1.7. Forgalomelmélet

Szerző: dr. Molnár Sándor

Lektor: dr.Jereb László

1.7.1. Bevezetés

A forgalomelmélet [1.7.1], [1.7.7] a távközlési hálózatok teljesítményanalízisének és tervezésének az alaptudománya. Agner Krarup Erlang dán matematikus (1878-1929) [1.7.6], volt az első, aki a XX. század elején megfogalmazta a matematikai problémát. Az elmélet a telefonhálózatok fejlesztésével párhuzamosan fejlődött, és lényegi elemévé vált a klasszikus távközlési hálózatok tervezésének [1.7.7].

A forgalomelmélet, a kapcsolási és hálózati technikák átalakulása következtében lényegi változáson ment át az elmúlt évtizedben, és magabá olvasztotta az operációkutatás és a sorbanálláselmélet legújabb eredményeit is. A forgalomelmélet folyamatos evolúciója figyelhető meg, ahogyan a különféle tudományágak erdeményeit magába integrálja és alkalmazza. A forgalomelmélet tárgya a forgalmi igények, hálózati erőforrások és a teljesítményjellemzők közötti kapcsolat matematikai modellezése. A forgalmi igények a természetük szerint statisztikusak, így a sztochasztikus folyamatok elméletéből származtathatóak.

Ebben a fejezetben először áttekintjük a hálózati forgalom legfontosabb jellemzőit. A forgalom természete alpavetően meghatározza azt a forgalomelméletet ami ma a rendelkezésünkre áll. Ezután áttekintjük a forgalomelmélet elemeit, beleértve a jelöléseket, a rendszerek típusait és a legfontosabb forgalomelméleti összefüggéseket. Az ismertetett alapok alkalmazása a 3.3 fejezetben található, ahol a forgalmi modelleket és tervezési elveket ismertetjük.

1.7.2. A hálózati forgalom jellemzői

A hálózati forgalom jellemzői a mai adathálózatokban (pl. Internet) teljesen eltérnek a klasszikus telefonhálózatok forgalmának jellemzőitől, és egy bonyolultabb problémával kell szembenéznünk [1.7.8], [1.7.9]. Az eltérés lényege, hogy a telefonforgalom lényegileg *statikus* jellegű. Ezért lehetett tipikus felhasználói viselkedést találni. A forgalmi jellemzők *korlátozott változékonysága* pedig lehetővé tette, hogy átlagértékekkel számoljunk, melyek jól leírták a forgalom jellegét.

A telefonhálózatok forgalmának statikus jellege tette lehetővé, hogy olyan "univerzális törvényeket" találjunk, mint a hívások keletkezésének *Poisson természete* [1.7.8], [1.7.9]. Ez a "szabály" azt mondja, hogy a hívások függetlenek és a hívások közötti idő exponenciális eloszlású. A Poisson hívásérkezési modellnek rendkívül nagy volt a népszerűsége az elmúlt ötven évben. A Poisson modell sikere az egyszerűségében rejlik, ami a gyakorlat számára egy fontos kritérium.

Egy hasonló "univerzális törvény" az, hogy a POTS forgalomban a hívások tartásideje közelítőleg *exponenciális eloszlású*. Ez a modell szintén a legtöbb esetben igen jól közelíti a valóságos folyamatokat. Ezenkívül értéke az egyszerűség és analitikus kezelhetőség. Annak ellenére használták ezt a modellt, hogy sokszor a telefonhívások tartásidejének eloszlása eltért az exponenciális jellegtől. Ez az eltérés azonban nem okozott jelentős hibákat az analízisben, ami szintén a Poisson hívásérkezési modellnek köszönhető. Ugyanis számos teljesítményjellemző Poisson érkezési folyamat esetén nem függ a tartásidő eloszlásától, csak annak átlagértékétől.

