TÖBBOSZTÁLYOS, ÖSSZEFÜGGŐ FORGALOMMAL RENDELKEZŐ SORBANÁLLÁSI HÁLÓZATOK HATÉKONY, MÁTRIX-ANALITIKUS MEGOLDÁSA

MTA doktori értekezés tézisei

HORVÁTH GÁBOR

Budapest, 2017

dc_1412_17

1 BEVEZETŐ

A sorbanállás-elmélet megjelenése és távközlési alkalmazásának kezdete a múlt század elejére tehető, amikor Erlang megmutatta, hogy a telefon hálózatok forgalma jól modellezhető Poisson folyamattal, és bevezette az első sorbanállásos rendszereket a hívásblokkolás, illetve -várakozás valószínűségének kiszámítására.

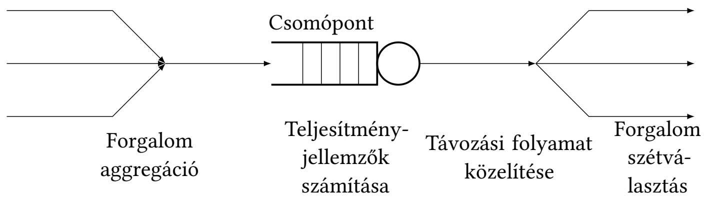
A csomagkapcsolt hálózatok megjelenése újabb lendületet adott a sorbanállásos rendszerekkel kapcsolatos kutatásoknak. Ezekben a hálózatokban a legfontosabb minőségi jellemzőket, többek között a csomagvesztési valószínűséget valamint a csomagok késleltetési idejét, azoknak a buffereknek a viselkedése határozza meg, melyek a hálózati eszközökben a csomagok – továbbítás előtti – átmeneti tárolására szolgálnak. Ennek megfelelően, teljesítmény- és megbízhatósági modellezés céljára alkalmas lehet egy csomagkapcsolt hálózatot bufferek hálózatának, sorbaállási hálózatnak tekinteni.

Sajnos matematikailag pontos algoritmusok csak bizonyos szigorú megkötéseknek eleget tevő sorbanállási hálózatokra állnak rendelkezésre: a forgalom csak Poisson folyamat lehet, a csomópontokban pedig a csomagkiszolgálási idők és a kiszolgálási diszciplína sem lehet tetszőleges. Ezek a megkötések túl szigorúnak bizonyultak a valós távközlési alkalmazásokhoz ([25]), hiszen a csomagok követési ideje a valóságban gyakran összefüggő, és a kiszolgálási idők sem exponenciális eloszlásúak. Sőt, egy modern hálózatban a forgalmat nem lehet egy homogén csomagfolyamként felfogni, a csomagokat a minőségi elvárásaik mentén osztályokba sorolják, és a hálózati eszközök ütemezői a különböző osztályba tartozó csomagokat különböző szolgáltatásban részesítik. A valós hálózatok ezen jellegzetességei lényegesen befolyásolják a minőségi jellemzőket, így ezeket mind az analitikus, mind a szimuláció alapú teljesítményelemzés során feltétlenül figyelembe kell venni.

Számos új modellezési megközelítést javasoltak a hálózati forgalom pontosabb leírására [27]. Az egyik kutatási irány a Poisson folyamat és az exponenciális eloszlás olyan kiterjesztéseit célozta meg, melyek továbbra is Markovi keretek között maradnak. Ennek eredményeképpen vezették be a fázis típusú (phase-type, PH) eloszlásokat és a Markovi érkezési folyamatokat (Markovian arrival process, MAP), valamint ez utóbbiak többosztályos változatát, a jelölt Markovi érkezési folyamatokat (marked Markovian arrival process, MMAP) [24]. Ezeknek a Markovi, de ennek ellenére rendkívül rugalmas forgalmi modelleknek a megjelenése nagy fejlődést hozott az alábbi három területen:

- Úgynevezett illesztő algoritmusok kifejlesztésében, melyek a lehető legpontosabb Markovi modellt állítják elő valós hálózati viselkedés jellemzésére,
- olyan sorok hatékony teljesítményanalízisében, melyek Markovi forgalmi modellel rendelkeznek,
- olyan sorbanállási hálózatok vizsgálatában, melyek forgalmát Markovi modellek írják le.

A tézisek eredményeinek fő motivációja a valós távközlési- és számítógép hálózatok hatékony teljesítményvizsgálata, azonban fontos megjegyezni, hogy az utóbbi év-



1. ábra. A csomópontok feldolgozásának lépése a forgalom alapú dekompozíció szerint

tizedekben a sorbanálláselmélet számos új területen is teret nyert, melyek a bemutatott eredmények további potenciális alkalmazási területei lehetnek. Ilyen új területek az egészségügyi rendszerek analízise, gyártási és logisztikai folyamatok modellezése, meghibásodási analízis, vagy akár a crowd-sourcing rendszerek és a közösségi hálók vizsgálata.

1.1 *Kutatási célkitűzések*

A *forgalom alapú dekompozíció* az egyik leggyakrabban alkalmazott megközelítés nyílt sorbanállási hálózatok vizsgálatára. Lényege, hogy a hálózat csomópontjait egyenként vesszük sorra, és egymástól elkülönítve dolgozzuk fel [21]. Minden egyes csomópont esetén négy lépést kell végrehajtani: az első lépés a különböző irányokból érkező bemenő forgalom aggregálása egy érkezési folyamattá, a második lépésben történik a csomóponttal kapcsolatos teljesítményjellemzők kiszámítása, a harmadik lépés a távozási folyamat jellemzése, míg az utolsó lépés a távozó forgalom szétbontása a forgalomirányításnak megfelelően (1. ábra). Egy csomópont akkor vehető sorra, ha minden csomópont feldolgozása, ahonnan bemenő forgalma származhat, már befejeződött. Kutatási tevékenységem a forgalom alapú dekompozíció lépéseinek továbbfejlesztéséhez kapcsolódik, magában foglalva

- a hálózati forgalom Markovi modellekkel való minél pontosabb reprezentálását,
- a hálózat csomópontjainak minél hatékonyabb teljesítményelemzését,
- a csomópontok távozási folyamatának pontos analízisét.

A forgalmi modellezés terén a fő kutatási célkitűzés a fázis típusú eloszlások és Markovi érkezési folyamatok határának, modellezési képességének tanulmányozása, valamint olyan illesztési eljárások kidolgozása, melyek segítségével ezek a Markovi modellek valós forgalmi adatokból előállíthatók.

A csomóponti analízis támogatásához a kapcsolódó (több osztályos) sorbanállási modellek teljesítményjellemzőinek minél hatékonyabb előállítására és a távozási folyamat minél pontosabb jellemzésére van szükség. Az általam kidolgozott algoritmusok megfelelő numerikus viselkedésére mindig nagy hangsúlyt fektetek, hogy jól skálázhatók (nagy méretű rendszerekre is alkalmazhatók), robusztusak és numerikusan stabilak legyenek.

A végső cél a nagy méretű, többosztályos forgalommal rendelkező nyílt sorbanállási hálózatok minél pontosabb analízise, melyekben a csomópontok kellően általánosak, nem feltétlenül exponenciális eloszlású kiszolgálási időkkel és különféle többosztályos csomagütemezővel. Meggyőződésünk, hogy egy ilyen analízis eszköz létrehozása számos gyakorlati, mérnöki probléma megoldásához járulna hozzá, nem feltétlenül kizárólag a távközlési és számítógép-hálózatok területén.

1.2 *A kutatás módszertana*

A disszertációban alkalmazott Markovi metodológia a múltban számos alkalommal bizonyította alkalmasságát mind a forgalmi modellezés, mint a sorbanállásos rendszerek teljesítményanalízise terén.

Számos modern, forgalmi modellezésre szolgáló sztochasztikus folyamat hátterében egy Markov lánc áll. A PH eloszlású valószínűségi változók megfeleltethetők egy tranziens Markov lánc nyelőbe jutási idejének, MAP-ok és MMAP-ok esetében pedig egy (pl. érkezési) esemény akkor következik be, ha ez a háttér Markov lánc bizonyos állapotátmenetek mentén vált állapotot. Ezek a Markovi forgalmi modellek számos előnyös tulajdonsággal rendelkeznek. A PH eloszlások *sűrűek*, ami azt jelenti, hogy tetszőleges pontossággal képesek közelíteni eloszlások egy meglehetősen széles körét. Zártsági tulajdonságaik szintén vonzóvá teszik őket: PH eloszlású véletlen változók összege, minimuma, és maximuma úgyszintén PH eloszlású. A MAP-ok és MMAP-ok pedig olyan pontfolyamatok, melyek korrelált érkezési időközök, és – több osztály esetén – az osztályok közötti kereszt-korreláció modellezésére is alkalmasak. A MAP-ok és MMAP-ok is sűrűek, valamint zártak a forgalom aggregáció és a (Bernoulli) forgalom szétválasztás műveletekre, ami ideálissá teszi őket egy sorbanállási hálózat forgalmának reprezentálására. A felsorolt tulajdonságokon felül mind a PH eloszlásokat, mind a MAP-okat és MMAP-okat viszonylag könnyű beépíteni mind analitikus, mind szimuláció alapú teljesítményelemzési eszközökbe. A disszertáció nagy mértékben támaszkodik ezekre a Markovi forgalmi modellekre, a tézisek pedig számos kapcsolódó új eredményt is bemutatnak.

