

12. Ütközés feloldási protokollok

Kommunikációs hálózatok teljesítményének elemzése

Horváth Illés

2024/11/27

1. Újraküldéses M/M/1 sor
2. Réselt ALOHA protokoll
3. Réselt CSMA protokoll
4. Réselt CSMA/CD protokoll

Újraküldéses M/M/1 sor

Az alap M/M/1 sornál egyetlen szerver van, és ha olyankor érkezik egy igény, amikor a szerver épp dolgozik, akkor azt a szerver eltárolja egy bufferben, majd amikor a szerver végzett a kiszolgálással, egyből elkezd a következő igényt kiszolgálni.

Ha a szervernek nincs buffere, akkor amíg a szerver épp dolgozik, a beérkező igények elvesznek.

Most egy olyan modellt vizsgálunk meg, amikor a szervernek nincs buffere, de mégis szeretnénk, hogy minden beérkező igény átmenjen.

Ezt újraküldéssel (retrial) oldjuk meg: ha olyankor érkezik egy igény, amikor a szerver épp dolgozik, akkor a szerver visszajelez a forrásnak, hogy az igényt nem tudja fogadni, a forrás pedig valamennyi (általában véletlen) idő múlva újból megpróbálkozik a küldéssel.

Újraküldéses M/M/1 sor

Ilyenkor a korábban visszautasított igények tulajdonképpen hasonlóan várakoznak, mint az alap M/M/1 sor esetén a bufferben, azonban amikor a szerver felszabadul, akkor nem kezdi el egyből kiszolgálni a következő igényt, hanem megvárja, amíg valamelyik forrás újraküldi egy igényt.

Az *újraküldéses M/M/1* sor a következő:

- ▶ a szerver egy igényt $\text{EXP}(\mu)$ idő alatt szolgál ki (azaz a szerver kiszolgálási rátája μ);
- ▶ az igények λ rátájú Poisson-folyamat szerint érkeznek;
- ▶ ütközés esetén egy igényt a forrás $\text{EXP}(\lambda_r)$ idő múlva próbál újraküldeni (azaz az újraküldés rátája λ_r).

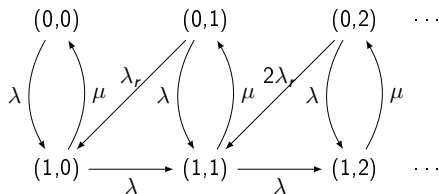
(Lényegében az érkezés, az újraküldés és a kiszolgálás is memóriamentes.)

Újraküldéses M/M/1 sor

Tekintsük a (C, N) párt, ahol N az újraküldésre váró igények száma, és C értéke 0 vagy 1 aszerint, hogy a szerver éppen szabad vagy dolgozik.

A korábbi feltételek mellett (C, N) -re teljesül a Markov-tulajdonság, azaz $(C(t), N(t))$ egy folytonos idejű Markov-lánc.

Az állapotgráfja a következő:



Újraküldéses M/M/1 sor

Tétel

Egy újraküldéses M/M/1 sor pontosan akkor stabil, ha

$$\rho = \frac{\lambda}{\mu} < 1.$$

Biz. (vázlat): Összehasonlítjuk az alap M/M/1 sorral.

Egy alap M/M/1 sor $0, 1, 2, 3, \dots$ állapotai megfelelnek az itteni $(0, 0), (1, 0), (1, 1), (1, 2), \dots$ állapotoknak.

Annyi a különbség, hogy az újraküldéses M/M/1 sor az $(1, n)$ állapotból két lépésben jut el az $(1, n - 1)$ állapotba:

$(1, n) \rightarrow (0, n) \rightarrow (1, n - 1)$, és ez plusz $\text{EXP}(n\lambda_r)$ időt jelent az alap M/M/1 sorhoz képest.

$n \rightarrow \infty$ esetén $\text{EXP}(n\lambda_r) \rightarrow 0$, azaz a plusz várakozás nagy n esetén elhanyagolható, és a stabilitási feltétel ugyanaz, mint az alap M/M/1 sorra, vagyis $\rho = \lambda/\mu < 1$.

