3.3. Forgalmi modellek és forgalmi méretezés

Szerző: dr. Molnár Sándor

Lektor: dr. Frajka Béla

3.3.1. Bevezetés

A hálózati forgalom természetéről, valamint a forgalomelméleti alapokról az 1.7. fejezetben adtunk áttekintést. Ezeket az alapokat felhasználva ez a fejezet áttekintést ad a legfontosabb forgalmi modellekről amelyek a hálózati forgalom modellezésére alkalmasak lehetnek. Áttekintjük a forgalmi modell típusokat a legegyszerűbbektől a legbonyolultabbakig. A gyakorlatban a modelltípus kiválasztásánál mindig kompromisszumot kötünk, hogy a modell kellően komplex legyen ahhoz, hogy lehetőleg a legjobban leírja a valóságos forgalom természetét, de ugyanakkor a lehető legegyszerűbb legyen, hogy matematikailag is jól tudjuk kezelni. A kiválasztott forgalmi modellt általában valamilyen sorbanállási feladatban alkalmazzuk, amelynek segítségével teljesítményjellemzőket határozhatunk meg.

A fejezetben áttekintést adunk a telefonhálózatok és az adathálózatok forgalomméretezési eljárásairól. Az adathálózatoknál külön hangsúllyal tárgyaljuk az Internet forgalmi méretezését.

3.3.2. Forgalmi modellek

A forgalom általában egyedi vagy több diszkrét egységekből áll (csomagok, cellák, stb.). Matematikailag a forgalmat a *pontfolyamatok* elméletével írhatjuk le. A pontfolyamatoknak két lehetséges leírásmódja létezik: a *számlálófolyamatok* segítségével vagy az *érkezések közötti idők sorozatának* leírásával [3.3.1]. A számláló folyamat $\{N(t)\}_{t=0}^{\infty}$ egy folytonos idejű, nem negatív, egész értékű sztochasztikus folyamat, ahol $N(t) = \max\{n : T_n \leq t\}$ az érkezések száma a (0,*t*] intervallumban. Az érkezések közötti idők sorozata egy valós értékű véletlen számsorozat $\{A_n\}_{n=1}^{\infty}$, ahol $A_n = T_n - T_{n-1}$ az *n*-ik és az azt megelőző érkezés közötti idő hossza. A forgalmat batch folyamatnak nevezzük kötegelt érkezések esetén. A batch folyamatok leírására definiálunk batch érkezési folyamatot $\{B_n\}_{n=1}^{\infty}$ ahol B_n az érkezések száma az *n*-dik batchben. Hasonlóan hasznos leírófolyamat az un. *munkahátralék folyamat* $\{W_n\}_{n=1}^{\infty}$. Ez leírja az *n*-dik érkezésben mennyi W_n munka érkezett a rendszerbe.

A következőkben forgalmi modelleket ismertetünk, melyek forgalom generálására használhatóak és jellemezhetőek a következő sorozatok valamelyikével: $\{N(t)\}_{t=0}^{\infty}, \{A_n\}_{n=1}^{\infty}, \{B_n\}_{n=1}^{\infty}, \{W_n\}_{n=1}^{\infty}$.

A **felújítási folyamatban** $\{A_n\}_{n=1}^{\infty}$ független és azonos eloszlású valószínűségi változók sorozata [3.3.1], [3.3.2]. Ez a modell egyszerű, de sok esetben nem valósághű, mert nem tudja modellezni a valódi forgalom erős korrelációs struktúráját.

A **Poisson folyamat** [3.3.1], [3.3.2] olyan felújítási folyamat, melynél az érkezések közötti idők $\{A_n\}_{n=1}^{\infty}$ exponenciális eloszlásúak λ paraméterrel. A Poisson folyamatot definiálhatjuk a számláló folyamat segítségével is, ahol $\{N(t)\}_{t=0}^{\infty}$ folyamatnak független és stacionárius növekményei vannak Poisson határeloszlással: $P\{N(t)=n\}=\exp(-\lambda t)(\lambda t)^n/n!$ A Poisson folyamatot egyszerűsége és néhány elegáns matematikai tulajdonsága miatt gyakran használják a forgalomelméletben. A telefonhívások érkezését a telefonhálózatokban általában Poisson folyamattal modellezik.

A Bernoulli folyamatok [3.3.1], [3.3.2] a Poisson folyamatok diszkrét idejű analóg folyamatai. Ebben a modellben az érkezés valószínűsége minden időrésben *p* és független a többi érkezéstől. A *k* darab időrésben vizsgált érkezések száma binomiális eloszlású: $P\{N(t)=n\}=\binom{k}{n}p^n(1-p)^{k-n}$ és az érkezések közötti idő geometriai eloszlású: $P\{A_n=j\}=p(1-p)^j$

A **fázis típusú felújítási folyamatok** [3.3.1], [3.3.2] olyan speciális felújítási folyamatok, melyeknél az érkezések közötti idő eloszlása fázis típusú eloszlás. Ez egy fontos modellosztály, mert analitikusan jól kezelhető és ugyanakkor minden eloszlás tetszőleges pontossággal közelíthető fázis típusú eloszlásokkal.

