

Rendszermodellezés gyakorló feladat – Teljesítmény-modellezés és adatelemzés (13 pont)

Adott egy keresési szolgáltatás, melyet 3 különböző komponens valósít meg. Egy komponens áll egy szoftverből (melynek működését itt nem modellezzük) és az azt kiszolgáló hardver erőforrásokból. Feltételezzük, hogy a terhelés szolgáltatásigénye *lineárisan skálázódik*.

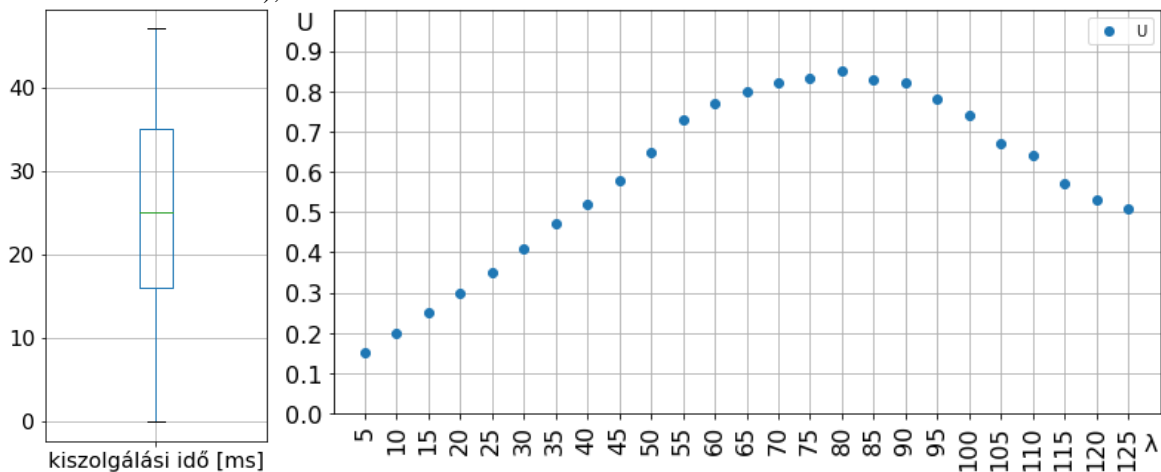
F1	/10
F2	/13
F3	/13
Szumma	/36

Minden felhasználói kérést egyetlen konkrét komponens szolgál ki, amelyet egy terhelélosztó választ ki. a feladatunk megtalálni az optimális terhelélosztást, mely esetében a rendszer áteresztőképessége maximális az adott erőforrások mellett.

Minden komponens két hardveres erőforrásával jellemezzük: CPU és Diszk. A különböző szoftverek különbözőképpen használják ezeket. Nem ismerjük a belső működést, ezért 15 kérés/s terheléssel megvizsgáljuk a rendszer működését, a komponensek közt egyenletes „*round robin*” terhelélosztást használva. Az oldalt látható átlagos erőforrás-használati időket mérjük *egyetlen kérés* kiszolgálására vetítve:

Komponens	CPU idő	Diszk idő
1	40,0 ms	5,0 ms
2	25,0 ms	20,0 ms
3	20,0 ms	12,5 ms

- Ha egyenletesen osztjuk el a kéréseket (round robin), mekkora lesz a rendszer által időegységenként kiszolgált kérések maximális száma? (2 p)
- Más terhelélosztást választva maximálisan mekkora átbocsátóképesség érhető el? Milyen terhelélosztási súlyok mellett működne optimálisan a rendszer? (2 p)
- Miért lényeges információ, hogy 15 kérés/s mellett mértük meg a kiszolgálási időket? (1 p)
- Valójában a kérések elosztását végző terhelélosztó nem csak az elosztást, hanem a naplózást is elvégzi, átlagosan 20 ms időt töltve minden kéréssel. Legalább mennyi kéréssel kell tudnia egyszerre/átlapolva foglalkoznia annak érdekében, hogy (az optimális terhelélosztást feltételezve) ne váljon maga a terhelélosztó szűk keresztmetszetté a b) pontban leírt esetében? (1 p)
- Óvatosak vagyunk, így nem az átlagos kiszolgálási idő, hanem a (a löketes terhelés miatt) a kiszolgálási idők 75 percentilise alapján akarjuk méretezni a terhelélosztó kapacitását (átlapoltan feldolgozott kérések száma), mekkorára méretezzünk?



A méréseket a fenti boxplot foglalja össze. (Dimenzió: kérés/s) (1 p)

- A 2018. májusában életbe lépő GDPR szabályozás miatt be kell iktatnunk a rendszerbe egy órakulmot, mely minden kérés kiadása előtt megvizsgálja, hogy a kérés nem sérti-e a felhasználói adatok rendelkezésére vonatkozó előírásokat. Előzetesen kimértük a komponens átlagos kihasználtságát többször, egyenletes terhelések mellett (ld. ábra jobbra fent). Várhatóan fogja-e korlátozni a a) és a b) pontban meghatározott rendszerek működését ez az órakulum? (2p)
- A rendszer valójában egy kritikus alkalmazást szolgál ki, így át kell tervezni a működést. A továbbiakban (terhelélosztó használata nélkül) mindig először az 1. komponens kapja a kéréseket. Amennyiben az 1. komponens által végzett lekérdezés nem jár eredménnyel (az esetek 20%-ban) a 2. és a 3. komponens párhuzamosan végrehajtja ugyanazt a lekérdezést. Amíg nem adnak hibahatáron belül egyező eredményt, ezen két komponens lekérdezését újra és újra meg kell ismételni. Adjon jólstruktúrált folyamatmodellt a rendszer működésére! (2 p)

- h) Mekkora eséllyel adhat vissza hibahatáron túl eltérő eredményt a 2. és a 3. komponens, ha a lekérdezés mint folyamat átlagos válaszideje nem haladhatja meg a 75 ms-t? Közelítésként feltételezzük, hogy az egyes komponensek válaszideje megegyezik a maximális erőforráshasználati idejükkel, és hogy a hibahatáron túli válasz valószínűsége nem változik a folyamat során. (2 p)