Újraküldéses M/M/1 sor

Tétel

Egy stabil újraküldéses M/M/1 sor stacionárius eloszlása

$$p_{(0,n)} = \frac{\rho^n}{n! \lambda_r^n} \prod_{i=0}^{n-1} (\lambda + i \lambda_r) (1 - \rho)^{\frac{\lambda}{\lambda_r} + 1}$$
$$p_{(1,n)} = \frac{\rho^{n+1}}{n! \lambda_r^n} \prod_{i=1}^n (\lambda + i \lambda_r) (1 - \rho)^{\frac{\lambda}{\lambda_r} + 1}$$

Nem biz., de csak be kell helyettesíteni a stacionárius eloszlás definíciójába.

Újraküldéses M/M/1 sor

További teljesítményjellemzők.

A rendszerben lévő igények átlagos száma (beleértve azt, amit éppen kiszolgál a szerver, és az újraküldésre váró igényeket is):

$$L = \sum_{n=0}^{\infty} np_{(0,n)} + (n+1)p_{(1,n)} = \frac{\rho(\lambda + \lambda_r)}{(1 - \rho)\lambda_r}.$$

Egy igény átlagos rendszerben töltött ideje (Little-formula alapján):

$$W = \frac{\lambda + \lambda_r}{(1 - \rho)\mu\lambda_r}.$$

Közeghozzáférés

Az újraküldéses M/M/1 sorban ha λ_r értéke nagy, akkor az újraküldések során gyakran lesznek ütközések. Ezt az újraküldéses M/M/1 sor nem bünteti, vagyis a szervert nem zavarja, ha kiszolgálás közben más forrás is próbál adni. Ez a gyakorlatban sok esetben nem realiztikus.

Megvizsgáljuk a másik végletet is: amikor ütközés esetén a kiszolgálás azonnal megszakad és az addigi kiszolgálás kárbavész. Ez realiztikus helyzet például rádiós kommunikáció esetén, amikor sok terminál próbál kommunikálni egy bázisállomással.

Ilyen helyzetekben különösen nagy jelentősége van a közeghozzáférési protokollnak (medium access control, MAC), vagyis annak, hogyan kezeljük az ütközéseket, és hogyan szabályozzuk, hogy melyik forrás hogyan fér hozzá a szerverhez/csatornához.

ALOHA protokoll

Az ALOHA protokoll két csatornát használ: a fő csatornán a források adnak a szervernek, a másik csatornán a szerver jelzi a sikeres átvitelt a forrásoknak.

Ha egy forrás adott határidőig nem kap visszajelzést, akkor feltételezi, hogy az átvitel sikertelen volt, és átvált újraküldés üzemmódba. Ilyenkor ugyanazt az üzenetet ismétli mindaddig, amíg sikeres nem lesz az átvitel.

Ha minden forrás azonnal újraküldene a határidő lejárta után, akkor pontosan ugyanazok a források ismét ütköznének. Ennek elkerülése érdekében a források újraküldés előtt véletlen ideig várakoznak.

Ennek következménye egy tradeoff: ha túl rövid a várakozás az újraküldés előtt, gyakori lesz az ütközés; ha túl nagy a várakozás, ritkábbak lesznek az ütközések, viszont az átvitel ideje hosszabb lesz.

ALOHA protokoll

Az ALOHA protokoll problémái.

Mivel a források az ütközésről nem az ütközés pillanatában értesülnek, csak a határidő lejártakor, ezért ütközés után még hosszú ideig adnak feleslegesen.

Ez különösen a hosszabb üzenetek esetében problémás, amik egyrészt eleve sokáig blokkolják a csatornát, másrészt az ütközés esélye is nagy.

Ezért a javított változatát, a réselt ALOHA protokollt fogjuk vizsgálni és elemezni.

Réselt ALOHA protokoll

A réselt (slotted) ALOHA protokollnál az időt rövid, T hosszúságú diszkrét ablakokra (résekre) osztjuk. A források minden ablak elején adhatnak.

A források felosztják az átküldendő forgalmat is olyan egységnyi csomagokra, amelyek egy ablakban átvihetők a csatornán.