A **Markov-i forgalmi modellekben** [3.3.1], [3.3.2] függőségi viszonyt modellezhetünk az A_n véletlen sorozatban. A modell konstrukciója a következő:

tekintsünk egy diszkrét állapotterű Markov folyamatot $M = \{M(t)\}_{t=0}^{\infty}$. *M* a következőképpen viselkedik: az *i*-dik állapotban marad λ_i paraméterű exponenciális eloszlású tartásideig. Vegyük észre, hogy az eloszlás paramétere egyedül az állapottól függ. Ezután *j*-dik állapotba ugrik p_{ij} valószínűséggel. A Markov folyamat minden ugrásához rendelünk egy érkezést, így az érkezések közötti idő exponenciális. Ez a legegyszerűbb *Markov-i forgalmi modell*.

A **Markov-i felújítási folyamat** [3.3.1], [3.3.2] sokkal általánosabb, mint az egyszerű Markov folyamat. Ennek ellenére ez a folyamat még mindíg analitikusan kezelhető. A Markov felújítási folyamatot $R = \{(M_n, \tau_n)\}_{n=0}^{\infty}$ az $\{M_n\}$ Markov lánc és ennek az ugrálásai közötti idő $\{\tau_n\}$ segítségével definiálhatjuk. A következő szabályt vezetjük be: az (M_{n+1}, τ_{n+1}) pár eloszlása csak a jelenlegi állapottól M_n függ és nem függ a korábbi állapotoktól vagy a korábbi ugrálások közötti időktől. Ebben a modellben szintén az ugrálásokhoz rendelünk érkezéseket.

A **Markov-i érkezési folyamatok** (Markov arrival processes) (MAP) [3.3.1], [3.3.2], [3.3.3] a Markov felújítási folyamatok egy nagy osztályát képezik. A MAP folyamatokban az érkezések közötti idő fázis típusú és az érkezések a Markov folyamat abszorpciós állapotaiba ugrálásakor történnek. A folyamatot olyan eloszlásból indítjuk újra, ami az adott abszorpciós állapottól függ. A MAP folyamatok analitikusan kezelhetőek és egy nagyon jó modellezési tulajdonságokkal bírnak.

A **Markov modulált folyamatoknál** egy Markov lánc pillanatnyi állapota határozza meg az érkezések generálásának szabályát [3.3.1], [3.3.2]. Tekintsünk egy folytonos idejű Markov folyamatot $M = \{M(t)\}_{t=0}^{\infty}$ ahol az állapottér: *1,2,…m*. Amíg az *M* folyamat *k* állapotban van, az érkezések generálásának szabálya kizárólag a *k* állapottól függ. Amikor *M* egy másik állapotba kerül pl. *j*, akkor amíg ebben az állapotban van egy új, csak ezen állapot által meghatározott generálási szabály lesz érvényes. Másképp fogalmazva az érkezési szabályokat *M* modulálja. A moduláló folyamat természetesen a Markov folyamatnál jóval bonyolultabb is lehetne, de akkor a modell analitikusan kevésbé volna jól kezelhető.

A **Markov modulált Poisson folyamat** (Markov Modulated Poisson Process) (MMPP) [3.3.1], [3.3.2], [3.3.3] a leggyakrabban használt Markov modulált forgalmi modell. Ebben a modellben, amíg a moduláló Markov folyamat *k* állapotban van, az érkezések Poisson folyamat szerint történnek, amelyeknek intenzitása λ_k. A legegyszerűbb Markov modell a kétállapotú MMPP, ahol az egyik állapotot “ON” bekapcsolt állapotnak értelmezzük valamilyen Poisson intenzitással, a másik állapotot “OFF” kikapcsolt állapotnak nulla intenzitással. Ezt a modellt megszakított Poisson folyamatnak is hívják. Az ON-OFF modellek gyakran használatosak beszédforgalom modellezésére, ahol az ON állapot a beszéd, az OFF állapot a szünet modellezésére szolgál.

A **Markov-i állapotátmenet modulált folyamatokban** [3.3.1], [3.3.2] a Markov folyamat $M = \{M(t)\}_{t=0}^{\infty}$ állapotátmenet ugrásai modulálják az érkezések generálásának szabályát. Az állapotátmeneteket egy állapotpár segítségével írhatjuk le: az ugrás előtti és az ugrás utáni állapottal. Az érkezések száma B_n az n időrésben teljesen szabályozva van a moduláló lánc ugrásaival: $P\{B_n = k | M_n = i, M_{n+1} = j\} = t_{ij}(k)$ és teljesen független a múlt bármely más állapotinformációjától.

Az **általánosan modulált determinisztikus folyamatokban** (Generally Modulated Deterministic Process) (GMDP) [3.3.3] a forrás bármely állapotban lehet a lehetséges N állapotból. Amikor j állapotban van, a forgalom állandó λ_j intenzitással generálódik. A j állapotban eltöltött időt tetszőleges eloszlás meghatározhatja, de a gyakorlatban legtöbbször geometriai eloszlást használnak, hogy a modell analitikusan könnyen kezelhető legyen. A kétállapotú GMDP, ahol az egyik állapotban nulla a generálási intenzitás, a diszkrét megfelelője a már tárgyalt ON-OFF modellnek.