Az 1980-as években, a nagyméretű, szabályos struktúrájú Markov láncok hatékony megoldására bevezetett *mátrix-analitikus algoritmusok* tették lehetővé számos, fázis-típusú eloszlásokra és Markovi érkezési folyamatokra alapozott sor numerikusan stabil egyensúlyi megoldását. A sorok teljesítményjellemzőinek kiszámítására három metodológia egyikét, azaz vagy a sorhossz folyamat alapú, vagy a munkahátralék folyamat alapú, vagy az életkor folyamat alapú megközelítést szokás követni. A három lehetőség közül a sorhossz folyamat alapú analízis a legismertebb, melyben egy Markov lánc követi a sorban álló igények számának alakulását véletlen, vagy bizonyos beágyazott időpontokban. A klasszikus sorbanállásos rendszerekkel foglalkozó szakirodalom (pl. [23]) erre a metodológiára épít. Az utóbbi évtizedekben azonban felismerték, hogy sok, összetett viselkedésű sor könnyebben elemezhető a munkahátralék folyamat alapján (amely a kiszolgáló hátralévő munkájának alakulását követi, [30]), vagy az életkor folyamat alapján (amely a rendszerben legrégebb óta tartózkodó igény életkorát követi, [29]). Mind a munkahátralék, mind az életkor folyamat folytonos folyamat, melyek ugrásokat is tartalmaznak. A sorbanálláselméletben az utóbbi évek egyik legnagyobb eredménye, hogy a mátrix-analitikus algoritmusokat ezeknek a folytonos folyamatoknak a megoldására is kiterjesztették (lásd [26] és [19]). A disszertáció sorokkal kapcsolatos részei és a vonatkozó tézisek nagy mértékben építenek a sorhossz, a munkahátralék és az életkor folyamatok mátrix-analitikus megoldására.

1.3 *A tézisek áttekintése*

A disszertáció három egymásra épülő részből áll, a tézisek ezek mentén három csoportra oszthatók.

Az első téziscsoport témái a Markovi forgalom modellezés legfontosabb eszközei, a fázis típusú eloszlások és a Markovi érkezési folyamatok. Az 1.1-es tézis megadja a három állapotú PH eloszlások kanonikus formáját, mely – többek között – lehetővé teszi hatékonyabb PH illesztési eljárások kifejlesztését is. Az 1.2-es tézis egy momentumillesztő eljárást mutat be, amely az erre a célra bevezetett rugalmas reprezentációnak köszönhetően automatikusan képes megnövelni az állapotok számát mindaddig, amíg a cél momentumok illesztése sikeres nem lesz. Az 1.3 tézis a MAP-ok és MMAP-ok néhány alapvető tulajdonságára világít rá, és eljárást ad az együttes momentumok illesztésére is, egy- és többosztályos esetben egyaránt. Ezt a tézist három heurisztikus eljárás egészíti ki, melyek feladata a momentumillesztés eredményének Markovi reprezentációba transzformálása. A téziscsoport eredményei lehetővé teszik a hálózati forgalom Markovi modellezését, melyet analitikus és szimulációs vizsgálatokban is fel lehet használni.

A második téziscsoport egy- és többosztályos, Markovi érkezési és kiszolgálási folyamattal rendelkező sorok analízisével kapcsolatos eredményeket foglal magában. A 2.4-es tézis egy új eljárást fogalmaz meg a preemptív és nem-preemptív prioritásos sorok teljes körű teljesítményanalízisére. Az eljárás a munkahátralék folyamat elemzésén alapul, és a rendszerben lévő igények számára valamint a rendszerben töltött időre adja meg a stacionárius eloszlást, illetve a momentumokat. A 2.1-es, 2.2-es és 2.3-as tézis témája az egyosztályos MAP/MAP/1, a többosztályos sorrendi kiszolgálással rendelkező MMAP[K]/PH[K]/1-FCFS, valamint a többosztályos prioritásos kiszolgálással rendelkező MMAP[K]/PH[K]/1-Prio sor távozási folyamatának analízise. Bár mindhárom tézis távozási folyamatokkal kapcsolatos, a három sor esetében mások a kihívások és alapvetően különböznek a megoldások is.

A harmadik téziscsoport, amely csupán egyetlen tézisből áll, a sorbanállási hálózatok analíziséről szól. A 3.1-es tézisben javasolt analízis módszer az első két téziscsoport eredményeinek integrálásával közelítést ad olyan sorbanállási hálózatok teljesítményjellemzőire, melyek forgalma az első téziscsoportban vizsgált Markovi érkezési folyamattal adott, csomópontjait pedig a második téziscsoportban tárgyalt sorok alkotják.

Motiváció

A valós hálózati forgalmat minél pontosabban reprezentáló Markovi forgalmi modell előállításához hatékony *illesztési* eljárásokra van szükség. Az illesztési eljárások valós mérési adatokból, vagy bizonyos statisztikai jellemzők alapján állítanak elő PH eloszlásokat vagy MAP, ill. MMAP érkezési folyamatokat.

A szakirodalomban nagy számú PH eloszlás illesztő eljárás lelhető fel, melyek közül a momentumillesztő algoritmusok különösen hasznosak, hiszen az illesztendő momentumok valós mérésekből, illetve analitikus modellekből egyaránt egyszerűen kinyerhetők. Sajnos az ismert momentumillesztő eljárások túlnyomó többsége nem elég rugalmas, gyakran előfordul, hogy negatív sűrűségfüggvénnyel rendelkező érvénytelen PH eloszlást eredményeznek. Csupán két olyan eljárásról tudunk ([22] és [16]), melyek mentesek ettől a problémától.

Az összefüggő érkezési időket modellezni képes Markovi érkezési folyamatokra azonban közel sem áll rendelkezésre olyan sok és olyan kiforrott illesztési eljárás, mint PH eloszlásokra. Sok, az illesztő eljárások szempontjából alapvető fontosságú kérdés még mindig nyitott, többek között az is, hogy hogyan lehet a legalkalmasabb módon mérni két összefüggő érkezési folyamat távolságát. Vannak illesztő eljárások, melyek a likelihood maximalizálásán alapulnak, míg mások a peremeloszlások és az autokorrelációs függvények távolságát igyekeznek minimalizálni.

Ebben a témában végzett kutatási tevékenységünk célja a PH eloszlások és a MAP-ok, illetve MMAP-ok korlátainak jobb megértése, illetve ezek ismeretében az eddigieknél hatékonyabb illesztő eljárások kifejlesztése.

1.1-es tézis

Bebizonyítottam, hogy megfelelő hasonlósági transzformáció segítségével minden három állapotú PH eloszlás az alábbi kanonikus formák egyikébe transzformálható:

$$\boldsymbol{\gamma}^{(1)} = \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_3 \end{bmatrix}, \qquad \boldsymbol{\gamma}^{(2)} = \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_3 \end{bmatrix}, \qquad \boldsymbol{\gamma}^{(3)} = \begin{bmatrix} \gamma_1 & 0 & \gamma_3 \end{bmatrix},$$

$$\boldsymbol{G}^{(1)} = \begin{bmatrix} -x_1 & 0 & 0 \\ x_2 & -x_2 & 0 \\ 0 & x_3 & -x_3 \end{bmatrix}, \boldsymbol{G}^{(2)} = \begin{bmatrix} -x_1 & 0 & x_{13} \\ x_1 & -x_1 & 0 \\ 0 & x_3 & -x_3 \end{bmatrix}, \boldsymbol{G}^{(3)} = \begin{bmatrix} -x_1 & 0 & x_{13} \\ x_2 & -x_2 & 0 \\ 0 & x_3 & -x_3 \end{bmatrix},$$

ahol $\boldsymbol{G}$ a tranziens Markov lánc generátor mátrixa, a $\boldsymbol{\gamma}$ vektor pedig a kezdőállapot eloszlása.

A kapcsolódó publikációk: [1] és [2].

A kanonikus forma egy olyan reprezentáció, amely egy PH eloszlást – minimális paraméterszám segítségével – egyértelmű módon meghatároz.

Az aciklikus PH eloszlások kanonikus formája régóta ismert ([18]), az általános (nem aciklikus struktúrákat is tartalmazó) PH osztályra azonban jóval nehezebb kanonikus formát adni.

Két állapot esetén az általános PH(2) osztályra is van ismert kanonikus alak. Két állapot esetén azonban könnyebbséget jelent, hogy a generátormátrix mindkét sajátértéke szükségszerűen valós (hiszen a domináns sajátérték mindig valós). A három állapotú általános PH eloszlások generátormátrixa viszont tartalmazhat komplex sajátértékeket, ami a fent megadott, komplikáltabb kanonikus formát eredményezi.

Jól ismert eredmény (amely a disszertáció 2.1.1-es fejezetében közölt, spektrális alakban felírt sűrűségfüggvényből is látszik), hogy a PH(3) eloszlások 5 paraméterrel rendelkeznek. A fent megadott három kanonikus forma úgyszintén 5 paramétert tartalmaz (mivel $\gamma_1 + \gamma_2 + \gamma_3 = 1$), így ezek minimális paraméterszámú reprezentációk.