Jelentős változás következett be ezen "univerzális törvények" érvényességét illetően amikor a telefonhálózatokat nem csak beszéd, hanem FAX és Internet hozzáférésre is használni kezdték. Ezen szolgáltatások statisztikus természete ugyanis jelentősen eltér a beszédforgalométól. Például a hívások időtartama sokkal hosszabb és sokkal változékonyabb, mint a beszédhívások. A web népszerűségének növekedésével pedig egyre több és több ember kezdte el használni a klasszikus telefonhálózatot az Internet elérésére. Ezek a változások késztették arra a kutatókat, hogy felülvizsgálják a régi modelleket.

Az adathálózatok forgalmának természete jelentősen eltér a telefonhálózatokétól, de minden olyan próbálkozás, hogy a telefonhálózatoknál jól bevált "univerzális törvényeket" itt is találjunk, eddig kudarcot vallott [1.7.8]. Ennek a legfőbb oka, hogy az adathálózatok forgalma sokkal változékonyabb mint a beszédforgalom. Leegyszerűsítve azt is állíthatjuk, hogy lehetetlen találni általános modellt, mert az adatkommunikációban minden egyes kapcsolat ideje a nagyon rövidtől az extrém hosszúig változhat és az adatsebesség is széles tartományban bármi lehet. Az adatforgalomnak nincs meg az a homogén természte, mint a beszédforgalomnak. Az adatforgalom börsztösségének az egyik fő oka az, hogy itt többnyire gépek kommunikálnak egymással és nem emberek.

Az adatforgalom nagy változékonyságát megtaláljuk mind az *idő* tartományában (a forgalom összefüggőségi kapcsolatai nem csengenek le exponenciális gyorsasággal, mint a beszédforgalom esetében hanem annál sokkal lassabban, hosszú idejű összefüggések is jelen vannak) mind a *méret* tartományában (a forgalmi egységek méretének eloszlása nem exponenciális, mint beszédforgalomnál, hanem a hosszú farkú eloszlások a tipikusak, pl. a web által letöltött objektumok mérete). Ezen tényezők miatt új modellek és technikák kifejlesztése fontossá vált. A hosszabb időskálájú összefüggőségi struktúrát statisztikusan leírhatjuk a *hosszú idejű összefüggőséggel* (long-range dependence, LRD) ahol az autokorrelációs függvénynek hatványalakú lecsengése van. Az extrém méretbeli változékonyságot pedig a végtelen szórású *hosszú farkú eloszlások* segítségével (heavy-tailed distributions*)* írhatjuk le, ami pl. a Pareto eloszlással való modellezést jelentheti. A hatványalakú lecsengési tulajdonságok mind térben és méretben gyakran okozzák azt, hogy a forgalomnak *fraktális* természete van [1.7.8].

A fraktáltulajdonsgok egyik fontos megjelenése az *önhasonlóság*. Ez azt jelenti, hogy a forgalom számos statisztikus jellemzője azonos marad több időskálán keresztül. Az egyszerű önhasonló modellek, amelyeket az elmúlt évtizedben széleskörben alkalmaztak úgy tűnik sikeresen tudják modellezni a fraktális adatforgalmat. A jelenlegi kutatások azonban azt mutatják, hogy a forgalomnak sokkal kifinomultabb börsztszerkezete van, amelyet inkább a *multifraktál* modellek segítségével lehet leírni. Ahol a monofraktál jellegű önhasonló modellek nem megfelelőek, ilyen multifraktál modellek alkalmazása válik szükségszerűvé [1.7.8].

Az adatforgalom jellegének nagy változékoysága mellett egyéb tényezők is megnehezítik az adatforgalom modellezését. Az Internet forgalma megduplázódik minden évben. Ez a gyors forgalom növekedés és az előre nem látható új, esetleg tömegesen jelentkező alkalmazások ("killer applications") felboríthat minden jövőre vonatkozó forgalmi predikciót. Mindamellet eddig az Internet történetében csak három ilyen alkalmazás volt, ami drasztikusan megváltoztatta az Internet forgalom természetét (az e-mail, a web és a manapság terjedő Napster (zenei műsorok cseréje) tipusú alkalmazások), de senki sem tudja, hogy mikor jelenik meg esetleg egy olyan népszerű alkalmazás ami alapvetően megváltoztatja az Internet forgalmának összetételét. A helyzet még bonyolultabbá válik, ha a különböző alkalmazások eltérő minőségi követelményeit (Quality of Service, QoS) is figyelembe vesszük, amelyek eltérő forgalmi jelleget is vonnak maguk után.

