



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



**FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ**

ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

VYUŽITÍ TEORIE HROMADNÉ OBSLUHY PŘI NÁVRHU A OPTIMALIZACI PAKETOVÝCH SÍTÍ

QUEUEING THEORY UTILIZATION IN PACKET NETWORK DESIGN AND OPTIMIZATION
PROCESS

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. ZDENĚK RÝZNER

VEDOUCÍ PRÁCE

SUPERVISOR

doc. Ing. VÍT NOVOTNÝ, Ph.D.

BRNO 2011



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav telekomunikací

Diplomová práce

magisterský navazující studijní obor
Telekomunikační a informační technika

Student: Bc. Zdeněk Rýzner

ID: 78366

Ročník: 2

Akademický rok: 2010/2011

NÁZEV TÉMATU:

Využití teorie hromadné obsluhy při návrhu a optimalizaci paketových sítí

POKYNY PRO VYPRACOVÁNÍ:

Prostudujte problematiku datových sítí z pohledu teorie hromadné obsluhy, seznamte se s jednotlivými typy obsluhových systémů pro modelování chování přepojovacích prvků a spojů v dnešních datových sítích. Navrhněte modely současných typů síťových prvků (přepínačů a směrovačů) s jednou či více frontami (podpora QoS), ve vybraném programovacím jazyce modely implementujte, navrhněte jednoduchý model sítě a realizujte simulace. Výsledky, jako jsou pravděpodobnostní rozložení zpoždění, jeho proměnlivosti a ztráty dat zhodnoťte a graficky znázorněte. Navrhněte laboratorní úlohu.

DOPORUČENÁ LITERATURA:

- [1] TRIVEDI, K., S., Probability and Statistics with Reliability, Queuing and Computer Science Applications. John Wiley & Sons, ISBN 0-471-33341-7, New York, USA, 2002
- [2] KESIDIS, G. An Introduction to Communication Network Analysis. Wiley-IEEE Press, ISBN: 978-0-471-37141-0, USA, 2007
- [3] KLIKA, O. a kol. Teorie a praxe provozního zatížení. Nakladatelství dopravy a spojů, OS-31-047-73, Praha, 1973

Termín zadání: 7.2.2011

Termín odevzdání: 26.5.2011

Vedoucí práce: doc. Ing. Vít Novotný, Ph.D.

prof. Ing. Kamil Vrba, CSc.

Předseda oborové rady

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

Anotace

Tato diplomová práce se zabývá teorií hromadné obsluhy a jejím využitím při návrhu modelů uzlů v paketové síti. Jsou zde popsány obecné principy pro vytváření modelů systémů hromadné obsluhy a matematické pozadí této problematiky. Dále byl v prostředí MATLAB vytvořen program pro simulaci zpoždění paketů v síti, který implementuje dva druhy popsaných modelů – M/M/1 a M/G/1. Program umožňuje simulovat jednotlivé prvky sítě a získat základní charakteristiky sítě jako je zpoždění či ztrátovost paketů. Byla také navržena laboratorní úloha, ve které se studenti seznámí se základy teorie hromadné obsluhy a vyzkouší si analytický a simulační přístup k řešení systémů hromadné obsluhy.

Klíčová slova: Kendallova klasifikace, laboratorní úloha, Markovův proces, Markovův řetězec, MATLAB, model M/G/1, model M/M/1, Poissonův proces, proces množení a zániku, simulace, stochastický proces, systémy hromadné obsluhy, teorie hromadné obsluhy

Abstract

This master's thesis deals with queueing theory and its utilization in designing node models in packet-switched network. There are described general principles of designing queueing models and its mathematical background. Furthermore, simulator of packet delay in computer network was created. This application implements two described models – M/M/1 and M/G/1. Application can be used for simulation of network nodes and for obtaining basic network characteristics like packet delay or packet loss. Laboratory training exercise was also created. In this exercise students familiarize themselves with basic concepts of queueing theory and examine both analytical and simulation approach to solving queueing systems.

Keywords: Birth-death process, continuous-time Markov chain, discrete-time Markov chain, Kendall's classification, laboratory training exercise, MATLAB, M/G/1 model, M/M/1 model, Poisson process, queueing systems, queueing theory, stochastic process

Bibliografická citace práce

RÝZNER, Z. *Využití teorie hromadné obsluhy při návrhu a optimalizaci paketových sítí*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2011. 56 s. Vedoucí diplomové práce doc. Ing. Vít Novotný, Ph.D..