A tézishez kacsolódik a disszertáció 1. algoritmusa, mely egy tetszőleges PH(3) eloszlást kanonikus formára alakít, akkor is, ha a kiinduló reprezentáció nem Markovi. Ha az algoritmus nem ad vissza Markovi reprezentációt, akkor egészen biztos, hogy a kiinduló reprezentáció nem PH(3) eloszlás volt. Ha az algoritmus PH(3) eloszlást kap, akkor minden esetben a fenti három kanonikus forma egyikét fogja kimenetként visszaadni. Például, az alábbi $(\underline{\sigma}, \boldsymbol{S})$ nem Markovi reprezentációt,

$$\underline{\sigma} = \begin{bmatrix} 0,1 & 0,2 & 0,7 \end{bmatrix}, \quad \boldsymbol{S} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -3 & 0,2 \\ 0 & -0,2 & -3 \end{bmatrix},$$

mely negatív átmeneti intenzitást tartalmaz, az algoritmus a $(\underline{\gamma}, \boldsymbol{G})$ paraméterekkel megadott harmadik kanonikus formába transzformálja,

$$\underline{\gamma} = \begin{bmatrix} 0,9683 & 0 & 0,0317 \end{bmatrix}, \quad \boldsymbol{G} = \begin{bmatrix} -3,0221 & 0 & 0,027123 \\ 2,9572 & -2,9572 & 0 \\ 0 & 1,0207 & -1,0207 \end{bmatrix},$$

ami már egy érvényes Markovi reprezentáció.

A kanonikus formák segítségével a PH illesztő eljárások hatékonyabbá tehetők. Ha az illesztő eljárás nem használ kanonikus formát, hanem a teljes $(\underline{\gamma}, \boldsymbol{G})$ (tudvalevőleg redundáns) reprezentáción optimalizál, akkor könnyen körbe-körbe járhat alig különböző eloszlások nagyon különböző reprezentációi között. A kanonikus formák nem redundánsak és egyértelműek, tehát „egyenesebb” az út az optimum felé. A különbség számszerűsítésére egy példát állítottunk össze, melyhez egy egyszerű, line search alapú eloszlás illesztő algoritmust implementáltunk, mely a relatív entrópiát ([17]) használja távolságfüggvényként. A numerikus vizsgálatokban a következő négy eloszlást illesztettünk: W1 (Weibull), U1 (egyenletes), ME (mátrix-exponenciális) és BC (egy valós hálózati mérésekből származó empirikus eloszlás). Mivel a PH(3) eloszlások három kanonikus formával rendelkeznek, az illesztést mindhárom struktúrával megismételtük. Az illesztő eljárás az optimalizálást 100 véletlen kezdő pontból indítva végezte el, és a legjobb eredményt választotta ki. Az illesztés nagyjából

	W1	U1	ME	BC
Teljes PH(3):	150,98	110,4	117,49	146,6
f1) forma:	52,04	43,34	40,56	119,62
f2) forma:	64,27	55,22	52,53	111,14
f3) forma:	108,49	71,33	71,84	122,93

1. táblázat. A PH eloszlás illesztéshez szükséges iterációk átlagos száma

ugyanazt az eredményt adta a teljes PH(3) reprezentációval és a kanonikus reprezentációkkal is. Az 1. táblázat azonban egyértelműen rámutat a kanonikus formák alkalmazásának előnyére: az eredmény eléréséhez kevesebb iteráció szükséges, és a végeredmény kevésbé függ a véletlen kiindulási ponttól.

1.2-es tézis

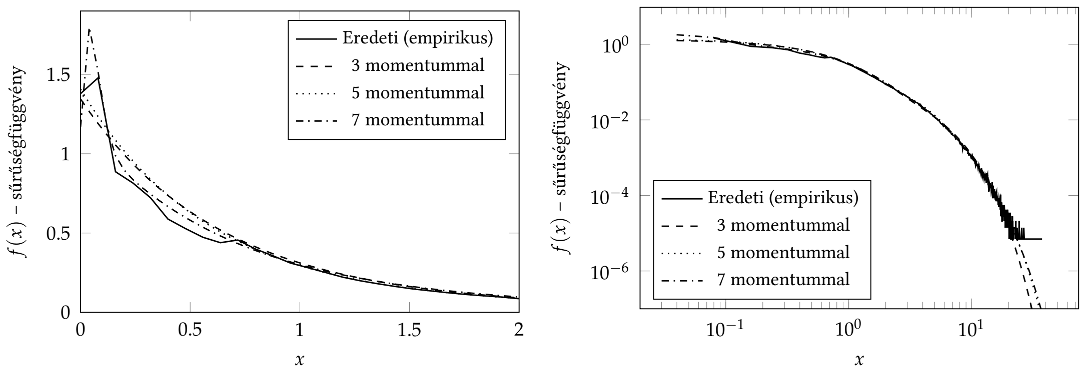
Bevezettem az általánosított hiper-Erlangnak elnevezett speciális PH struktúrát, és javasoltam hozzá egy rugalmas momentumillesztő eljárást, ami a reprezentáció méretét az illesztendő momentumoknak megfelelően automatikusan állítja be.

A tézis eredményeit a [3]-as publikáció tartalmazza.

A momentumillesztő eljárások túlnyomó része annyi momentumot illeszt, ahány szabad paraméterrel az ehhez használt PH eloszlás rendelkezik. Például, 5 momentumot tipikusan PH(3) eloszlással illesztenek, hiszen annak éppen 5 szabad paramétere van, ahogy erre az 1.1 tézis tárgyalásakor rámutattunk. Sajnos a PH(3) eloszlások által megvalósítható momentumok halmaza nagyon szűk, az illesztés sok esetben érvénytelen (negatív sűrűségfüggvénnyel rendelkező) PH eloszlást eredményez.

Ezekben a helyzetekben a megoldást az úgynevezett rugalmas PH struktúrák alkalmazása jelenti, melyek az illesztendő momentumok számánál több szabad paraméterrel rendelkeznek. A paraméterek egy része a momentumok illeszkedését biztosítja, az érvényes PH eloszlás előállításához pedig a fennmaradó paraméterek adják a mozgásteret. A szakirodalomban csupán két ehhez hasonló, rugalmas illesztő eljárás lelhető fel. Az egyik azonos fokszámú Erlang eloszlásokból építkező hiper-Erlang struktúrával, a másik pedig egy exponenciális és egy Erlang keverékével illeszt momentumokat. A rugalmasságot mindkét esetben az Erlang komponens(ek) fokszáma biztosítja. Mindkét eljárásnak megvan a hátránya: előbbi túlságosan nagy méretű PH eloszlást, és furcsa formájú sűrűségfüggvényt eredményezhet, az utóbbi pedig csak 3 momentum illesztésére alkalmas.

Az általánosított hiper-Erlang eloszlásokat (generalized hyper-Erlang distributions, GHErD) azért vezettük be, hogy segítségével a korábbiaknál jobb tulajdonsá-



2. ábra. A momentumillesztés eredményeinek összehasonlítása

gokkal rendelkező momentumillesztő eljárást adhassunk rá. A GHErD sűrűségfüggvényt

$$f(x) = \sum_{i=1}^{K} \sigma_i \frac{(\lambda_i x)^{r_i-1}}{(r_i-1)!} \lambda_i e^{-\lambda_i x}$$

definiálja, ahol a λ_i és a σ_i paraméterek komplexek is lehetnek. Ezek az eloszlások viszonylag egyszerű momentum formulával rendelkeznek, melynek köszönhetően az illesztési feladat egy polinomiális egyenletrendszer megoldására vezethető vissza. Ilyen egyenletrendszerek *összes* megoldásának előállítására pedig léteznek numerikus algoritmusok. Annak ellenőrzésére, hogy az egyenletrendszer egy adott megoldása érvényes PH eloszlást ad-e, az eredményt monociklikus reprezentációba transzformáljuk a disszertáció 2.1.4-es fejezetében leírt eljárás segítségével. A struktúra rugalmasságát az $r_i, i = 1, \ldots, K$ paraméterek adják. Rögzített r_i paraméterek mellett a polinomiális egyenletrendszernek több jó megoldása is lehet, és az is lehetséges, hogy nincs egy megoldás sem. Ez utóbbi esetben az r_i paramétereket meg kell növelni. Belátható, hogy elegendően nagy r_i paraméterek mellett minden (pozitív eloszláshoz tartozó) momentum sorozat sikeresen illeszthető GHErD eloszlással.

A következő numerikus példában az LBL-TCP-3 mérési adatsor momentumait illesztjük a bemutatott eljárás segítségével. Amennyiben az algoritmus több, az adatsor momentumait pontosan illesztő PH eloszlást is talál, azok közül azt választja ki, amely a mérési adatsor empirikus sűrűségfüggvényére vonatkozó relatív entrópiát minimalizálja. 3 momentum illesztése esetén a hiper-exponenciális eloszlás bizonyult a legjobbnak ($r_1 = 1, r_2 = 1$), mely mellett a relatív entrópia $0,3024$. 5 momentum illesztésével a relatív entrópia jelentősen kisebb lesz ($0,0953$), és a legjobb eredményt a $r_1 = 1, r_2 = 1, r_3 = 2$ paraméterek adják. A 2. ábra jobb oldalán látható, hogy az 5 momentumos illesztés sokkal jobban közelíti a log-log skálán ábrázolt farok viselkedést, mint a 3 momentumos. Az illesztett momentumok számának 7-re növelésével az eredmények tovább javulnak ($0,0727$, a paraméterek: $r_1 = 1, r_2 = 1, r_3 = 3, r_4 = 3$), és a baloldali ábrán tetten érhető, hogy már a sűrűségfüggvény törzse is jobban hasonlít az empirikus sűrűségfüggvény törzsére.

Ez a momentumillesztő eljárás a következő tézisben bemutatott MMAP illesztő eljárásoknak egy nagyon fontos összetevője.