Annak érdekében, hogy leírhassuk a forgalom jellemzőit mind a streaming mind az elasztikus forgalom esetében számos modellt fejlesztettek ki. A megfelelő modell alapján pedig kidolgozható az alkalmas méretezési technika. A 3.3. fejezetben a fontosabb modelleket és méretezési elveket ismertetjük.

1.7.3. A forgalomelmélet alapfogalmai

Ebben a fejezeten a legfontosabb forgalomelméleti alapelveket ismertetjük [1.7.1].

1.7.3.1. Fogalmak és jelölésrendszer

A hálózatban egy kapcsolat felépítési igényt *hívásnak* nevezünk, amelyet egy *előfizető* kezdeményez. A hívás időtartama a *tartási idő* vagy *kiszolgálási idő*. A *forgalom* az egyégnyi időre eső teljes tatásidő. A forgalom egysége az *erlang* (erl vagy E) amelyet a forgalomelmélet megalapítójáról neveztek el.

A forgalomnak a következő fontos tulajdonságai vannak:

1. A forgalom (felajánlott forgalom) $a=ch$ (erl) ahol c az egységnyi idő alatt kezdeményezett hívások száma és h az átlagos tartásidő.
2. A forgalom (felajánlott forgalom) egyenlő az átlagos tartásidő alatt kezdeményezett hívások számával.
3. A forgalom (átvitt forgalom) amit egy trönk továbbít az egyenlő a trönk foglaltságának valószínűségével.
4. A forgalom (átvitt forgalom) amit egy trönkcsoport továbbít az egyenlő a csoportban levő foglalt trönkök számának átlagával.

1.7.3.2. Forgalomelméleti rendszerek osztályozása

Kapcsolórendszer a bemeneti portokat köti össze a megfelelő (kijelölt, kiválasztott) kimeneti potrtokkal. Egy rendszert *teljesen elérhetőnek* nevezünk, ha bármelyik bemeneti port bármelyik kimeneti porthoz csatlakoztatható. *Torlódásnak* nevezzük a rendszer azon állapotát amikor valamilyen kapcsolat nem létesíthető a foglalt kimeneti portok vagy foglalt belső utak miatt. A rendszert *várakozásos rendszernek* nevezzük, ha torlódás esetén a bejövő hívás várakozni tud. Amennyiben nincs várakozási lehetőség torlódás esetén, a rendszert *veszteséges rendszernek* nevezzük.

A teljes elérhetőségű rendszereket a kövekezőképpen írhatjuk le [1.7.1]:

1. Bemeneti folyamat: Ez a hívások érkezésének folyamatát írja le.

2. Kiszolgáló mechanizmus: Ez leírja a kimenetek számát, a kiszolgálasi idő eloszlását, stb.

3. Sorbanállási diszciplína: Ez specifikálja a hívás kezelésének módját torlódás esetén. Késleltetéses rendszerekben a legegyszerűbb sorbanállási szabály a “first-in first-out” (FIFO), “last-in first-out” (LIFO), prioritásos rendszerek, processzor megosztás, stb.

A teljes elérhetőségű rendszerek osztályozására a *Kendall jelölésrendszert* használjuk [1.7.1], [1.7.3], [1.7.4] David A. Kendall brit statisztikus tiszteletére:

A/B/C/D/E-F

ahol

- *A* jelöli az érkezések közötti idő eloszlását,
- *B* jelöli a kiszolgálási idő eloszlását,
- *C* jelöli a párhuzamos kiszolgálók számát,
- *D* jelöli a rendszer kapacitását,
- *E* jelöli a véges felhasználó populáció nagyságát
- *F* jelöli a sorbanállási szabályt.