Prohlášení

Prohlašuji, že svou diplomovou práci na téma využití teorie hromadné obsluhy při návrhu a optimalizaci paketových sítí jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením tohoto projektu jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

V Brně dne

.....

podpis autora

Poděkování

Rád bych poděkoval doc. Ing. Vítu Novotnému, Ph.D. za odborné vedení, za pomoc a rady při zpracování této diplomové práce.

OBSAH

Úvod	8
1 Teorie hromadné obsluhy	10
1.1 Základní prvky a parametry systému hromadné obsluhy	11
1.1.1 Vstupní proud požadavků	12
1.1.2 Doba obsluhy	13
1.1.3 Čekání ve frontě	13
1.1.4 Režim výstupu z fronty	13
1.1.5 Režim obsluhy	14
1.2 Klasifikace systémů hromadné obsluhy	15
2 Stochastické procesy	17
3 Markovovy modely	18
3.1 Markovovy řetězce	18
3.2 Markovovy procesy se spojitým časem	20
3.2.1 Poissonův proces	21
3.2.2 Procesy množení a zániku	23
4 Modelování síťových prvků a síťového provozu	27
4.1 Model M/M/1	28
4.1.1 Model M/M/1 s prioritami (2 prioritní třídy)	31
4.1.2 Model M/M/1 s konečnou délkou fronty	32
4.2 Model M/G/1	33
4.2.1 Model M/G/1 s prioritami (2 prioritní třídy)	35
4.2.2 M/G/1/K s konečnou délkou fronty	35
5 Programová realizace	36
5.1 Popis programu	36
5.1.1 Úvodní popis	36
5.1.2 Práce s programem	37
5.1.3 Spuštění simulace	40
5.1.4 Doplnující poznámky	43
5.2 Popis řešení funkcí, které realizují výpočty	43
5.2.1 Výpočet modelu M/M/1	44

5.2.2	Výpočet modelu M/M/1 s prioritami.....	45
5.2.3	Výpočet modelu M/G/1	45
5.2.4	Výpočet modelu M/G/1 s prioritami	45
5.3	Orientační porovnání simulace a výpočtu pomocí vzorců.....	46
6	Návrh laboratorní úlohy	47
	Závěr	48
	Seznam použité literatury	50
	Seznam obrázků.....	53
	Seznam použitých zkratk, veličin a symbolů	54
	Seznam příloh	56

ÚVOD

Teorie hromadné obsluhy (v zahraniční odborné literatuře často také nazývaná jako teorie front) vznikla jako matematická disciplína počátkem 20. století v souvislosti s rozvojem telefonních ústředen, pro jejichž návrh byla zpočátku určena.

Teorie hromadné obsluhy zkoumá modely, v nichž do nějakého systému, který může mít jednu či více linek obsluhy vstupují jednotky, které mají být těmito linkami obslouženy. Typickým příkladem pro oblast telekomunikací jsou datové sítě, kde obslužnými prvky jsou především přepojovací prvky – směrovače a přepínače, ale také samotné spoje mezi přepojovacími prvky. Tomu velmi podobná je problematika dopravy. Dalším obsluhovým systémem může být například prodejna, kam přicházejí zákazníci a musejí projít přes některou pokladnu, kde zaplatí, nebo zařízení hromadného stravování, jako jídelny či restaurace, kde máme vývařovnu a skupinu číšníků obsluhující, a našlo by se mnoho dalších příkladů. Cílem teorie hromadné obsluhy je navrhnout matematický aparát pro popis a následnou analýzu systému za účelem zjištění výkonnostních ukazatelů systému, jako jsou doba, kterou požadavek stráví v systému, doba strávená ve frontách a v obslužných jednotkách, úroveň obsazenosti front, podíl neobsloužených požadavků (ztrát) a řadu dalších parametrů, které pak slouží k revizi počátečního návrhu a nebo k optimalizaci již vybudovaného systému, aby byl systém ekonomický, a přitom nebyly překročeny klíčové parametry jako jsou zpoždění či procento povolených ztrát.

Způsob implementace a forma obsluhy systémů front tvoří důležitou část moderních telekomunikačních systémů a jedná se o nedílnou součást všech moderních telekomunikačních zařízení. V paketově přepínané síti zpravidla přenášené datové toky patří různým službám s různými požadavky na parametry přenosu. Aby mohly být jednotlivé služby provozovány v odpovídající kvalitě, je nutné, aby síť, resp. její uzlové prvky, zajišťovaly určitou míru garance přenosových parametrů (kvalitu služeb). Kvalitu služeb lze sledovat z různých pohledů – nejběžněji a nejčastěji sledovanými parametry jsou zpoždění a chybovost (ztrátovost). Sledované parametry jsou zpravidla náhodnými veličinami a výsledky mohou být prezentovány například jejich střední hodnotou, rozptylem či přesně pomocí distribuční funkce.