1.3-as tézis

Rámutattam, hogy minden N-ed rendű MAP egyértelműen megadható N^2 paraméterrel. Bevezettem egy momentum illesztő eljárást, mely $2N-1$ momentum és $(N-1)^2$ egylépéses együttes momentum alapján létrehoz egy N-edrendű MAP-ot. Az eredményeket MMAP-ra is kiterjesztettem: megmutattam, hogy a C osztályos, N-ed rendű MMAP-okat $C \cdot N^2$ paraméter egyértelműen meghatározza, és megadtam a momentum illesztő eljárás többosztályos változatát is.

Az egyosztályos eredmények [4]-ben, a többosztályosok [5]-ben jelentek meg, de a tézishez szorosan kapcsolódnak a [6], a [7], a [8] és a [9] publikációk is.

Mielőtt a tézisben megfogalmazott eredmények ismertek lettek volna, a MAP-ok által generált érkezési időközök összefüggőségének jellemzésére szinte kizárólag az autokorrelációs függvényt használták. Rámutattunk, hogy ez nem elegendő, mivel végtelen sok, esetenként egymástól nagyon különböző forgalmi viselkedést leíró MAP létezhet azonos autokorrelációs függvénnyel. Az autokorrelációs függvény helyett az egylépéses együttes momentumok (két egymást követő érkezési időköz együttes momentumai) viszont már egyértelműen leírják a MAP-ok viselkedését.

A momentumillesztő eljárást egy példán keresztül mutatjuk be. Az m_k-val jelölt momentumokat és az $\eta_{i,j}$-vel jelölt egylépéses együttes momentumokat az LBL mérési adatsorból nyertük ki (ez az adatsor két órányi TCP forgalom csomagérkezési időközeit rögzíti), és $m_1 = 1$-re normalizáltuk:

$$\begin{bmatrix} m_1 & m_2 & m_3 & m_4 & m_5 \end{bmatrix} = \begin{bmatrix} 1 & 2,942 & 16,84 & 150,73 & 1876,8 \end{bmatrix},$$
$$\boldsymbol{N} = \begin{bmatrix} 1 & m_1 & m_2 \\ m_1 & \eta_{1,1} & \eta_{1,2} \\ m_2 & \eta_{2,1} & \eta_{2,2} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 2,942 \\ 1 & 1,3013 & 4,5056 \\ 2,942 & 4,5 & 17,416 \end{bmatrix}.$$

Az első lépés az érkezési időközök eloszlásának illesztése (egyelőre korreláció nélkül), amihez a [32]-ban megjelent momentumillesztő eljárással készítünk egy 3 állapotú PH eloszlást $m_k, k = 1, \dots, 5$ alapján. A PH eloszlás tranziens generátora lesz a $\boldsymbol{D_0}$ mátrix, és a kezdeti állapot valószínűség vektor, $\underline{\alpha}$, pedig a MAP érkezési időkre beágyazott fáziseloszlása:

$$\underline{\alpha} = \begin{bmatrix} 0,33333 & 0,33333 & 0,33333 \end{bmatrix},$$
$$\boldsymbol{D_0} = \begin{bmatrix} -1,7356 & 0,34074 & -0,95214 \\ -0,18575 & -0,63031 & -0,042169 \\ -0,48092 & -0,036353 & -0,6245 \end{bmatrix}.$$

A következő lépés a $\boldsymbol{D_1}$ mátrix előállítása. Az együttes momentumok illesztése 4 lineáris egyenletet ad a $\boldsymbol{D_1}$ mátrix elemeire:

$$\eta_{i,j} = i!j!\underline{\alpha}(-\boldsymbol{D_0})^{-i-1}\boldsymbol{D_1}(-\boldsymbol{D_0})^{-j}\mathbb{1}, \quad i,j = \{1,2\}.$$

Továbbá, a háttér Markov lánc generátorában a sorösszegeknek nullát kell adnia, tehát $\boldsymbol{D}\mathbb{1} = (\boldsymbol{D_0} + \boldsymbol{D_1})\mathbb{1} = \underline{0}$ további 3 lineáris egyenletet jelent:

$$\boldsymbol{D_1}\mathbb{1} = -\boldsymbol{D_0}\mathbb{1}.$$

Végül, 2 újabb lineáris egyenlet írható fel az érkezésekre beágyazott fáziseloszlásra:

$$\underline{\alpha}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1} = \underline{\alpha}.$$

Összességében 9 lineáris egyenlet áll rendelkezésre a $\boldsymbol{D_1}$ mátrix 9 elemének meghatározására, melynek megoldása

$$\boldsymbol{D_1} = \begin{bmatrix} 1,4612 & -0,24037 & 1,1263 \\ 0,1905 & 0,53436 & 0,13337 \\ 0,47616 & 0,1323 & 0,5333 \end{bmatrix}.$$

A kapott $\boldsymbol{D_0}$ és $\boldsymbol{D_1}$ mátrixokkal a momentumok és együttes momentumok illesztése tökéletes, viszont ez a mátrix pár sajnos nem Markovi reprezentáció, mivel negatív állapotátmeneti rátákat tartalmaz. Megfelelő hasonlósági transzformációval azonban sok esetben lehetséges a mátrixokat Markovi reprezentációba transzformálni. A disszertáció 3.2.4 fejezete egy egyszerű heurisztikus algoritmust mutat be, mely elemi transzformációs lépések sorozatával igyekszik a Markovi reprezentációt megtalálni. Ez az algoritmus ez esetben sikerrel járt, és az alábbi mátrix párt eredményezte:

$$\boldsymbol{G_0} = \begin{bmatrix} -2{,}0161 & 0{,}091134 & 0{,}029499 \\ 0{,}014955 & -0{,}63564 & 0{,}0079826 \\ 0{,}08378 & 0{,}062089 & -0{,}33873 \end{bmatrix}, \boldsymbol{G_1} = \begin{bmatrix} 1{,}8439 & 0{,}04524 & 0{,}0063419 \\ 0{,}063835 & 0{,}54292 & 0{,}0059518 \\ 0{,}047433 & 0{,}003395 & 0{,}14203 \end{bmatrix},$$

mely pontosan ugyanazt az érkezési folyamatot írja le, mint $\boldsymbol{D_0}$ és $\boldsymbol{D_1}$ (így a momentumok illesztése tökéletes), de ezt Markovi reprezentációval éri el.

Ez a transzformációs eljárás azonban esetenként nem jár sikerrel. Ha a momentumillesztő eljárás olyan mátrixokat állít elő, melyek érvénytelen sztochasztikus folyamatot definiálnak (negatív az együttes sűrűségfüggvény), akkor azon semmilyen hasonlósági transzformáció nem tud segíteni. Ezekben az esetekben a cél az érvénytelen (de a momentumokat pontosan illesztő) MAP-hoz lehető „legközelebb” álló érvényes MAP előállítása. A disszertáció 3.3-dik fejezete két ilyen algoritmust is javasol, melyek mindegyike a [6]-ban bevezetett, úgynevezett kétlépéses illesztő eljárások családjába tartozik. Az első lépés az érkezési időközök eloszlásának illesztése (amire alkalmasak az 1.2-es tézis eredményei), a második lépés pedig a korreláció illesztése. A disszertáció 3.3.3 fejezetében ismertetett eljárás az egylépéses együttes momentumok L_2 távolságát, a 3.3.4 fejezet megoldása pedig az együttes sűrűségfüggvények L_2 távolságát minimalizálja a második lépésben.

A tézis eredményeinek segítségével MAP alapú forgalmi modelleket lehet előállítani együttes momentumokból. Ezek az együttes momentumok származhatnak mérésekből származó adatsorokból, vagy, ahogy azt később a 3.1-es tézis megmutatja, sorok távozási folyamatának analíziséből.

Motiváció

A sorbanállási hálózatok forgalmi dekompozíció alapú megoldása során a csomópontok analízisének két szerepe van:

- a rendszerben tartózkodó igények számával és a rendszerben töltött idővel kapcsolatos teljesítményjellemzők meghatározása,
- a csomópont távozási folyamatának (mely a soron következő csomópontok bemenő folyamata lesz) jellemzése.

A MAP bemenő folyamattal és PH eloszlású kiszolgálási idővel rendelkező egyosztályos sorok teljesítményanalízise mára sztenderdnek tekinthető, régóta ismert, kiforrott eszközök állnak hozzá rendelkezésre [24]. A többosztályos sorokról ugyanez nem mondható el. Még az olyan egyszerű sort, mint a többosztályos, sorrendi kiszolgálással rendelkező MMAP[K]/PH[K]/1-FCFS sort sem lehet a hagyományos megközelítéssel megoldani, mivel a sor hosszát reprezentáló Markov lánc szerkezetére jelenleg nincs ismert hatékony stacioner megoldás. Erre a konkrét sorra az életkor folyamatra alapozott megközelítés bizonyult célravezetőnek [20]. Hasonlóan, az MMAP[K]/G/1 prioritásos sor jellemzőit is sokkal könnyebb kiszámítani, ha a sorhossz folyamat helyett a munkahátralék folyamatra építjük a megoldást (lásd [30] és [31]). A prioritásos sorok analízise azonban még ezzel együtt is rendkívül számításigényes. Mivel prioritásos ütemezővel rendelkező csomópontokat is tartalmazó sorbanállási hálózatokat szeretnénk vizsgálni, új, az eddigieknél lényegesen hatékonyabb számítási eljárások kidolgozására van szükség.

Az egyosztályos MAP/MAP/1 sorok távozási folyamatának MAP-pal történő közelítésére számos eljárás található a szakirodalomban. Az 1.3-as tézisben megfogalmazott MAP karakterizációs eredmények azonban felvetették egy újabb közelítő eljárás lehetőségét, mely a távozó forgalom MAP modelljét a távozási folyamat egylépéses együttes momentumai alapján, momentumillesztéssel állítja elő. Ehhez ki kell számítani a távozási folyamat együttes momentumait, mivel erre az irodalomban még nincs vonatkozó eredmény.