A **folyadék forgalmi modellezési technikában** a forgalmat egy folyamatosan áramló folyadéknak modellezzük és eltekintünk a valódi forgalom diszkrét egységekből álló természetétől [3.3.1], [3.3.2], [3.3.3]. Ez a modellezés akkor elfogadható, ha a forgalomnak a vizsgált időskálán rengeteg diszkrét egységét, pl. csomagokat figyelhetünk meg. A modell előnye az, hogy sokkal egyszerűbb, mint azok a modellek, melyek a forgalom diszkretizált egységei közötti struktúrát próbálják leírni. A legegyszerűbb folyadékmodellek két állapotot feltételeznek: egy ON állapotot, amikor a forgalom egy állandó λ rátával áramlik, és egy OFF állapotot. amikor nincs forgalom továbbítás. Az egyszerű analitikus kezelhetőség érdekében sokszor az ON és OFF állapotokat független exponenciális eloszlású valószínűségi változókkal modellezzük. Ebben az esetben a modell egy alternáló felújítási folyamat.

Az **autoregresszív forgalmi modellben** a múltban bekövetkezett forgalmi változások explicit függvénye határozza meg a jelenben bekövetkező forgalmi változásokat [3.3.1], [3.3.2], [3.3.3]. Ezen modellcsalád tipikus példái a *lineáris autoregresszív (AR) folyamatok*, a *mozgó átlag (MA) folyamatok*, az *autoregresszív mozgó átlag (ARMA) folyamatok* és az *autoregresszív integrált mozgó átlag (ARIMA) folyamatok.* Ezek a modellek nagyon hasznosnak bizonyultak VBR videó forgalom medellezésére.

A **TES modellek** (*Transform-Expand-Sample*) [3.3.1], [3.3.2] olyan modellcsalád, mely a következő követelményeknek felel meg: a határeloszlást és az autokorrelációs függvényt illeszti az empirikus adatokra. A TES modellek jól használhatóak pl. MPEG videó modellezésére.

A **fraktális Gauss zaj** (Fractional Gaussian Noise, FGN) [3.3.1] egy másodrendben önhasonló folyamat *H* hasonlósági paraméterrel, ahol *½<H<1.* Ez egy stacioner Gauss folyamat $X=\{X_k\}_{k=1}^{\infty}$, a következő autokorrelációs függvénnyel: $\rho_X(k)=\frac{1}{2}(|k+1|^{2H}-2|k|^{2H}+|k-1|^{2H}), k\geq 1$. Az FGN egyúttal hosszú idejű összefüggő folyamat (LRD) *H* paraméterrel: $\rho_X(k)\approx H(2H-1)|k|^{2H-2}, k\to\infty$. Az FGN egy jó forgalmi modell aggregált LRD forgalom modellezésére.

A **fraktális ARIMA folyamatok** (fractional ARIMA, FARIMA) [3.3.1] a klasszikus ARIMA(*p,q,d*) modellre épülnek, de a differencia operátor *d* parametere tört értékeket is felvehet. A FARIMA modellek sokkal rugalmasabbak, mint az FGN modellek, mert nem csak az LRD struktúrát, de egyúttal a rövid idejű összefüggőségi struktúra (short-range dependent, SRD) is jól modellezhető vele.

Az **M/Pareto modellben** λ intenzitású Poisson folyamattal érkeznek Pareto eloszlású börsztök [3.3.4]. A börszt ideje alatt az állandó intenzitású *r* rátával történnek érkezések. A börszt hosszát a Pareto eloszlás határozza meg, melynek paraméterei: $1<\gamma<2, \delta>0$: $P\{X>x\}=\begin{cases}\left(\frac{x}{\delta}\right)^{-\gamma}, & x\geq\delta\\ 1, & \text{egyébként}\end{cases}$. A modell LRD forgalmat generál $H=(3-\gamma)/2$ paraméterrel.

Az **alkalmazásorientált forgalmi modellek** azok, melyekkel közvetlenül valamilyen gyakorlati alkalmazás forgalmát modellezhetjük. Az eddig felsorolt forgalmi modellek (vagy ezek kombinációi) alkalmazhatóak számos alkalmazás forgalmának modellezésére. A következő táblázat tervezési irányelveket ad arra, hogy a legnépszerűbb alkalmazások modellezésére milyen típusú forgalmi modelleket használhatunk [3.3.5].

Alkalmazás	Modell	Eloszlás
TELNET	igény érkezések közötti idő	Exponenciális
	teljes hossz	Lognormális
	csomag érkezések közötti idő	Pareto
	csomag méret	többnyire 1 byte-os csomagok
FTP	igény érkezések közötti idő	Exponenciális
	elemek száma	Empirikus
	elemek mérete	Lognormális
CBR hang	igény érkezések közötti idő	Exponenciális
	teljes hossz	Exponenciális
	csomag érkezések közötti idő	Állandó
	csomag méret	Állandó
VBR videó telekonferencia	keret érkezések közötti idő	Állandó
	keret méret	Gamma
MPEG videó	keret érkezések közötti idő	Állandó
	jelenet hossz	Geometriai
	keret méret	Lognormális
WWW	igény érkezések közötti idő	Exponenciális
	dokumentum méret	Pareto

3.3.1. táblázat. Forgalmi modellek felhasználási területe és jellemzői

3.3.3. Klasszikus telefonhálózatok forgalomelméleti méretezése

A forgalomelmélet kulcsszerepet játszott a kezdetektől fogva a telefonhálózatok tervezésében és méretezésében. A telefonhívások modellezésére, a stacioner Poisson folyamatot feltételezve, a forgalom és a teljesítőképesség közötti

$$B = E(a,n) = \frac{a^n / n!}{\sum_{i=0}^{n} a^i / i!}$$

kapcsolatot a jól ismert Erlang veszteségi formula fejezi ki. Ez a formula megadja a hívásblokkolás valószínűségét *B*, ha a felajánlott forgalom *a* és a rendelkezésre álló vonalak száma *n*.