A sikeres átvitelről ideális esetben az ablak végén egyből kap értesítést a forrás (és így a sikertelen átvitel is kiderül ugyanakkor), de általában lehet, hogy csak D_{ack} ablakkal később.

Ütközés esetén a késleltetés ideje véletlenszerű, egész többszöröse az ablakméretnek.

Réselt ALOHA protokoll

Tegyük fel, hogy a csomagok λ paraméterű Poisson-folyamat szerint érkeznek a forrásokból. Minden csomag megvárja a következő időablak elejét.

Ha az időablak mérete T , akkor az egy időablakban érkező új csomagok száma $X \sim \text{POI}(\lambda T)$.

Egy ablakban akkor van sikeres átvitel, ha pontosan egy csomag érkezik alatta. Ha nem lennének újraküldött csomagok, hanem csak az új csomagok érkeznének, akkor egy időablakban a sikeres átvitel valószínűsége

$$\rho = P(X = 1) = \lambda T e^{-\lambda T}$$

lenne. (Ez az érték egyúttal a csatorna kihasználtsága is.)

De vannak újraküldött csomagok is, ezeket is figyelembe kell vennünk.

Réselt ALOHA protokoll

Az újraküldött csomagok figyelembevételéhez azt tesszük fel, hogy az új és az újraküldött csomagok együttesen is Poisson-folyamatot alkotnak (ez az ún. zero order feltevés).

Ez a feltevés nem teljesen jogos, de jó közelítésnek tekinthető. Az új érkezések valóban Poisson-folyamatot alkotnak, az újraküldött csomagok közül viszont azok, amik azonos ablakban ütköztek, nem teljesen függetlenek, de ha a késleltetés aránylag nagy szórású, akkor közel függetlenek.

Réselt ALOHA protokoll

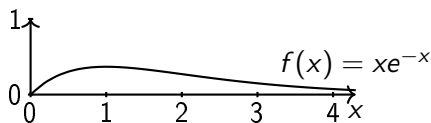
Legyen az új és újraküldött csomagok együttes (aggregált) rátája λ_a .

A következő két kérdést vizsgáljuk:

- ▶ Hogyan állítsuk be λ ismeretében az időablakok T hosszát?
- ▶ Hogyan lehet λ és T ismeretében kiszámítani λ_a értékét?

A két kérdést könnyebb egyszerre megválaszolni.

Az első észrevétel, hogy az xe^{-x} függvénynek a maximuma az $x = 1$ pontban van és a maximum értéke $1/e \approx 0.3679$.



Réselt ALOHA protokoll

Ha az aggregált érkezési ráta λ_a , akkor egy küldésre a sikeres átvitel valószínűsége

$$\lambda_a T e^{-\lambda_a T}.$$

Mivel az xe^{-x} függvény maximuma $x = 1$ -ben van, ezért ez akkor lesz maximális, ha $\lambda_a T = 1$, azaz

$$T = 1/\lambda_a.$$

Ekkor minden küldésnél $1/e$ valószínűséggel lesz sikeres az átvitel, így egy új érkezés összesen $\text{GEO}(1/e)$ küldés után fog átmenni. Ennek várható értéke $\frac{1}{1-1/e}$, azaz

$$\lambda_a = \frac{\lambda}{1 - 1/e},$$
$$T = \frac{1 - 1/e}{\lambda}.$$

Réselt ALOHA protokoll

Főbb teljesítményjellemzők.

Optimálisan beállított réselt ALOHA protokollra a csatorna kihasználtsága

$$\rho = 1/e \approx 0.3679.$$

Egy csomag rendszerben töltött ideje az az idő, amennyi az érkezésétől a sikeres átvitel visszajelzéséig kell. Ennek várható értéke

$$D = \frac{1}{1 - 1/e} T + \left(\frac{1}{1 - 1/e} - 1 \right) (D_{\text{ack}} + E(W)),$$

ahol $E(W)$ az ütközés utáni késleltetés várható értéke.

Réselt ALOHA protokoll

Mivel az xe^{-x} függvény elég lapos a maximumhelye közelében, ezért nem nagy baj, ha az időablak hosszát nem teljesen pontosan állítjuk be a forgalmak érkezési rátájához.