Cílem této diplomové práce je prostudovat problematiku datových sítí z pohledu teorie hromadné obsluhy, navrhnout modely síťových prvků, vytvořit simulační program, ve kterém

jsou aplikovány získané poznatky a navrhnout laboratorní úlohu na téma teorie hromadné obsluhy.

Úvodní kapitola práce je věnována obecným vlastnostem systémů hromadné obsluhy. Stěžejním pilířem modelů, využívajících teorii hromadné obsluhy jsou stochastické modely. Ty pracují s náhodnými procesy a veličinami. Teorie hromadné obsluhy využívá teorii pravděpodobnosti, matematické statistiky a teorii náhodných proměnných. Stěžejní části matematického aparátu dané problematiky – Markovovým procesům – se věnuje druhá a třetí kapitola práce.

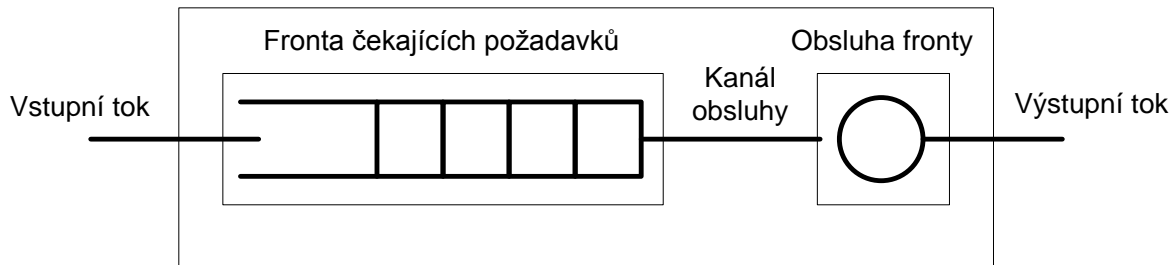
Způsob, jakým budou modelovány síťové uzly a datové toky, je popsán ve čtvrté kapitole.

Aplikaci teoretických poznatků z předchozích kapitol do praxe, kterou v mé práci představuje realizace programu pro simulaci parametrů uzlů paketové sítě, se věnuje pátá kapitola. V prostředí MATLAB byl vytvořen program s grafickým uživatelským rozhraním, který umožňuje simulovat prvky sítě z jedné, pevně dané topologie sítě. Uživatel volí použitý simulační model, délku simulace a počet cyklů. Po provedení simulace získá řadu výstupních parametrů ve formátu středních hodnot, rozptylu a směrodatné odchylky. Také jsou vygenerovány grafy, zobrazující průběh zpoždění, rozložení zpoždění a stav zaplnění čekací fronty jednotlivých uzlů. V této kapitole je taktéž uveden popis toho, jak konkrétně byly jednotlivé výpočetní funkce vytvořeny. Poslední částí práce je návrh laboratorní úlohy, ve které se studenti seznámí se základy systémů hromadné obsluhy a vyzkouší si řešení takových systémů jak analyticky, tak s pomocí vytvořeného simulačního programu.

1 TEORIE HROMADNÉ OBSLUHY

Teorie hromadné obsluhy je matematická disciplína, jenž zkoumá obslužné systémy. Těmito obslužnými systémy může být například dopravní systém, obchod, telekomunikační síť, telekomunikační služba, opravna, jídelna a mnoho dalších systémů. Obslužný systém může mít různou konfiguraci, obecně však sestává ze vstupů, kterými přicházejí požadavky, dále front pro čekání na obsluhu, bloku obsluhy požadavků a výstupů. Čekací část může sestávat z jedné či více vstupních front s různou velikostí a s různým způsobem řazení nebo obsluhou požadavků, případně doplněných i o systém ochrany před zahlcením. Blok obsluhy může mít jednu, či více obslužných jednotek. V závislosti na poměru intenzity příchodu požadavků a výkonnosti bloku obsluhy může docházet k pozdržení požadavků ve frontách, případně dokonce i k odmítnutí požadavku (ztrátě), hrozí-li (nebo už i došlo k) zahlcení fronty či doba čekání požadavku na obsluhu překročila časový limit závislý na typu služby.