A formula kifejezi, hogy a blokkolási valószínűség a felajánlott forgalom egyszerű függvénye. Fontos megjegyeznünk, hogy a blokkolási valószínűség nem érzékeny a forgalom egyéb jellemzőire, pl. a hívás tartási idejének eloszlására. (A formula minden *M/G/n/n* sorbanállási rendszerre érvényes.) Ezt a formulát használták legtöbbet a forgalomelmélet történetében. A telefonhívásokat egymástól független módon, véletlenszerűen kezdeményezik, ezért a véletlen modellek, amelyek a forgalmas órákra stacioner forgalmat feltételeznek, alkalmasak voltak a mérnöki tervezésre. Mivel a telefonhívások pont-pont közötti állandó sávszélességet lefoglaló kapcsolatok, az Erlang formula kiválóan alkalmas a telefonhálózatok tervezési feladataira.

Az Erlang formulának számos továbbfejlesztett változata van a legkülönbözőbb hálózati helyzetekre adaptálva, de az Erlang veszteségi formula (és a kapcsolódó Erlang késleltetési formula) mind a mai napig a mérnökök jól felhasználható eszköze. Kétség nélkül állíthatjuk, hogy az Erlang formula aratta eddig a legnagyobb sikert a forgalomelmélet történetében.

Az Erlang veszteségi és késleltetési formulákon kívül számos technikát fejlesztettek még ki a telefonhálózatok tervezési problémáinak megoldására. Ilyenek például az ekvivalens véletlen módszer, mely Wilkinson nevéhez fűződik, a forgalom börsztösségének különféle leírásai a csúcsosság funkcionállal és diszperziós indexekkel, az Engset modell, stb. [3.3.11].

3.3.4. Az Internet forgalomelméleti méretezése

Jelenleg épp az Internet forgalomelméletének a születési pillanatait látjuk annak újdonságaival és szülési fájdalmaival együtt. Az Internet hálózatának tervezésében ma még a forgalomelmélet nem játszik jelentős szerepet és sokszor a hálózatok tervezői csak néhány tervezési ökölszabályt alkalmaznak. Amint azt a 1.7.2. fejezetben tárgyaltuk, az adatforgalom természete jelentősen eltér a beszédforgalom természetétől, és olyan hasonló univerzális szabályokat nem lehet találni, mint amilyenek a beszédforgalom modellezésénél nagy segítséget jelentettek. Új technikákat és modelleket kell kifejleszteni, hogy megbirkózzanak ezekkel a kihívásokkal. A következőkben áttekintjük az Internet forgalomelméleti tervezésének két valószínűsíthető alternatíváját. Az egyiket “nagy sávszélesség filozófiájának”, a másikat a “menedzselt sávszélesség filozófiájának” fogjuk hívni [3.3.12].

3.3.4.1 A nagy sávszélesség filozófiája

Egy széleskőrben elterjedt és jelentős álláspont manapság az, hogy nincs is igazából szükség komplikált forgalomelméletre az Internethez. Ennek az iskolának a képviselői azt állítják, hogy annak ellenére, hogy az Internet forgalma drasztikusan növekszik, a linkek kapacitása és a kapcsoló eszközök annyira olcsóak lesznek a jövőben, hogy az erőforrások túlméretezése kivitelezhető lesz. Ez a “*nagy sávszélesség filozófiája*”*.* Érdemes egy kicsit részletesebben megvizsgálnunk ennek az álláspontnak a realitását.

Az állítás képviselőinek a várakozása szerint az Internet átviteli és információs technológiája követni tudja az Internet forgalmának évenkénti duplázódási trendjét [3.3.10], továbbá ez a technológia olcsó megoldásokat tud majd kínálni. Technológiai szempontból ez a várakozás reálisnak mondható, legalább is a közeljövőt illetően. Gondoljunk bele, hogy ha a mai Internetben csak annyi változás történne, hogy a linkek kapacitása megnövekedne, akkor egy olyan csomagkapcsolt hálózat állna rendelkezésünkre, amely lehetővé tenne valós idejű kommunikációt mindenféle QoS támogatás nélkül! A jelenlegi “best effort ” típusú Internet erre képes lenne!

Egy másik fontos tényező az adatok lokalitásának elve. Ez azt jelenti, hogy rengeteg adat helyileg lesz fontos és így a “caching” fontossága is megnő [3.3.10]. Amennyiben elképzeljük azt, hogy az összes bitet, amit ma hard drive-okon tárolnak, továbbítani szeretnénk, akkor ehhez 20 évre lenne szükség. Ez a lokalitási trend egy relatív csökkenést fog várhatóan okozni a teljes továbbítandó információ tömegben.