Szuboptimálisan beállított ALOHA protokoll esetén a λ érkezési ráta és T időablak hossz függvényében λ_a értékét a

$$\lambda_a = \frac{\lambda}{1 - \lambda_a T e^{-\lambda_a T}}$$

implicit egyenlet (numerikus) megoldása adja meg.

A csatorna kihasználtsága

$$\rho = \lambda_a T e^{-\lambda_a T}.$$

Egy csomag átlagos rendszerben töltött ideje

$$D = \frac{1}{1 - \lambda_a T e^{-\lambda_a T}} T + \left(\frac{1}{1 - \lambda_a T e^{-\lambda_a T}} - 1 \right) (D_{\text{ack}} + E(W)).$$

Réselt ALOHA protokoll

A zero order feltevés elfedi annak a jelentőségét, hogy az ütközés után mennyi a késleltetés, pedig a gyakorlatban ez is fontos.

Gyakran szoktak exponenciális back-off-ot alkalmazni, ami azt jelenti, hogy első ütközésnél a késleltetés egyenletes egy $[0, D_0]$ intervallumból, a második ütközésnél egyenletes a $[0, 2D_0]$, majd a $[0, 4D_0]$ stb. intervallumból. Egy széles intervallumból vett egyenletes eloszlás azért praktikus, mert ebben az esetben a zero order feltevés jó közelítéssel teljesül.

Az egyre növvő intervallumok miatt a Markov-tulajdonság csak akkor teljesül a rendszerre, ha az újraküldött igényeknél azt is tudjuk, hány ütközésen vannak túl. De ezt általában már nem szoktuk figyelembe venni.

CSMA és CSMA/CD protokollok

Vezetékes rendszereknél előfordul, hogy van lehetőség arra, hogy az ütközést maguk a források is észleljék (esetleg késleltetéssel), és reagáljanak rá. A lehetséges reakciók:

- ▶ ha egy forrás tudja még küldés előtt észlelni, hogy másik forrás is ad, akkor ő nem kezd el adni, ez a CSMA (carrier sense multiple access);
- ▶ ha az ütközést csak késleltetve veszik észre, akkor még mindig megtehetik, hogy az ütköző források leállítják az adást, ez a CD (collision detection).

Réselt CSMA protokoll

A réselt (slotted) CSMA protokoll esetén bármelyik aktív forrás τ idő alatt észlelni tudja a további forgalmakat a csatornában.

Tipikusan $\tau < T$, és feltesszük, hogy T egész többszöröse τ -nak (pl. $T = 10\tau$). Tulajdonképpen ilyenkor τ lesz a legkisebb időegység, amit vizsgálunk.

A protokoll működése a következő:

- ▶ Ha egy forrás ad, akkor ütközés csak a következő τ hosszú időablak végéig lehetséges.
- ▶ Ütközés esetén T hosszú időszak elvész, utána a csatorna szabad, az ütköző források pedig majd véletlen késleltetés után próbálnak újraküldeni.
- ▶ Ha viszont nincs ütközés a következő τ hosszú időablak végéig, akkor összesen T ideig a többi forrás nem ad, hanem megvárják, amíg az aktív forrás befejezi az adást.

Réselt CSMA/CD protokoll

A réselt CSMA/CD protokoll esetén bármelyik aktív forrás τ idő alatt észlelni tudja a további forgalmakat a csatornában. Tipikusan $\tau < T$, és feltesszük, hogy T egész többszöröse τ -nak (pl. $T = 10\tau$). Tulajdonképpen ilyenkor τ lesz a legkisebb időegység, amit vizsgálunk.

A protokoll működése a következő:

- ▶ Ha egy forrás ad, akkor ütközés csak a következő τ hosszú időablak végéig lehetséges.
- ▶ Ütközés esetén τ hosszú időszak elvész, utána a csatorna szabad, az ütköző források pedig majd véletlen késleltetés után próbálnak újraküldeni.
- ▶ Ha viszont nincs ütközés a következő τ hosszú időablak végéig, akkor összesen T ideig a többi forrás nem ad, hanem megvárják, amíg az aktív forrás befejezi az adást.