A következő jelölések használatosak:

- M: exponenciális (markovi)
- E_k: k-állapotú Erlang
- H_n: n-rendű hiperexponenciális
- D: determinisztikus
- G: általános
- GI: általános független
- MMPP: markov modulált Poisson folyamat
- MAP: Markov érkezési folyamat

Például az $M/M/1/\infty/\infty$-FCFS jelölés egy olyan sorbanállási rendszert reprezentál, melyben Poisson érkezések vannak és a kiszolgálási idő exponenciális eloszlású. A rendszernek egy kiszolgálója van és végtelen méretű tárolója. A felhasználók populációja végtelen és “először be először szolgál” a kiszolgálási szabály.

1.7.3.3. Alapösszefüggések

PASTA

Poisson érkezési folyamat estén (exponenciális az érkezések közötti idők eloszlása) állandósult állapotban egy tetszőleges pillanatban az aktív kapcsolatok számának eloszlása egyenlő az érkezés pillanatában levő kapcsolatok számának eloszlásával. Ennek az összefüggésnek *PASTA* (Poisson arrivals see time averages) a neve [1.7.1] mert ez a valószínűség egyenlő az aktív kapcsolatok átlagos időhányadával ha hosszú időn keresztűl vizsgáljuk a rendszert.

Markov tulajdonság

Exponenciális érkezések közötti idők eloszlása esetén egy tetszőleges időpillanat után a hátralévő idő eloszlása is exponenciális ugyanazzal a paraméterrel, az a modell, melynél az érkezések közötti idő és a kiszolgálási idő is exponenciális *Markovi modell* [1.7.1]. Ettől eltérő esetekben a modellt *nem Markovi modellnek* nevezzük.

Little Formula

Az $N=\lambda W$ összefüggést *Little formulának* [1.7.1], [1.7.3], [1.7.4] nevezzük, ahol N a rendszerben tartozkodó igények átlagos száma, λ az igények átlagos érkezési intenzitása és W az átlagos várakozási idő a rendszerben. A Little formula érvényes minden stacionárius rendszerre ahol a rendszerben igények nem születnek és nem vesznek el.

Veszteségi Formula

Annak a valószínűsége, hogy egy tetszőleges igény elvész a rendszerben [1.7.5]

$$P_{loss} = 1 - \frac{1-\phi}{\rho},$$

A veszteségi formula többkiszolgálós rendszerekre is érvényes, ahol ρ az egy kiszolgálóra jutó átlagos kihasználtság és ϕ a valószínűsége annak, hogy egy tetszőleges kiszolgáló szabad.

Hátralévő munka és a rendszerben levő igények száma

Állandó kiszolgálási idejű egy kiszolgálós rendszerben a következő az összefüggés a hétralévő munka (unfinished work) V_t és a rendszerben levő igények száma X_t között: $X_t = \lceil V_t \rceil$. Ez alapján a következő komplementer eloszlásfüggvény írható fel [1.7.5]:

$P(X_t > n) = P(V_t > n)$, ahol *n* egész.

Csomagvesztési valószínűség és sorhossz farokeloszlás

Tekintsünk egy diszkrét idejű *G/D/1* rendszert állandó csomaghosszal (cellák). A cellavesztési valószínűségre a következő felső korlátot írhatjuk fel [1.7.5]

$$\rho P_{loss} \leq P(X_t^{\infty} > K),$$

ahol ρ a kihasználtság, X_t^{∞} a sorhossz egy hipotetikus végtelen kapacitású sorban.

Az általánosított Beneŝ formula

Tekintsünk egy kiszolgálási rendszert végtelen kapacitású tárolóval. Tételezzük fel, hogy a rendszer stacionárius, tehát a 0 időpont egy tetszőleges időpontot reprezentálhat. A kiszolgálási kapacitás 1 egységnyi munka egységnyi idő alatt. A rendszerben levő munka komplementereloszlás függvénye a 0 időpntban [1.7.5]

$$P(V_0 > x) = \int_{u>0} P(\xi(u) \geq x > \xi(u+du) \qquad \text{és} \qquad V_{-u} = 0),$$

ahol, *u* egy idő, $\xi(t) = A(t) - t$, $t \geq 0$, és *A(t)* a rendszerbe érkező munka idő egységben kifejezve a [-*t*,0) intervallumban. Ez az eredmény érvényes a legtöbb számunkra érdekes sorbanállási rendszerre és rendkívűl hasznosnak bizonyult a forgalomelméletben.