Egy további tényező az, hogy a “streaming” tripusú forgalom, ami lényeges QoS támogatást igényel, várhatóan nem lesz domináns az Interneten [3.3.10]. Sokan várták azt, hogy ezen forgalom nőni fog, de mindez ideig ezek a várakozások nem teljesültek be és várhatóan nem is fognak. A statisztikák azt mutatják, hogy az igény ez iránt a forgalom iránt nem nő annyira, mint ahogy a linkek kapacitása nő. Vizsgáljunk meg egy egyszerű példát: legyen 1% “streaming” forgalmunk, ami QoS támogatást igényel. Két megoldási alternatíva kínálkozik. Kiépíthetünk egy QoS architektúrát, amivel megoldjuk a QoS támogatást vagy egyszerűen csak megnöveljük a linkek kapacitását 5%-al. A “nagy sávszélesség filozófia” hívői szerint a második megoldás olcsóbb. Továbbá azzal is érvelnek, hogy a multimédia alkalmazások “tárolj és továbbíts” technikát fogják használni a valós idejű “streaming” helyett. Ezt az álláspontot támogatja az is, hogy a mágneses tárolók kapacitása ugyanolyan arányban növekszik, mint a linkek kapacitása. Mindezeken túl a különféle hálózati szűk keresztmetszetek miatt (pl. vezeték nélküli átvitel) is fontos lesz az adatok lokális tárolása.

Nagyon tanulságos az is, ha az elmúlt évek kapacitásnövekedési trendjeinek okait vizsgáljuk. Megfigyelhetjük azt, hogy az emberek általában nem azért fizetnek ADSL vagy modem hozzáférésért, mert annyi adatot akarnak továbbítani, hogy szükségük van a nagy kapacitásra, hanem azért, mert amikor ráklikkelnek egy hiperlinkre, akkor azt az oldalt azonnal akarják látni a képernyőjükön! Tehát a nagy kapacitás nem a sok adat továbbítása miatt kell, hanem a gyors hozzáférés (kis késleltetés) miatt. Ugyanez az oka annak, hogy az elmúlt tíz évben a LAN-ok átlagos kihasználtsága tizedére csökkent. Az emberek a nagy kapacitást a kis késleltetésű hozzáférés miatt fizetik meg!

Valóban az erőforrások túlméretezése lesz a megoldás a jövő Internetében? Ezt ma még senki sem tudja. Azért is meglehetősen nehéz jóslásokba bocsátkozni, mert ez a kérdés nem csak technológiai tényezőktől függ, hanem politikai és gazdasági tényezők is jelentősen befolyásolják. Azért, ha egy óvatos becslést e sorok írója megtehet, akkor az mondható el, hogy ha az erőforrások túlméretezése megoldás is lesz a gerinchálózatokban, az kevésbé valószínű, hogy a hozzáférési hálózatokban ugyenez megtörténik. Azokban az esetekben pedig, ahol a túlméretezés nem megoldás, a rendelkezésre álló erőforrásokkal kell ügyesen bánnunk. Ez vezet a második alternatívához, a *“menedzselt sávszélesség elvéhez”.*

3.3.4.2 A menedzselt sávszélesség elve

Amennyiben korlátozott hálózati erőforrás áll rendelkezésünkre, valamilyen forgalmi szabályozásra van szükség a megfelelő link kapacitás és router memória hozzárendeléséhez ahhoz, hogy a kívánt QoS követelményeket minden forgalom számára biztosítani tudjuk. A QoS követelményeknek három fő kategóriája van: transzparencia, hozzáférhetőség és throughput [3.3.7]. A *transzparencia* a továbbított adatok időbeli és szemantikai integritását fejezi ki. Például az adatátvitel szemantikai integritása általában követelmény, de a késleltetés nem annyira fontos. A *hozzáférhetőség* méri a hozzáféréskérés elutasítási valószínűségét és a kapcsolat felépítési késleltetést blokkolás esetén. Ebben a kategóriában a blokkolási késleltetés a tipikus példa, ami egy gyakran használt jellemző telefonhálózatokban. A *throughput* a legfontosabb QoS mérték adathálózatokban. Például a mai Interneten egy 100Kbit/s throughput biztosítani tudja a legtöbb web lap kvázi azonnali továbbítását (egy másodperc alatt).

A forgalmi típusok természetük szerint két nagy csoportba oszthatóak: stream forgalom és elasztikus forgalom [3.3.7]. A *stream forgalom* olyan folyamokkal írható le, amelyeket időtartamuk és sebességük jellemez. A stream forgalom tipikus példái az audio és a valós idejű videó alkalmazások: telefon, interaktív videó szolgáltatások és videó konferencia. A stream forgalom idő integritását fontos megőrizni. A veszteség, a késleltetés és a dzsitter fontos QoS jellemzők ezen forgalomtípus esetén.

Az *elasztikus forgalom* digitális objektumokból (dokumentumok) áll, melyeket egyik helyről egy másikra szeretnénk továbbítani. A forgalom elasztikus, mert a folyam sebessége változhat külső hatások következményeképpen (pl. szabad kapacitás). Tipikusan elasztikus alkalmazások a web, az e-mail és a file transzfer. Elasztikus forgalom esetén a szemantikus integritást fontos megőrizni. Az elasztikus forgalom leírható az igények érkezési folyamatával és az objektumok méretének eloszlásával. A throughput és a válaszadási idő a tipikus QoS mértékei ennek a forgalomtípusnak.

A következő két alfejezetben áttekintjük a stream és az elasztikus forgalom menedzselésének alapelveit.

3.3.4.3 A streaming forgalom nyílt hurkú szabályozása

A streaming típusú forgalmat általában forgalmi szerződésre alapuló *nyílt hurkú, megelőző forgalmi szabályozás* módszereivel szabályozzák [3.3.7]. A forgalmi szerződés egy olyan sikeres megállapodás a felhasználó és a hálózat között, melyben a felhasználó forgalmi igényeit leíró paraméterek, és az igényelt QoS paraméterek szerepelnek. Ezen információk segítségével a hálózat végrehajt egy belépés szabályozási algoritmust, amely a kapcsolatot csak akkor fogadja el, ha az igényelt QoS biztosítható és a meglévő kapcsolatok QoS paraméterei sem sérülnek.