Réselt CSMA protokoll elemzése

A réselt CSMA protokoll elemzéséhez ismét a zero order feltevással dolgozunk, tehát feltesszük, hogy az új és újraküldött érkezések együtt λ_a rátájú Poisson-folyamat szerint érkeznek.

A csatornán háromféle időszakot különböztetünk meg:

- ▶ tétlen (idle, I): ilyenkor a csatorna szabad,
- ▶ ütközés (collision, L),
- ▶ sikeres küldés (success, S): ilyenkor a csatornán átmegy egyetlen csomag, és amíg át nem ér, a többi forrás nem ad.

Az egyes τ hosszú időablakok is besorolhatóak a 3 állapot valamelyikébe. Ugyan az időablakok sorozatára nem teljesül a Markov-tulajdonság, de az egyes szakaszok végén lévő pillanatokban a rendszer memóriamentes, tehát ha az egyes szakaszok végén lévő pillanatokba beágyazott folyamatot nézzük, az már egy diszkrét idejű Markov-lánc. Kiszámítjuk az átmenet-valószínűség mátrixot.

Réselt CSMA protokoll elemzése

Egy tétlen időszak akkor ér véget, amikor először ad legalább egy forrás egy τ hosszú ablakban. Jelölje X ezen források számát. Ezen belül ha $X = 1$, akkor a küldés sikeres, ha $X \geq 2$, akkor ütközés van. Ezek alapján

$$p_{I \rightarrow S} = P(X = 1 | X \geq 1) = \frac{\lambda_a \tau e^{-\lambda_a \tau}}{1 - e^{-\lambda_a \tau}},$$

$$p_{I \rightarrow L} = P(X \geq 2 | X \geq 1) = \frac{1 - (1 + \lambda_a \tau)e^{-\lambda_a \tau}}{1 - e^{-\lambda_a \tau}}.$$

(A feltételes valószínűség annak felel meg, hogy a tétlen időszak végén nézünk rá a rendszerre.)

Általában tétlen időszakot vagy ütközés, vagy sikeres küldés követ, utána mindig ismét tétlen időszak következik. Más átmenet nem lehetséges. Ez alapján

$$p_{S \rightarrow I} = p_{L \rightarrow I} = 1,$$

$$p_{S \rightarrow S} = p_{L \rightarrow L} = p_{L \rightarrow L} = p_{L \rightarrow S} = p_{S \rightarrow L} = 0.$$

Réselt CSMA protokoll elemzése

A beágyazott Markov-lánc átmenetmátrixa az előbbi értékekből összerakva

$$P = \begin{bmatrix} 0 & \frac{1-(1+\lambda_a\tau)e^{-\lambda_a\tau}}{1-e^{-\lambda_a\tau}} & \frac{\lambda_a\tau e^{-\lambda_a\tau}}{1-e^{-\lambda_a\tau}} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

az ehhez tartozó (beágyazott) stacionárius eloszlás

$$u_{\text{st}} = \left(\frac{1}{2} \quad \frac{1}{2} \cdot \frac{1-(1+\lambda_a\tau)e^{-\lambda_a\tau}}{1-e^{-\lambda_a\tau}} \quad \frac{1}{2} \cdot \frac{\lambda_a\tau e^{-\lambda_a\tau}}{1-e^{-\lambda_a\tau}} \right).$$

Az L és S időszakok hossza T , az I szakaszok pedig $\text{GEO}(1-e^{-\lambda_a\tau})$ darab τ hosszú ablakból állnak, aminek várható értéke $\frac{\tau e^{-\lambda_a\tau}}{1-e^{-\lambda_a\tau}}$.