1.7.4. Az M/G/1 sorbanállási rendszer

Az az egykiszolgálós sorbanállási rendszer melyben az igények Poisson folyamat szerint érkeznek és a kiszolgálási idő tetszőleges eloszlású (*M/G/1*) egy fontos kategória. A következőkben áttekintjük ezen rendszer lényeges jellemzőit [1.7.1], [1.7.2], [1.7.3], [1.7.4].

A következő jelöléseket használjuk:

- *W*: várakozási idő a sorban
- *T*: a rendszer válaszadási ideje
- N_q: igények száma a sorban
- *N*: igények száma a rendszerben
- *S*: kiszolgálási idő

Az átlagos várakozási idő és az igények átlagos száma a sorban az *M/G/1* rendszerben a következőképpen számítható [1.7.2]:

$$\overline{W} = \frac{\rho E(S)(1+c_S^2)}{2(1-\rho)}, \quad \overline{N}_q = \frac{\rho^2(1+c_S^2)}{2(1-\rho)}$$

ahol ρ a kihasználtság és c_S^2 a kiszolgálási idő relatív szórásnégyzete: $c_s^2 = \frac{E^2(S)}{S^2}$.

A rendszerben levő igények számának eloszlása a Pollaczek-Khinchin egyenlet segítségével számolható [1.7.2]:

$$G_N(z) = L_S(\lambda(1-z))\frac{(1-\rho)(1-z)}{L_S(\lambda(1-z))-z},$$

ahol $G_N(z)$ az *N* generátorfüggvénye, $L_X(s)$ az *X* Laplace transzformáltja és λ a Poisson folyamat érkezési intenzitása. Ez alapján a legfontosabb *M/G/1* rendszerekre a következő eredmények adódnak.

Sorbanállási rendszer	Sorhossz eloszlás P(N=n)
M/M/1	$(1-\rho)\rho^n$
M/H$_2$/1	$q(1-\alpha_1)\alpha_1{}^n + (1-q)(1-\alpha_2)\alpha_2^n$
M/D/1	$(1-\rho)\sum_{k=0}^{n} e^{k\rho}(-1)^{n-k}\frac{(k\rho+n-k)(k\rho)^{n-k-1}}{(n-k)!}$
M/E$_k$/1	$(1-\rho)\sum_{j=0}^{n}(-1)^{n-j}\frac{\alpha^{n-j-1}}{(1-\alpha)^{kj}}\left[\binom{kj}{n-j}\alpha+\binom{kj}{n-j-1}\right]$

A táblázatban H_2 a hiperexponeniális eloszlást jelenti α_1, α_2 és q paraméterekkel. E_k jelöli a k állapotú Erlang eloszlást α és k paraméterekkel. Ez a két eloszlás azért fontos, mert segítségükel tetszőleges *M/G/1* rendszert közelíthetünk. Ahol a kiszolgálási idő relatív szórásnégyzete 1 vagy az alatt van az *M/E$_k$/1* rendszert, ahol ez az érték 1 vagy annál nagyobb az *M/H$_2$/1* rendszert használhatjuk az approximációhoz.

1.7.5. Általános sorbanállási rendszerek

Az általános sorbanállási rendszerek (*G/G/n* sorok) rendszerint nehezen kezelhetőek, de van néhány olyan alosztály amely analitikusan is jól kezelhető. Például az *G/M/1* tipusú sorbanállási rendszerek kevésbé fontosak mint az *M/G/1* duálpárjuk, de az analízisük mégis egyszerűbb. Egy fontos eredménye a *G/G/1* rendszerek analízisének a *Lindley's integrál egyenlet* [1.7.1], [1.7.3] amivel a stacionárius várakozási idő eloszlása számítható:

$$F_W(t) = \int_{-\infty}^{t} F_W(t-v)dF_U(v) \,,$$

ahol *U =S-A* és *A* jelöli az időt két egymás után érkező igény között.