A szabályozás hatékonysága nagymértékben függ attól, hogy a teljesítményjellemzők milyen pontosan becsülhetőek a *forgalom leíró paraméterek* segítségével [3.3.6]. A gyakorlat azt mutatja, hogy nem könnyű jó forgalmi jellemzőket találni. A forgalmi jellemzőknek *egyszerűnek* (érthetőek legyenek a felhasználó számára), *hasznosnak* (erőforrás allokációs szempontból) és *szabályozhatónak* (a hálózat számára ellenőrizhetőnek) kell lenniük. A gyakorlat azt mutatja, hogy mindhárom szempontból ideális forgalomleírót lehetetlen találni. Például az ATM és az Internet szabványosítási törekvéseiben az elfogadott *token bucket forgalomleírók* jól szabályozható és ellenőrizhető forgalomleírók, de kevésbé hasznosak erőforráshozzárendelés szempontjából. A felhasználók használhatnak eljárásokat (pl. forgalomsimítás), amikkel biztosíthatják a deklarált forgalmi jellemzők betartását. A hálózat pedig a belépési pontokon alkalmazhat eljárásokat (forgalom felügyelet), amik ellenőrzik a deklarált forgalmi jellemzőket. Ezek az eljárások legtöbbször a már említett token bucket módszerekre épülnek.

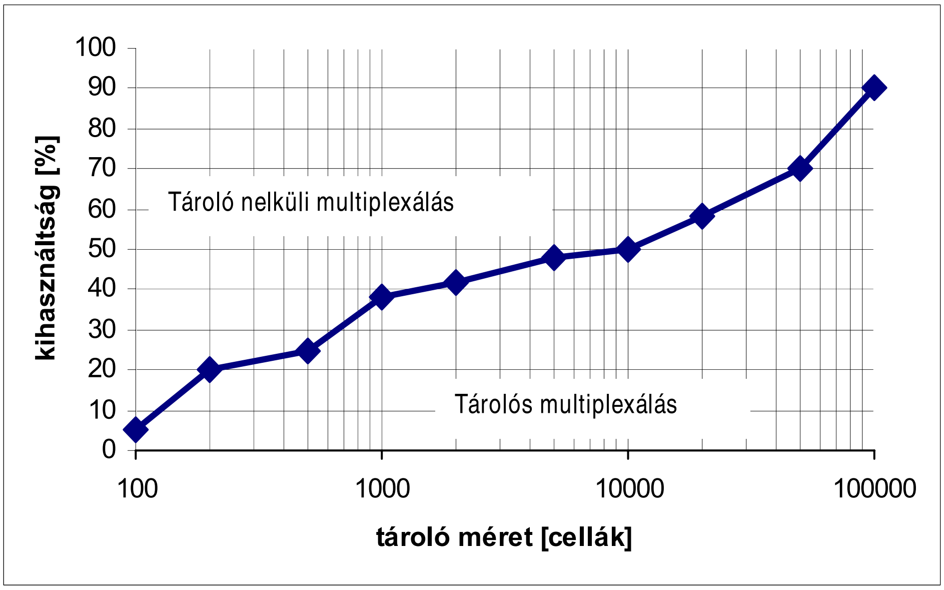
A legfontosabb típusai a nyílt hurkú forgalomszabályozásnak jelentősen eltérnek attól függően, hogy statisztikus multiplexálási nyereséget akarunk-e elérni vagy nem [3.3.7]. A következő tábla mutatja a fő kategóriákat:

Eljárás	**Tároló megosztás**	**Sávszélesség megosztás**
cúcssebességű kapacitás hozzárendelés	NEM	NEM
sebesség burkoló multiplexálás	NEM	IGEN
kapacitás megosztás	IGEN	IGEN

Amennyiben multiplexálási nyereséget nem kívánunk elérni, a legegyszerűbb eljárást, a csúcssebességű kapacitás kiutalást alkalmazhatjuk. Ez a módszer egyszerűen minden kapcsolat számára annak maximális sebességének megfelelő sávszélességet foglal le. A módszer előnye, hogy az egyetlen forgalmi jellemző a csúcssebesség. A beléptetési szabályozás nagyon egyszerű, csak azt kell ellenőrizni, hogy az igényelt csúcssebességek összege meghaladja-e a teljes kapacitást. A fő hátránya ennek a módszernek, hogy nagyon pazarló a kapacitással, mert statisztikailag csak az idő igen kis hányadában fordul az elő, hogy az összes kapcsolat csúcssebességgel továbbít.