Réselt CSMA protokoll elemzése

A beágyazott stacionárius eloszlást újrásúlyozva az egyes időszakok átlagos hosszával megkapjuk az egyes időszakok hosszú távú időátlagát (ami lényegében a stacionárius eloszláshoz hasonló olyan folyamatokra, amikre nem teljesül a Markov-tulajdonság):

$$\begin{aligned}V_I &= \frac{\tau e^{-\lambda_a \tau}}{T(1 - e^{-\lambda_a \tau}) + \tau e^{-\lambda_a \tau}}, \\V_L &= \frac{T(1 - (1 + \lambda_a \tau)e^{-\lambda_a \tau})}{T(1 - e^{-\lambda_a \tau}) + \tau e^{-\lambda_a \tau}}, \\V_S &= \frac{T\lambda_a \tau e^{-\lambda_a \tau}}{T(1 - e^{-\lambda_a \tau}) + \tau e^{-\lambda_a \tau}}.\end{aligned}$$

A CSMA protokoll fő előnye, hogy a τ rövidebb időablakokban jelentősen nagyobb annak az esélye, hogy csak egy forrás ad egyszerre, és emiatt sokkal gyakoribb lesz a sikeres küldés.

Réselt CSMA protokoll elemzése

A réselt CMSA protokoll esetén a csatorna kihasználtsága

$$\rho = V_S = \frac{T\lambda_a\tau e^{-\lambda_a\tau}}{T(1 - e^{-\lambda_a\tau}) + \tau e^{-\lambda_a\tau}}.$$

Általában τ és λ a csatornára jellemző, adott érték, amikből λ_a a

$$\lambda_a = \frac{\lambda}{1 - \lambda_a\tau e^{-\lambda_a\tau}}$$

implicit egyenletből számolható ki. T értékét pedig mi állítjuk be numerikus optimalizálás alapján.

Csak egy példa: $\lambda = 0.3, \tau = 1$ esetén $\lambda_a = 0.413$, és $T = 10$ választással

$$\rho = 0.676.$$

Általában minél kisebb τ , annál magasabb ρ -t lehet elérni.

Réselt CSMA/CD protokoll elemzése

A CSMA/CD protokollnál az egyetlen különbség a CSMA-hoz képest, hogy ütközés esetén az elvesző időszak hossza nem T , hanem csak τ . Ilyenkor az u_{st} beágyazott stacionárius eloszlást ennek megfelelően súlyozzuk:

$$\begin{aligned}v_I &= \frac{\tau e^{-\lambda_a \tau}}{\tau(1 - \lambda_a \tau e^{-\lambda_a \tau}) + T \lambda_a \tau e^{-\lambda_a \tau}}, \\V_L &= \frac{\tau(1 - (1 + \lambda_a \tau)e^{-\lambda_a \tau})}{\tau(1 - \lambda_a \tau e^{-\lambda_a \tau}) + T \lambda_a \tau e^{-\lambda_a \tau}}, \\V_S &= \frac{T \lambda_a \tau e^{-\lambda_a \tau}}{\tau(1 - \lambda_a \tau e^{-\lambda_a \tau}) + T \lambda_a \tau e^{-\lambda_a \tau}}.\end{aligned}$$

A CSMA protokoll fő előnye, hogy a τ rövidebb időablakokban jelentősen nagyobb annak az esélye, hogy csak egy forrás ad egyszerre, és emiatt sokkal gyakoribb lesz a sikeres küldés.

Réselt CSMA protokoll elemzése

A réselt CMSA/CD protokoll elemzése analóg a CMSA-val, például

$$\rho = V_S = \frac{T\lambda_a\tau e^{-\lambda_a\tau}}{\tau(1 - \lambda_a\tau e^{-\lambda_a\tau}) + T\lambda_a\tau e^{-\lambda_a\tau}}.$$

A λ_a -ra vonatkozó egyenlet is identikus:

$$\lambda_a = \frac{\lambda}{1 - \lambda_a\tau e^{-\lambda_a\tau}}.$$

T értékét pedig mi állítjuk be numerikus optimalizálás alapján, csak ezúttal a fenti ρ formulát kell optimalizálni.

Összességében a CSMA-hoz képest amiatt, hogy itt csak τ idő veszik el ütközésnél (a CSMA-nál T), a kihasználtsággal még magasabbra is fel lehet menni, pl. 0.8 fölé. Általában minél kisebb τ , annál magasabb ρ -t lehet elérni. (Speciálisan $\tau \rightarrow 0$ esetén $\rho \rightarrow 1$.)