Az egyosztályos esettel ellentétben a többosztályos sorok távozási folyamatának közelítésére nem áll rendelkezésre ismert eredmény. Azonban az 1.3-as tézis momentumillesztő eljárásaira alapozott megközelítés több forgalmi osztályra (MMAP-ra) is megvalósítható, motivációt szolgáltatva a többosztályos sorok távozási folyamatának vizsgálatára.

2.1-es tézis

Megadtam a MAP/MAP/1 sor távozási folyamatának egylépéses együttes momentumait.

A vonatkozó eredmények [13]-ban jelentek meg.

Az alábbiakban tömören összefoglaljuk a momentumok és együttes momentumok számításának mikéntjét.

A fő észrevétel, hogy egy távozási időköz kétféle lehet: vagy egy kiszolgálási idővel egyenlő, vagy pedig egy hátralévő érkezési idő és egy kiszolgálási idő összegével. Az egylépéses együttes momentumok kiszámításához két egymást követő távozási időköz együttes viselkedését kell megvizsgálni, melyhez három esetet kell számításba venni:

- 0: a távozó igény üres sort hagyott maga után,
- 1: a távozó igény pontosan egy igényt hagyott maga után a rendszerben,
- 2+: a távozó igény legalább két igényt hagyott maga után. Ebben az esetben a következő két távozási időköz két kiszolgálási idő lesz, mivel a rendszer közben nem ürülhet ki.

Ezeknek az eseteknek a stacioner valószínűségeit (melyeket az $\underline{x}_0, \underline{x}_1$ és $\underline{x}_{2+}$ vektorok jelölik) a sorhossz folyamatot modellező kvázi születési-halálozási folyamat (quasi birth-death process, QBD) stacioner eloszlásából lehet származtatni. Ezt követően egy Markov láncot definiálunk, amely a sorhossz viselkedését írja le a második szintig bezárólag. A Markov lánc átmeneteit ezután két csoportra bontjuk, az egyik csoport azokból az átmenetekből áll, melyek nem járnak igény távozással (ezeket a rátákat a $\boldsymbol{H_0}$ mátrix tartalmazza), míg a másik csoport azokból, melyek igény távozással járnak ($\boldsymbol{H_1}$ mátrix). A mátrixok alakja a következő:

$$\boldsymbol{H_0} = \begin{bmatrix} \boldsymbol{L_0} & \boldsymbol{F} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{L} & \boldsymbol{F} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{L}+\boldsymbol{F} \end{bmatrix}, \quad \boldsymbol{H_1} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{B} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B} & \boldsymbol{0} \end{bmatrix},$$

ahol a $\boldsymbol{B}, \boldsymbol{L}, \boldsymbol{F}$ mátrixok a QBD szintvisszalépéshez, szinten maradáshoz, és szintelőrelépéshez tartozó blokkjai. A Markov lánc három állapotcsoportja a fent leírt 0, 1 és 2+ eseteknek felelnek meg. A Markov láncot a $\begin{bmatrix} \underline{x}_0 & \underline{x}_1 & \underline{x}_{2+} \end{bmatrix}$ eloszlásból indítva a következő két távozási időköz ($\mathcal{H}_0$ és $\mathcal{H}_1$) együttes momentumai az alábbi módon számíthatók ki:

$$E(\mathcal{H}_0^i \mathcal{H}_1^j) = i!j! \cdot \begin{bmatrix} \underline{x}_0 & \underline{x}_1 & \underline{x}_{2+} \end{bmatrix} \cdot (-\boldsymbol{H_0})^{-i-1} \boldsymbol{H_1} (-\boldsymbol{H_0})^{-j} \mathbb{1}.$$

A tézis eredményei adják az együttes momentum alapú sorbanállási hálózat analízis eljárás alapjait. A 3.1 tézisben javasolt algoritmus az imént leírt módon kiszámított együttes momentumokból illeszt MAP-ot a távozási folyamat közelítése érdekében.

2.2-es tézis

Megadtam a kétosztályos MAP/MAP/1 preemptív prioritásos sor távozási folyamatának együttes momentumait.

A tézis eredményeit az [5]-ös és a [10]-es publikációk tartalmazzák.

A megoldás menete hasonlít a 2.1-es tézisnél látottakhoz. A 2.1-es és a 2.2-es tézis szétválasztásának oka, hogy a 2.1-es tézis egy újszerű megközelítés első megtestesülése, míg a 2.2-es tézis ugyanannak a megközelítésnek egy lényegesen összetettebb rendszerre való alkalmazása, amely a prioritásos sorok stacioner analízisében is új eredményeket igényelt.

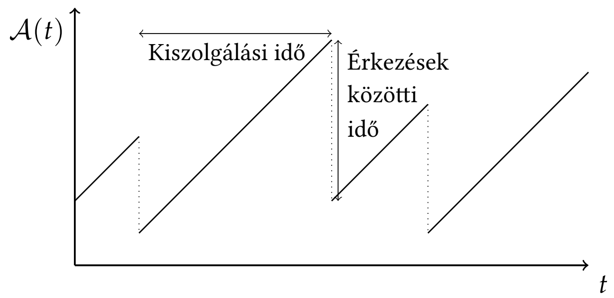
A kétosztályos prioritásos esetben három helyett hat esetet kell megkülönböztetni a (több osztályos) együttes momentumok kiszámításához:

- $0,0$: a távozó igény üres rendszert hagyott hátra,
- $1,0$: a távozó egy magas, és nulla alacsony prioritású igényt hagyott a rendszerben,
- $1,1+$: a távozó egy magas, és legalább egy alacsony prioritású igényt hagyott hátra,
- $2+,0+$: a távozás legalább 2 magas prioritású igényt hagyott hátra,
- $0,1$: a távozás után nulla magas, és egy alacsony prioritású igény maradt a rendszerben,
- $0,2+$: a távozás után nulla magas, és legalább két alacsony prioritású igény maradt a sorban.

Az igazán nagy kihívást a hat eset stacioner eloszlásának kiszámítása jelenti. Míg egy osztály esetén a sorhosszt egy QBD-vel lehetett modellezni (a 2.1-es tézisben), a kétosztályos esetben két sor van. A két sor együttes viselkedését leíró Markov lánc is QBD (ahogy azt [14]-ben megmutatták), de a fázisok száma végtelen. A [14]-ben publikált stacioner analízis eljárás szép megoldás, de sajnos több numerikusan szempontból előnytelen lépést is tartalmaz: végtelen elemszámú mátrix sorozatok kiszámítását és végtelen összegzéseket igényel, melyeket végessé téve az eredmény pontossága csorbul.

A Markov lánc speciális szerkezetének kihasználásával a tézisben bemutatott eljárás több ponton is alapvető mértékben javít az irodalomban leírtakhoz képest. Bebizonyítottuk, hogy a prioritásos rendszer analízisében kulcsfontosságú szerepet játszó két mátrix (melyet a disszertáció 7.4-es fejezete $\boldsymbol{Z}$-vel és $\boldsymbol{G}_{H_0}$-al jelöl), kommutál. Ez a felismerés megnyitotta az utat egy korábbiaknál lényegesen gyorsabb és pontosabb stacioner analízis eljárás kifejlesztése előtt.

A fenti hat eset stacioner valószínűségének kiszámítása után a távozási folyamat együttes momentumainak származtatása a 2.1-es tézisnél látottakhoz hasonlóan történik. Az együttes momentumok alapján előállítható a távozási folyamat MMAP modellje, mely lehetővé teszi a prioritásos sorokat tartalmazó sorbanállási hálózatok megoldását.



3. ábra. Az életkor folyamat időbeni alakulása

2.3-as tézis

Eljárást adtam a többosztályos MMAP[K]/PH[K]/1-FCFS sor távozási folyamatának teljes körű analízisére. Az eljárás új megközelítést követ: a sorhossz folyamat helyett az életkor folyamaton alapul.

Az eredményeket [11] tartalmazza.

A sorrendi kiszolgálással rendelkező FCFS sor analízise első pillantásra könnyebbnek tűnik, mint a prioritásos soré, de valójában számos teljesítményjellemző kiszámítása (pl. a rendszerben lévő igények száma) lényegesen nehezebb annál. Ennek az az oka, hogy a rendszer állapotának jellemzéséhez nem elegendő a sor hosszának és a fázisnak a nyomon követése, azt is tudni kell, hogy a sorban milyen típusú igények állnak, és milyen sorrendben. Ennélfogva, a sorhossz folyamatra alapozott analízis helyett, melyet a 2.1-es és 2.2-es tézisek követtek az egysoros MAP/MAP/1 és a prioritásos esetben, új megközelítésre van szükség.

A többosztályos FCFS sor esetén az *életkor folyamat* segítségével számítjuk ki a távozási folyamat jellemzőit: az egymástól n távolságra lévő távozási időközök együttes eloszlásának Laplace-Stieltjes transzformáltját, és momentumait. Az életkor folyamat, $\{\mathcal{A}(t), t \geq 0\}$, a kiszolgálóban lévő igény rendszerben töltött idejét követi (3. ábra). A kiszolgálási időszakokban 1-es meredekséggel nő (hiszen a kiszolgálóban lévő igény egyre öregebb lesz), és a távozási időpontokban van egy ugrása az x tengely irányába, az ugrás mértéke pedig egy érkezési időközzel egyenlő (hiszen a következő igény, mely a kiszolgálóba került, ennyivel fiatalabb, mint az előző).

