológiai korlátok ☹️
- **Horizontális skálázás** (Scale-out):
  - A feldolgozóegységek **számát** növeljük
  - Pl. több CPU mag, több szerver
  - Elvileg korlátlanul növelhető 😊
  - Plusz bonyolultság ☹️

# Slace-out a hétköznapiakban



# Scale-out a Neptunban



# K erőforráspéldány használata

- Ha az összes kérés közül egyszerre **legfeljebb K** futhat
  - Pl. K darab fürtözött szerver futtathatja őket
  - A többi folyamatpéldány **sorban áll**
- Ekkor:
  - Legyen T az **átlagos végrehajtási idő**

$$X_{max}^{(K)} = K \times X_{max}^{(1)} = \frac{K}{T}$$

# K erőforráspéldány használata

- Ha az összes kérés közül egyszerre **legfeljebb K** futhat
  - Pl. K darab fürtözött szerver futtathatja őket
  - A többi folyamat
- Ekkor:
  - Legyen T az átlag

Több erőforráspéldánnyal és párhuzamosítással a rendszer **skálázható**.

$$X_{max}^{(K)} = K \times X_{max}^{(1)} = \frac{K}{T}$$



# Vízvezeték analógia

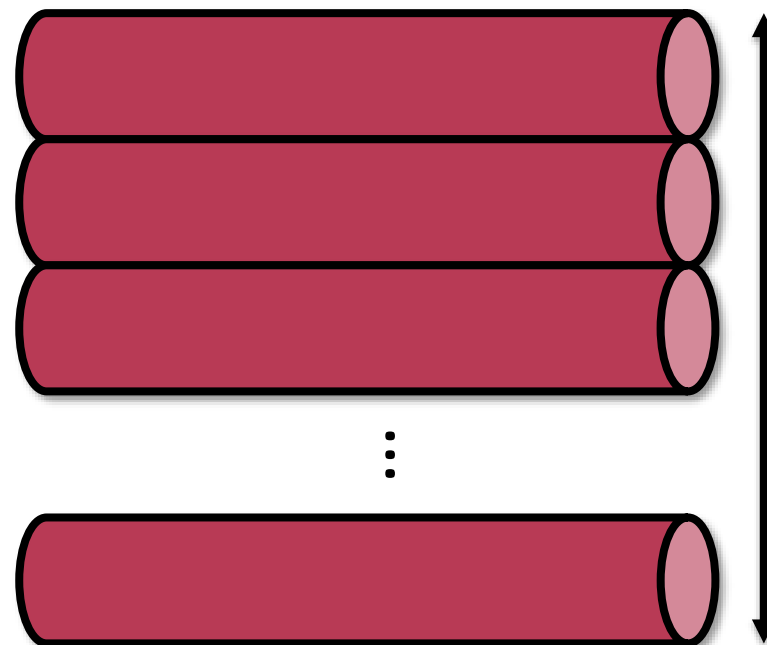
Egyetlen kizárólagos erőforráspéldány



Átbocsátóképesség:

$$X_{max}^{(1)}$$

K szabadon választható erőforráspéldány



Átbocsátóképesség:

$$X_{max}^{(K)} = K \times X_{max}^{(1)}$$

# K erőforráspéldány kihasználtsága

- Az előző gondolatmenet analógiájára:

$$X_{max}^{(K)} = \frac{K}{T} \quad \Rightarrow \quad X_{max}^{(K)} \times T = \textcircled{K} ?$$

# K erőforráspéldány kihasználtsága

- Az előző gondolatmenet analógiájára:

$$X_{max}^{(K)} = \frac{K}{T} \Rightarrow X_{max}^{(K)} \times T = \mathbf{K = K \times U}$$

- A kihasználtság képlete ebben az esetben:

$$U = \left( \frac{X}{K} \right) \times T$$

- Intuitíven:

„K időegység hányadrészében dolgozna 1 példány?”

„Az **egy példányra eső átlagos átbocsátás** mellett mekkora egy erőforráspéldány kihasználtsága?”

# K erőforráspéldány kihasználtsága

- Az előző gondolatmenet analógiájára:

$$X_{max}^{(K)} = \frac{K}{T} \Rightarrow X_{max}^{(K)} \times T = K = K \times U$$

- A kihasználtság képlete ebben az esetben:

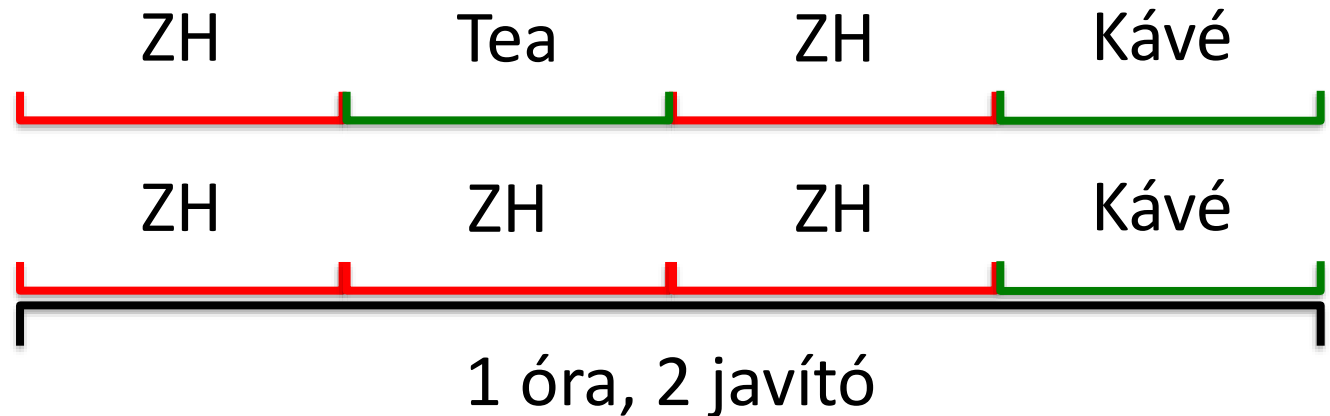
$$U = \frac{X}{K} \times T$$

- Intuitíven:

$$X = 5 \frac{1}{h}$$

$$T_{ZH} = 15 \text{ perc}$$

$$U = 62,5\%$$



# Összefoglalás

## ■ Egyensúlyi állapot:

- Átlagos értékekkel számolunk
- $\lambda = X$  (érkezési ráta = átbocsátás)

## ■ Átbocsátóképesség:

- Az elérhető maximális átbocsátás
- $X_{\max} = \frac{K}{T}$  (K erőforráspéldány esetén)

## ■ Kihasználtság:

- Az átbocsátás és az átbocsátóképesség aránya
- $U = \frac{X}{K} \times T$  (K erőforráspéldány esetén)