1.7.6. Forgalomelméleti technikák

A klasszikus sorbanállási módszereken kívül számos approximációt, korlátot és technikát dolgoztak ki a forgalomelméleti rendszerek kezelésére. Ebben a fejezetben áttekintünk néhányat a legfontosabbak közül.

A *folyadékmodell approximáció* [1.7.3] egy nagyon hasznos approximáció, amikor a vizsgált időskálán rengeteg forgalmi egység talalható (pl. csomagok).

Ebben az esetben a forgalmat tekinthetjük egy folytonos folyamnak, hasonlóan ahogy a folyadék áramlik egy csőrendszerben. Legyen *A(t)* és *D(t)* a (0,*t*) intervallumban érkező, illetve távozó igények száma. Az igények száma a rendszerben adott *t* időpontban *N(t)=A(t)-D(t)*, ahol feltételezzük, hogy a rendszer eredetileg üres volt. A nagy számok gyenge törvénye alapján amint *A(t)* növekszik egyre közelebb kerül annak átlagértékéhez, és ugyanez érvényes *D(t)* értékére is. A folyadékmodell approximációban egyszerűen az *A(t)* és a *D(t)* véletlen változókat helyettesítjük azok átlagával. Így végeredményben egy folytonos determinisztikus folyamatot kapunk. A folyadékmodelleket nagyon gyakran használják a forgalomelméletben.

A folyadékmodell approximáció az átlagértékeket használja, de az érkezési folyamat változékonysága ezen értékek körül nincs számításba véve. A *diffúziós approximáció* [1.7.3] kiterjeszti ezt a modellt és a központ határeloszlás tétel alapján az átlag körüli változásokat normális eloszlással modellezi. A diffúziós approximáció segítségével bonyolult sorbanállási rendszereket analizálhatunk. Például a complex *G/G/1* rendszer sorhossz elszlása számítható diffúziós approximáció segítségével.

Egy érdekes módszer amit a *maximum entrópia módszerének* [1.7.3] hívnak az információelméletből származik. Az alapelv a Bernoulli féle elégtelen információk elve, amely azt állítja, hogy ha semmilyen információnk nincs egy valószínűségi változóról, akkor egyenletes eloszlásúnak feltételezhetjük. Egy valószínűségi változó entrópiája minimum (zéro) amikor értéke ismert, bizonyos. Az entrópia maximális amikor a változó egyenletes eloszlású, mert egy esemény kimenetelének maximális a bizonytalansága. Az ötlet az, hogy az entrópiát maximalizálhatjuk külső feltételek ismeretében. A módszert sikerrel alkalmazzák a sorbanálláselméletben.

Egyéb módszereket, mint pl. a sorbanállási hálózatok számos megoldási technikával (fixpontos módszer, dekompozíciós technikák) szintén szép számmal fejlesztettek ki. Az érdeklődő olvasók az irodalomjegyzékben találnak referenciákat ezekre.

Irodalomjegyzék

[1.7.1] H. Akimaru. K. Kawashima: Teletraffic, Theory and Applications, Springer-Verlag, 1999.

[1.7.2] R. Nelson: Probability, Stochastic Processes, and queueing Theory, Springer-Verlag, 1995.

[1.7.3] P. G. Harrison, N. M. Patel: Performance Modelling of Communication Networks and Computer Architectures, Addison-Wesley, 1993.

[1.7.4] R. Jain: The Art of Computer Systems Performance Anaysis, Wiley, 1991.

[1.7.5] J. Roberts, U. Mocci, J. Virtamo (eds.), Broadband Network teletraffic, Springer-Verlag, 1996.

[1.7.6] E. Brockmeyer, F. L. Halstrom, A. Jensen: The Life and Works of A. K. Erlang, Acta Polytechnica Scandinavica, 1960.

[1.7.7] R. Syski, Introduction to Congestion Theory in Telephone Systems, Oliver and Boyd Ltd. 1960.

[1.7.8] W. Willinger, V. Paxson: Where Mathematics Meets the Internet, Notices of the American Mathematical Society, vol.45, no.8, Aug. 1998, pp. 961-970.

[1.7.9] J. Roberts, Traffic Theory and the Internet, IEEE Communications Magazine, January 2000.