A tézis arra az észrevételre alapul, hogy a távozási időközök vizsgálatakor elkülöníthető esetek eloszlása (a $\underline{x}_0, \underline{x}_1$ és $\underline{x}_{2+}$ a MAP/MAP/1 sor esetében, ill. a 6 eset valószínűsége a prioritásos esetben) az életkor folyamatból is levezethető, a sorhossz eloszlás ismerete nélkül. Az eredmények általánosak: nincs megkötés az igénytípusok számára, és az együttes momentumokat tetszőleges távolságra lévő távozási időközökre is sikerült kiszámolni, nem csak a szomszédosakra.

Demonstrációs célra vegyünk egy példát, melyben az igényeket generáló MMAP mátrixai

$$\boldsymbol{D_0} = \begin{bmatrix} -2 & 1 \\ 0 & -5 \end{bmatrix}, \ \boldsymbol{D_1} = \begin{bmatrix} 0 & 1 \\ 0,1 & 0 \end{bmatrix}, \ \boldsymbol{D_2} = \begin{bmatrix} 0 & 0 \\ 1,9 & 3 \end{bmatrix},$$

és a kiszolgálási időket az alábbi paraméterekkel rendelkező PH eloszlás definiálja:

$$\underline{\sigma}_1 = \begin{bmatrix} 0,8 & 0,2 \end{bmatrix}, \quad \boldsymbol{S_1} = \begin{bmatrix} -2 & 1,5 \\ 0 & -1 \end{bmatrix}, \quad \underline{\sigma}_2 = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad \boldsymbol{S_2} = \begin{bmatrix} -25 & 5 \\ 0 & -25 \end{bmatrix}.$$

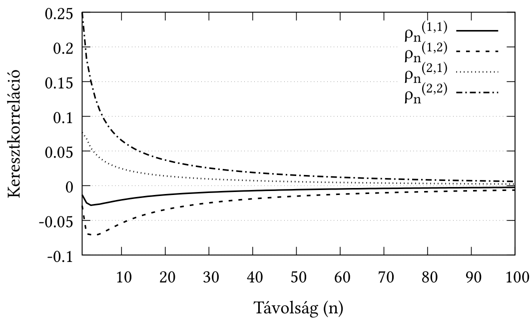
Ezekkel a paraméterekkel a rendszer kihasználtsága $0,7776$.

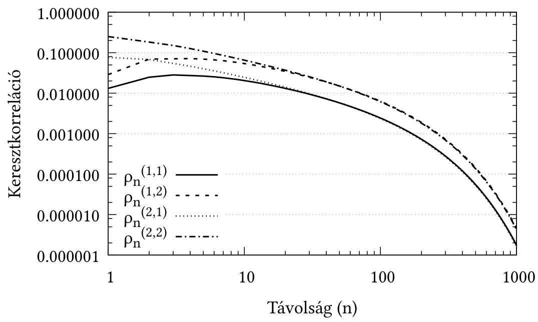
A 2. táblázatban látható az érkezési időközök (kereszt-) korrelációja. Mivel 2-es típusú igény csak a második fázisban érkezhet, ezért a 2-es típusú érkezési időközök független, azonos eloszlásúak. A táblázat így csak az 1-es igénytípus érkezési időközeinek ($\tilde{\rho}_n^{(1,1)}$), illetve az 1-es és 2-es érkezési időközök korrelációit ($\tilde{\rho}_n^{(1,2)}$) tartalmazza, a távozási időközök távolságának (n) függvényében.

n	$\tilde{\rho}_n^{(1,1)}$	$\tilde{\rho}_n^{(1,2)}$
1	$-1,7864 \times 10^{-2}$	$-2,9264 \times 10^{-2}$
2	$1,1135 \times 10^{-3}$	$6,9050 \times 10^{-3}$
3	$-2,5826 \times 10^{-4}$	$-1,3331 \times 10^{-3}$
4	$5,0054 \times 10^{-5}$	$2,6848 \times 10^{-4}$
5	$-1,0073 \times 10^{-5}$	$-5,3621 \times 10^{-5}$
6	$2,0121 \times 10^{-6}$	$1,07271 \times 10^{-5}$

2. táblázat. A sor bemenetét generáló MMAP kereszt-korrelációi

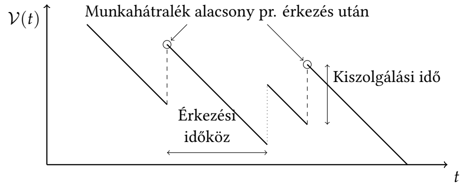
A távozási folyamat (kereszt-) korrelációi a 4. ábrán láthatók, lineáris és logaritmikus skálán egyaránt (ez utóbbi az abszolút értékeket ábrázolja). Jól megfigyelhető, mennyire megváltoztatta a sorbanállás a forgalom viselkedését: a távozási folyamat korrelációjának lecsengése sokkal lassabb, és az alternáló előjel is megszűnt. A korreláció csak az egymástól $n = 1000$ távolságban lévő időközökre csökkent 10^{-5} alá.



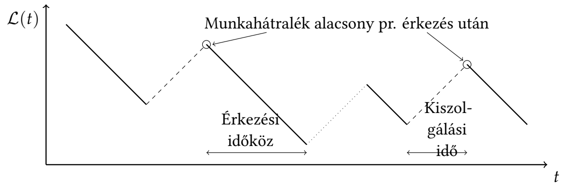


4. ábra. A távozási folyamat (kereszt-) korrelációi

A tézis eredményei lehetővé tették az FCFS ütemezővel rendelkező többosztályos sorok beillesztését az 3.1-es tézis által bevezetett együttes momentum alapú sorbanállási hálózat analízis algoritmusba.



5. ábra. A prioritásos sor munkahátralék folyamata



6. ábra. A prioritásos sor transzformált munkahátralék folyamata

2.4-es tézis

Egy új eljárást fejlesztettem ki a MMAP[K]/PH[K]/1 prioritásos sor teljesítményanalízisére, mind a preemptív, mind a nem preemptív kiszolgálás esetére. Hatékony eljárásokat adtam a rendszerben tartózkodó igények számával, mind a rendszerben töltött idővel kapcsolatos mennyiségek kiszámítására.

Az eljárást [12]-ben publikáltam.

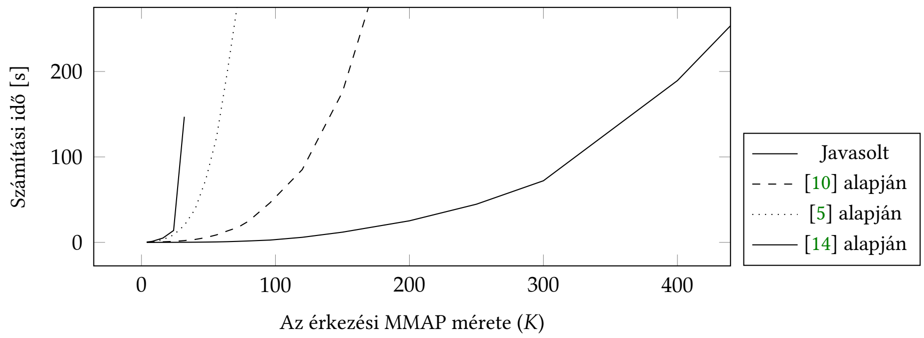
A prioritásos sorok analitikus megoldását a múlt század közepe óta kutatják. Az utóbbi két évtizedben MAP és MMAP érkezési folyamatokkal rendelkező prioritásos sorokra is születtek eredmények. A MAP/G/1 preemptív prioritásos sor munkahátralék folyamatra alapozott megoldása [30]-ben jelent meg, a [31] pedig a nem-preemptív esetet tárgyalja. Ezek a modellek tetszőleges kiszolgálási idő eloszlást megengednek, ami a sor analízisét nagyon összetetté, az eljárás numerikusan hatékony implementálását pedig nehézzé teszi.

A nehézségek megoldására a tézisben általános helyett PH eloszlású kiszolgálási időket feltételezünk. Ez az eljárás is a munkahátralék folyamaton alapul, de a PH eloszlású kiszolgálási idők mellett egy sokkal egyszerűbb, intuitívebb, és numerikusan sokkal előnyösebben viselkedő analízis eljárás bevezetése vált lehetővé.

Az eljárás első lépése a munkahátralék stacioner eloszlásának meghatározása egy alacsony prioritású érkezés pillanatában. A munkahátralék, $\mathcal{V}(t)$, azt az időt jelenti, amennyire a kiszolgálónak szüksége lenne az összes rendszerben lévő igény kiszolgálásához abban az esetben, ha az érkezések megszűnnének. Az érkezési időpontok között $\mathcal{V}(t)$ 1-es meredekséggel csökken, az érkezési időpontokban pedig megugrik az új érkező által behozott többlet munka mennyiségének megfelelően (5. ábra). Esetünkben a PH eloszlású kiszolgálási idők lehetővé teszik, hogy a munkahátralék folyamatot egy Markovi folyadékmodellé transzformáljuk (6. ábra), melynek stacioner analízisére vannak létező hatékony módszerek.

A tézisben javasolt módszer újdonsága, hogy a rendszer legfontosabb teljesítményjellemzőit sikerült ennek, illetve ehhez szorosan köthető Markov folyadékmodellek aktív periódusával kapcsolatba hozni. A Markov folyadékmodellek aktív periódusának vizsgálatára az utóbbi években bevezetett új eljárások segítségével pedig lehetővé vált a prioritásos sorok hatékony analízise is.