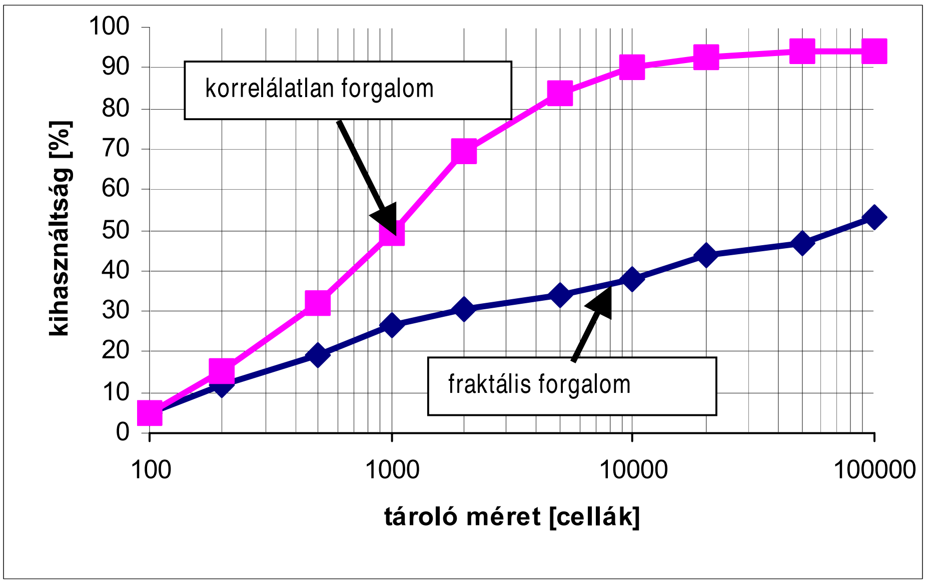
Ha a sávszélességet statisztikailag megosztjuk a kapcsolatok között, de a tárolókapacitást még nem, akkor a *rate envelope multiplexing* (sebesség-burkoló multiplexálás) esetéhez jutunk [3.3.6], [3.3.7], [3.3.8]. Az eljárást *tárolónélküli multiplexálásnak* is hívják, mert a folyadékmodelles analízisében nincs tárolóra szükség. A rate envelope multiplexing módszer esetében a cél az, hogy az aggregált folyam sebessége a teljes kapacitás alatt maradjon. Annak az eseménynek a valószínűsége, hogy az aggregált sebesség meghaladja a kapacitást, adott érték alatt kell maradjon, $P(\Lambda_t > c) < \varepsilon$, ahol Λ_t az aggregált forgalom sebessége, *c* a link kapacitása és ε a megengedett túlcsordulási valószínűség. A valóságban tárolók mindig kellenek, hogy az egyszerre érkező csomagok ne vesszenek el (cella szintű torlódás). A túlcsorduló forgalom elveszik és az átlagos veszteségi ráta: $E[(\Lambda_t - c)^+ / E(\Lambda_t)]$. A veszteségi ráta csak Λ_t stacioner eloszlásától függ, és nem függ annak időbeli függőségi struktúrájától. Ez egy fontos tényező, mert ez azt jelenti, hogy a korrelációs struktúrának nincs hatása a veszteségre. Ennek pedig az a fontos következménye, hogy a forgalommodellezés azon nagyon nehéz feladatára, hogy a korrelációs összefüggőségeket pontosan leírjuk, nincs szükség. A forgalmi struktúrának valójában van hatása a teljesítményjellemzőkre, de ezek elhanyagolhatóak, ha a veszteség kicsi. Például LRD forgalom eredményezhet hosszabb veszteségi periódusokat mint SRD forgalom, de ez a hatás elhanyagolható kis veszteségnél. A legfőbb hátránya ennek a módszernek, hogy az elérhető kihasználtság még mindig nem olyan jó.

Amennyiben a link kihasználtságot tovább akarjuk növelni a tárolót is statisztikusan meg kell osztanunk, lásd a 3.3.1. ábrát. Ez a *rate sharing* (kapacitás megosztás) módszere [3.3.6], [3.3.7], [3.3.8] amit *tárolós multiplexelésnek* is hívnak. A módszer ötlete az, hogy a tároló segítségével is eliminálhatunk néhány túlcsordulási veszteségi periódust. A cél a tárolóban levő sorhossz adott alacsony valószínűség alatt tartása, $P(Q > q) < \varepsilon$, ahol *q* a megengedett sorhossz is, *Q* az aktuális sorhossz és ε a megengedett valószínűsége annak, hogy a sorhossz meghaladja a megengedett sorhosszt. Ezzel a módszerrel sokkal nagyobb multiplexálási nyereség és kihasználtság érhető el.



3.3.1. ábra. Statisztikus multiplexálási alternatívák

A rate sharing (kapacitás megosztás) legfőbb hátránya az, hogy a veszteség adott tárolóméretnél és link kapaciásnál elég bonyolult módon függ a forgalmi jellemzőktől, beleértve a korrelációs struktúrát is. Például a veszteségi és késleltetési jellemzők nagyon bonyolultan számíthatóak, ha a forgalom LRD. Ez az oka annak, hogy a hívás belépést engedélyező eljárások sokkal bonyolultabbak, mint a rate envelope multiplexing esetében [3.3.8]. További probléma, hogy a komplex forgalomszabályozás dacára az elérhető kihasználtság még ennél az esetnél sem olyan nagy, ha a forgalom fraktális természetű, erős SRD és LRD tulajdonságokkal, lásd a 3.3.2. ábrát.



3.3.2. ábra. A korrelációs struktúra hatása

Sokféle hívés belépést engedélyező eljárást dolgoztak ki mind a rate envelope multiplexing, mind a rate sharing esetére [3.3.8]. A tapasztalat azt mutatja, hogy a legeredményesebb eljárások a mérés alapú módszerek, ahol az egyetlen forgalomleíró a csúcssebesség és a pillanatnyi elérhető sebességet valós időben becsüljük.