A hatékonyság érzékeltetésére három algoritmust hasonlítunk össze a javasolt módszerrel, melyek mind koruk legjobbjai voltak. Az egyik ilyen algoritmus a [14] folytonos időre átültetett változata, a másik ennek továbbfejlesztése ([5]), a harma-



7. ábra. Az analízishez szükséges számítási idő összehasonlítása

dik pedig a [10]-ben publikált eljárás. A 7. ábrán látható az alacsony prioritású igények számának első 10 momentumához szükséges számítási idő az érkezési folyamat méretének függvényében, preemptív ütemezéssel. A javasolt módszer legalább egy nagyságrenddel gyorsabb a többinél, lényegesen nagyobb rendszerek megoldására is alkalmas. Még a legnagyobb fázisszámú esetben sem volt tapasztalható numerikus probléma. Továbbá, [5]-el és [10]-el ellentétben a javasolt módszerrel a rendszerben töltött idővel kapcsolatos teljesítményjellemzők is számolhatók, valamint a nem preemptív ütemezés is vizsgálható.

Fontos megjegyezni, hogy a 2.4-es tézisben használt megközelítést nem lehet felhasználni a távozási folyamat vizsgálatára, mert csak egyes forgalmi osztályokra ad teljesítményjellemzőket. A távozási folyamat vizsgálatához a forgalmi osztályok együttes viselkedését kell tekinteni, e célra a 2.2-es tézis egy teljesen más eljárást alkalmaz.

4 SORBANÁLLÁSI HÁLÓZATOK ANALÍZISE

Motiváció

Olyan sorbanállási hálózatok megoldására, melyekben az igények érkezése nem Poisson folyamat, és a kiszolgálási idők nem exponenciálisak, a múlt évszázad második felében tettek először kísérletet. Az első próbálkozások az érkezési és kiszolgálási időközök második momentumának figyelembe vételére irányultak. Erre a megközelítésre alapul a széles körben elterjedt közelítő algoritmus, a QNA [33, 34] is. Ezek az eljárások azonban – implicit módon – feltételezik, hogy az egymást követő érkezési és kiszolgálási időközök függetlenek egymástól. A többosztályos sorbanállási hálózatokat is régóta vizsgálják ([15]), de a forgalmi osztályok közötti függőségeket nem voltak képesek figyelembe venni.

A sokoldalúan használható Markovi érkezési modellek (MAP-ok, MMAP-ok) megjelenése új lendületet adott az összefüggő forgalommal rendelkező sorbanállási hálózatok analízisének [21, 28]. Ezeknek az új megoldásoknak azonban van egy hátrányos tulajdonsága: a hálózat belső forgalmát leíró MAP modellek mérete csomópontról csomópontra nő, és hamar eléri a gyakorlati kiszámíthatóság korlátait.

Az általunk javasolt megközelítés is MAP-ot, illetve MMAP-ot használ a belső forgalom leírására, de a korábbiaktól teljesen eltérő módon: kiszámítjuk a belső forgalom egylépéses együttes momentumait, és az 1.3-as tézis eredményeire támaszkodva ezekre illesztünk MAP-ot, illetve MMAP-ot. Ennek a megközelítésnek az a fő előnye, hogy a forgalmat modellező MAP-ok mérete nem nő minden alkalommal, amikor áthalad egy csomóponton, mivel a figyelembe vett együttes momentumok száma állandó. Továbbá, a figyelembe vett együttes momentumok számának megfelelő megválasztása lehetőséget biztosít a modellméret és a pontosság közötti kompromisszum kívánt beállítására.

3.1-es tézis

Bevezettem egy új eljárást többosztályos, összefüggő forgalommal rendelkező sorbanállási hálózatok együttes momentum alapú közelítő analízisére.

A kapcsolódó eredményeket [13] és [5] tartalmazza.

A dekompozíció alapú sorbanállási hálózat analízis legkritikusabb kérdése a hálózati forgalom megfelelő modellezése. Az 1.3-as tézis alapján a hálózati forgalmat együttes momentumok segítségével jellemezzük, ami egy igen kompakt, kevés paraméterrel rendelkező reprezentáció. A hálózat analízise során minden csomópont feldolgozása négy lépésből áll:

1. A csomópont érkezési folyamatának előállítása a különböző irányokból érkező forgalmak aggregálásával. Mind a külső, mind a belső forgalmat MAP, illetve MMAP írja le, ezek a forgalmi modellek pedig zártak az aggregáció műveletére, tehát az aggregáció eredménye is MAP, illetve MMAP lesz.

2. A csomópont sorbanállással kapcsolatos teljesítményjellemzőinek (várakozási idő, sorhossz, stb.) kiszámítása. A disszertáció 5. fejezete ismerteti az egyosztályos MAP/MAP/1, a 6. fejezete a többosztályos MMAP[K]/PH[K]/1 FCFS, a 7. fejezete (és a 2.4-es tézis) pedig a prioritásos sor teljesítményanalízisét.

3. A távozási folyamat momentumainak és együttes momentumainak kiszámítása. Erre a célra a 2.1-es, a 2.2-es, és a 2.3-as tézis eredményei szolgálnak.

4. A MAP reprezentáció előállítása a momentumok és együttes momentumok alapján, az 1.3-as tézis eredményeinek és a kapcsolódó (a disszertáció 3.2-es és 3.3-as fejezeteiben található) algoritmusok segítségével.

A javasolt eljárás pontosságát a disszertáció 8.2-es és 8.3-as fejezetében található összehasonlító tanulmány vizsgálja. A legegyszerűbb, egyosztályos két állomásos tandem hálózat esetére az összehasonlításban szereplő algoritmusok néhány érdekes tulajdonságát a 3. táblázat foglalja össze. A táblázat második oszlopa a hálózati forgalmat modellező MAP mérete, mely a Poisson és az együttes momentum alapú algoritmusnál nem függ az érkezési és kiszolgálási folyamat fázisszámától. Az ún. ETAQA csonkolás eredményezi a legnagyobb forgalmi modellt, a minimális csonkolási szint

Távozási folyamatot közelítő algoritmus	Fázisszám	Hiba
Poisson folyamat alapú	1	0,3569/0,1396
Lelassított kiszolgálási folyamat alapú	2	0,1477/0,0961
Aktív periódus alapú	12	0,119/0,219
ETAQA csonkolás 2-es szinten	18	0,1364/0,068
ETAQA csonkolás 50-es szinten	306	0,0034/0,007
Együttes momentum alapú, 4 momentummal	2	0,1116/0,0695
Együttes momentum alapú, 9 momentummal	3	0,076/0,0511

3. táblázat. Különböző távozási folyamatot közelítő algoritmusok összehasonlítása az egyosztályos, két csomópontos példában

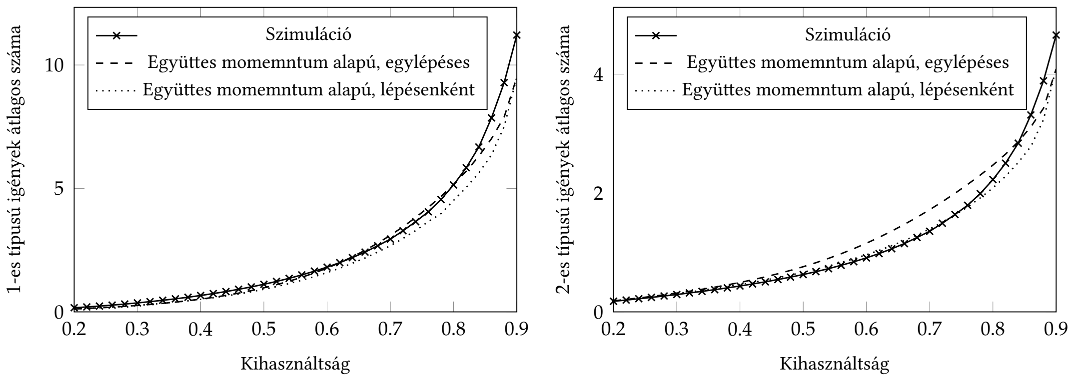
mellett 18 állapotra, a jobb pontosság elérése érdekében 50-re növelt csonkolási szint mellett már 306 állapotra van szükség. Az utolsó oszlop a második csomópontban a rendszerben lévő igények átlagos számára és annak relatív szórásnégyzetére számolt abszolút hiba szimulációs eredményekhez viszonyított relatív értéke. A legpontosabbnak az 50-es csonkolási szinttel előállított ETAQA távozási folyamat modell bizonyult, de az óriási fázisszám és a nem Markovi reprezentáció miatt további számítások elvégzésére ez a modell nem alkalmas. Az együttes momentum alapú megközelítés elfogadható pontosságot hozott, miközben a két csomópont közötti forgalmat modellező MAP Markovi, és rendkívül kompakt reprezentációval rendelkezik, a momentumok számától függően csupán 2, illetve 3 fázisra volt szükség.

A 8. és a 9. ábrán egy kétosztályos, két csomópontos hálózat eredményei láthatók, a grafikonok a második csomópontban tartózkodó igények számát ábrázolják FCFS és prioritásos kiszolgálás mellett. Ebben a példában az 1.3-as tézis momentumillesztő eljárása nem adott vissza Markovi reprezentációt, ezért közelítő momentumillesztést ([7]) kellett alkalmazni.

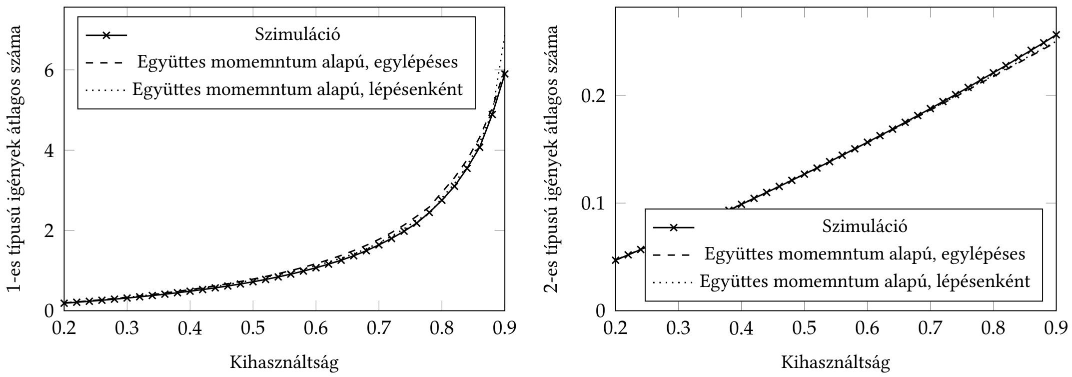
A kiértékelt példák alapján elmondható, hogy több fogalmi osztály esetén az eljárás pontossága valamivel romlik. Azt azonban fontos megjegyezni, hogy a bemutatott együttes momentum alapú megközelítés mind a mai napig az egyetlen lehetőség kellően általános többosztályos sorbanállási hálózatok közelítő analízisére; az egyosztályos hálózatokhoz bevezetett algoritmusokat vagy egyáltalán nem lehet többosztályos esetre kiterjeszteni, vagy csak úgy, hogy nem lesznek képesek a különféle többosztályos kiszolgálási elvek figyelembe vételére.

5 AZ EREDMÉNYEK HASZNOSÍTÁSA

A bemutatott tézisek, illetve azok bizonyos változatai, előzményei, szinte mind felhasználásra kerültek konkrét ipari alkalmazásokban. Főbb ipari partnereink a Nokia és a T-Systems voltak, akik hálózattervező és -optimalizáló algoritmusaikba építették be ezeket az analitikus elemeket.



8. ábra. A második csomópontban tartózkodó igények száma a kétosztályos, két-csomópontos példában, FCFS kiszolgálással



9. ábra. A második csomópontban tartózkodó igények száma a kétosztályos, kétcsomópontos példában, prioritásos kiszolgálással

A TÉZISEKHEZ KAPCSOLÓDÓ PUBLIKÁCIÓK

[1] G. Horváth , M. Telek. „A Canonical Representation of Order 3 Phase Type Distributions," in *Formal Methods and Stochastic Models for Performance Evaluation: Fourth European Performance Engineering Workshop* (Berlin, Germany), pp. 48–62, Springer, 2007.

[2] G. Horváth , M. Telek. „On the canonical representation of phase type distributions," in *Performance Evaluation*, vol. 66:8, pp. 396–409, August 2009.

[3] Gábor Horváth. „Moment matching-based distribution fitting with generalized hyper-Erlang distributions," in *Analytical and Stochastic Modeling Techniques and Applications*, pp. 232–246. Springer, 2013.

[4] M. Telek and G. Horváth, „A minimal representation of Markov arrival processes and a moments matching method," *Performance Evaluation*, vol. 27:9, pp. 1153–1168, October 2007.

[5] A. Horváth, G. Horváth, and M. Telek, „A traffic based decomposition of two-class queueing networks with priority service," *Computer Networks*, vol. 53:8, pp. 1235–1248, June 2009.

[6] G. Horváth, P. Buchholz, and M. Telek, „A MAP fitting approach with independent approximation of the inter-arrival time distribution and the lag correlation," in *Quantitative Evaluation of Systems, 2005. Second International Conference on the*, pp. 124–133, September 2005.

[7] B. Falko, G. Horváth, „Fitting Markovian Arrival Processes by Incorporating Correlation into Phase Type Renewal Processes," in *Quantitative Evaluation of Systems (QEST), 2010 Seventh International Conference on the*, (Williamsburg, Virginia, USA), pp. 97–106, September 2010.

[8] G. Horváth, „Measuring the Distance Between MAPs and Some Applications," in *Analytical and Stochastic Modelling Techniques and Applications: 22nd International Conference*, pp. 100–114, October 2015.

[9] G. Horváth and M. Telek, „Fitting Methods Based on Distance Measures of Marked Markov Arrival Processes," in *Seminal Contributions to Modelling and Simulation: 30 Years of the European Council of Modelling and Simulation*, pp. 159–163, 2016.

[10] G. Horváth, „Efficient analysis of the queue length moments of the MMAP/MAP/1 preemptive priority queue," *Performance Evaluation*, vol. 69:12, pp. 684–700, December 2012.

[11] G. Horváth and B. Van Houdt, „Departure process analysis of the multi-type MMAP[K]/PH[K]/1 FCFS queue," *Performance Evaluation*, vol. 70:6, pp. 423–439, June 2013.

[12] G. Horváth, „Efficient analysis of the MMAP[K]/PH[K]/1 priority queue," *European Journal of Operational Research*, vol. 246:1, pp. 128–139, October 2015.

[13] A. Horváth, G. Horváth, and M. Telek, „A joint moments based analysis of networks of MAP/MAP/1 queues," *Performance Evaluation*, vol. 67:9, pp. 759–778, September 2010.

HIVATKOZÁSOK

[14] Attahiru Sule Alfa, Bin Liu, and Qi-Ming He. Discrete-time analysis of MAP/PH/1 multiclass general preemptive priority queue. *Naval Research Logistics (NRL)*, 50(6):662–682, 2003.

[15] Yonathan Bard. Some extensions to multiclass queueing network analysis. In *Proc. of the Third Int. Symposium on Modelling and Performance Evaluation of Computer Systems*, pages 51–62, Amsterdam, The Netherlands, 1979. North-Holland.

[16] A. Bobbio, A. Horváth, and M. Telek. Matching three moments with minimal acyclic phase type distributions. *Stochastic Models*, 21(2-3):303–326, 2005.

[17] A. Bobbio and M. Telek. A benchmark for PH estimation algorithms: results for Acyclic-PH. *Stochastic Models*, 10(3):661–677, 1994.

[18] A. Cumani. On the Canonical Representation of Homogeneous Markov Processes Modelling Failure-time Distributions. *Microelectronics and Reliability*, 22:583–602, 1982.

[19] Tessa Dzial, Lothar Breuer, Ana da Silva Soares, Guy Latouche, and Marie-Ange Remiche. Fluid queues to solve jump processes. *Performance Evaluation*, 62(1):132–146, 2005.

[20] Qiming He. Analysis of a continuous time SM[K]/PH[K]/1/FCFS queue: Age process, sojourn times, and queue lengths. *Journal of Systems Science and Complexity*, 25(1):133–155, 2012.

[21] A. Heindl, Q. Zhang, and E. Smirni. ETAQA truncation models for the MAP/-MAP/1 departure process. In *QEST '04: Proceedings of the The Quantitative Evaluation of Systems, First International Conference on (QEST'04)*, pages 100–109, Washington, DC, USA, 2004. IEEE Computer Society.

[22] M.A. Johnson and M.R. Taaffe. Matching moments to phase distributions: Mixtures of Erlang distributions of common order. *Stochastic Models*, 5(4):711–743, 1989.

[23] Leonard Kleinrock. *Queueing systems, volume I: theory*. Wiley Interscience, 1975.

[24] Guy Latouche and Vaidyanathan Ramaswami. *Introduction to matrix analytic methods in stochastic modeling*, volume 5. SIAM, 1999.

[25] Vern Paxson and Sally Floyd. Wide area traffic: the failure of Poisson modeling. *IEEE/ACM Transactions on Networking (ToN)*, 3(3):226–244, 1995.

[26] V. Ramaswami. Matrix analytic methods for stochastic fluid flows. In *Teletraffic Engineering in a Competitive World - Proc. of the 16th International Teletraffic Congress (ITC 16)*, pages 1019–1030. Elsevier Science B.V., 1999.

[27] J. Roberts, U. Mocci, and J. Virtamo (eds.). *Broadband Network Teletraffic*. Springer, 1996.

[28] R. Sadre and B.R. Haverkort. Characterizing traffic streams in networks of MAP/-MAP/1 queues. In *Proceedings 11th GI/ITG Conference on Measuring, Modelling and Evaluation of Computer and Communication Systems (MMB 2001)*, pages 195–208. VDE Verlag, 2001.

[29] Bhaskar Sengupta. Markov processes whose steady state distribution is matrix-exponential with an application to the GI/PH/1 queue. *Advances in Applied Probability*, pages 159–180, 1989.

[30] Tetsuya Takine. The workload in the MAP/G/1 queue with state-dependent services: Its application to a queue with preemptive resume priority. *Stochastic Models*, 10(1):183–204, 1994.

[31] Tetsuya Takine. The nonpreemptive priority MAP/G/1 queue. *Operations Research*, 47(6):917–927, 1999.

[32] A. van de Liefvoort. The moment problem for continuous distributions. Technical report, University of Missouri, WP-CM-1990-02, Kansas City, 1990.

[33] W. Whitt. Approximating a point process by a renewal process, I : Two basic methods. *Operations Research*, pages 125–147, 1982.

[34] W. Whitt. Approximations for departure processes and queues in series. *Naval Research Logistics Quarterly*, pages 499–521, 1984.