3.3.4.4 Az elasztikus forgalom zárt hurkú szabályozása

Az elasztikus forgalmat általában *reaktív zárt hurkú forgalomszabályozó módszerekkel* menedzselik [3.3.6], [3.3.7]. Ez az elve a TCP-nek az Interneten és az ABR-nek az ATM-ben. Ezek a protokollok a maximális szabad sávszélesség kihasználtságra törekszenek úgy, hogy a fennálló kapcsolatok között a sávszélességet fair módón osszák szét. Most a TCP-t vizsgáljuk meg, ami az Internet általánosan használt átviteli protokollja. A TCP-ben egy additív növelési és multiplikatív csökkentési algoritmust implementáltak. Amíg nincs csomagvesztés, a sebesség lineárisan nő, de csomagvesztés esetén a csomagtovábbítási sebességet az algoritmus felezi. Az algoritmus megpróbál egy olyan átlagsebességre beállni, ami a kapacitás és a pillanatnyi forgalom jellemzőitől függ. Az elérhető sávszélesség közelítőleg fair módon oszlik meg a TCP folyamok között.

A TCP-nek az egyszerű modellje [3.3.9], ami a legfontosabb viselkedését leírja, a jól ismert négyzetgyökös összefüggés a throughput (*B*) és a csomagvesztés (*p*) között:

$$B(p) = \frac{c}{RTT\sqrt{p}},$$

ahol *RTT* a TCP folyam körülfordulási ideje, és *c* egy konstans. Fontos megjegyeznünk, hogy ez az egyszerű formula csak számos feltétel mellett igaz: *RTT* állandó, *p* kicsi (1% alatt van) és a TCP forrásnak mindig van adata továbbításra. A TCP eljárásról feltételezzük, hogy a fast retransmit és recovery eljárások működnek (nincs timeout) és a slow-start fázist nem modellezzük. Ennél sokkal bonyolultabb TCP modellek is ismertek, de a négyzetgyökös összefüggés egy elég általános szabálya a TCP-nek.

Hívásbelépési eljárások kifejlesztése elasztikus forgalmakra egy nem lezárt kutatási terület [3.3.6], [3.3.7]. Ezekben az eljárásokban a szabályozásoknak úgy kell működniük, hogy biztosítsák a megfelelő throughput-ot túlterheléses állapotokban is, de másrészről elkerüljék a folyamok visszautasítását normális terhelési viszonyok között.

3.3.5. Záró gondolatok a forgalmi méretezésről

A megfelelő forgalmi modell választásának az a tétje, hogy milyen pontosan tudjuk megragadni a forgalom legfontosabb jellemzőit. Az így kiválasztott a forgalmi modellt alkalmazzuk a legtöbb forgalomelméleti rendszerben, ami a leggyakrabban egy sorbanállási rendszer. A legfontosabb kérdés az alapvető összefüggés a forgalmi jellemzők, a hálózati erőforrások és a teljesítményjellemzők között. Több sorbanállási rendszer analitikusan kezelhető (pl. Poisson, MMPP, MAP, stb.), de van több olyan is, amik nagyon nehezen analizálhatóak analitikusan (pl. ARIMA, TES, FGN, stb.). Jelenleg is fontos kutatási feladat olyan új technikák és modellek kifejlesztése, amelyek jól kezelhetőek és ugyanakkor jól megragadják a valóságos rendszer jellemzőit.

A fenti áttekintésünk mutatja, hogy az Internet forgalmi méretezése még nem megoldott feladat és számos nyitott probléma vár megoldásra. Ellentétben a telefonhálózatok forgalomelméletével, ami egy kiforrott és jól megértett terület, az

Internet forgalomelmélete és méretezési eljárásai a jövő kutatási feladatai közé tartoznak.

Irodalomjegyzék

[3.3.1] D. L. Jagerman, B. Melamed, W. Willinger: Stochastic modeling of traffic processes, In J. Dshalalow, ed., Frontiers in Queueing: Models, Methods and Problems. CRC Press, 1997. pp. 271-320.

[3.3.2] V. S. Frost, B. Melamed: Traffic Models for Telecommunications Networks, IEEE Communications Magazine, March 1994. pp 70-81.

[3.3.3] G. D. Stamoulis, M. E. Anagnostou, A. D. Georganas: traffic sources models for ATM networks: a survey, Computer Communications, vol. 17, no. 6, June, 1994. pp. 428-438.

[3.3.4] R. G. Addie, M. Zukerman, T. D. Neame: Broadband Traffic Modeling: Simple Solutions to Hard Problems, IEEE Communications Magazine, August 1998. pp. 88-95.

[3.3.5] B. O. Lee, V. S. Frost, R. Jonkman: NetSpec 3.0 source Models for telnet, ftp, voice, video and WWW traffic, 1997.

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

[3.3.7] J. Roberts, Engineering for Quality of Service, in the book of Self-Similar Network Traffic and Performance Evaluation, (eds. K. Park, W. Willinger), Wiley, 2000.

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

[3.3.9] J. Padhye et al. Modeling TCP Throughput: A Simple Model and Its Empirical validation, Proc. SIGCOMM'88, ACM, 1998.

[3.3.10] A. Odlyzko: The history of communications and its applications for the Internet, available at http://www.research.att.com/~amo/doc/complete.html, 2000.

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

[3.3.12] S. Molnár, IP hálózatok forgalmi méretezése, Magyar Távközlés, September 2000.