U modelů hromadné obsluhy je často cílem měření vzájemných vazeb mezi základními ukazateli charakterizujícími kvalitu a efektivitu systému hromadné obsluhy a nalezení optimálního provozního režimu vzhledem ke zvolenému kritériu – optimalizace systému zpravidla vede podstatnému snížení možnosti vzniku front, nebo snížení celkových ztrát z důvodu čekání. Na druhou stranu je také nutné se zabývat i ekonomickou stránkou problému – snížení pravděpodobnosti vzniku front je ve většině případů dosaženo rozšířením kapacity obslužného zařízení, tedy za cenu navršení prostředků na systém. Cílem je tedy nalézt optimální relaci mezi určitou mírou kvality poskytovaných služeb (představovanou dobou obsluhy, ztrátami a dalšími kvalitu snižujícími parametry) a náklady na dosažení dané kvality služby (kdy se náklady pak promítnou do ceny služby pro zákazníka). Příklad systému hromadné obsluhy s jednou frontou a jedním obslužným prvkem je uveden na obr. 1:



Obr. 1: Příklad systému hromadné obsluhy

Se systémy hromadné obsluhy (SHO) se setkáváme v mnoha oblastech reálného života. Příkladem systému hromadné obsluhy může být prodejna, kde proud požadavků tvoří zákazníci, kteří přicházejí nakupovat a obslužnými zařízeními jsou prodavači, kteří je obsluhují. Podobného charakteru je i zpracování informací použitím výpočetní techniky, kterým se zabývá tato práce. Mnoho úloh hromadné obsluhy se vyskytuje též ve výrobě. Převážná část systémů hromadné obsluhy má stochastický charakter. Možnosti řešení jsou dvojího typu:

- a) Analytické – na základě známých parametrů modelu jsou pomocí nástrojů teorie pravděpodobnosti či jiných matematických odvětví vypočteny nebo odhadnuty ty parametry modelu, které nás zajímají (např. průměrný počet zákazníků ve frontě, apod.).
- b) Simulační – využije se známých parametrů modelu k nasimulování dané situace pomocí vhodného software. Parametry modelu, které nás zajímají, jsou odhadnuty na základě těchto simulací.

Výsledky získané na základě aplikace určitého modelu systému hromadné obsluhy jsou použitelné jenom tehdy, jestliže model dostatečně věrně odráží základní rysy reálného systému. Využití teorie pravděpodobnosti v řešení praktických úloh hromadné obsluhy se objevilo v Erlangových studiích, které řešily problematiku automatizovaného telefonního provozu. Systematickým zpracováním a klasifikací problémů hromadné obsluhy se věnoval D. G. Kendall [12].

1.1 Základní prvky a parametry systému hromadné obsluhy

Při popisu SHO matematickým modelem je nutné vyjít z analýzy jeho chování. V první fázi se zkoumají příčinné oboustranné vazby základních prvků systému – tj. specifikace **vstupního proudu požadavků**, způsob a mechanismus jejich přijímání a předzpracování (**režim fronty**) a **charakter dob trvání obsluhy**. Na základě logické analýzy a klasifikace zkoumaného systému hromadné obsluhy je následně možné formulovat odpovídající model.

Další fází modelování systému hromadné obsluhy je statistická analýza jeho jednotlivých prvků zahrnující i statistické testování výchozí hypotézy o charakteru rozdělení odpovídajících veličin i předpoklady o režimu fungování zkoumaného systému. Po získání potřebných statistických dat lze chování SHO popsat matematickým modelem. Výsledkem jeho řešení jsou hodnoty základních, zpravidla průměrných charakteristik, jejichž znalost

umožňuje udělat si představu o kvalitě a efektivitě provozu modelovaného systému. Z hlediska provozovatele systému hromadné obsluhy jsou důležitými průměrnými charakteristikami například vytížení a prostoje jednotlivých zařízení obsluhy, či počet nebo podíl požadavků čekajících na obsluhu [13]. Zákazníka zase budou zajímat parametry, jako rozložení doby obsluhy, proměnlivost doby obsluhy během jedné relace a ztrátovost požadavků v systému.

Aby bylo možné vypočítat základní charakteristiky systému hromadné obsluhy, je potřeba adekvátním způsobem popsat i jeho hlavní prvky. Stochastický charakter mají především dva prvky systému – vstupní proud požadavků a doba trvání obsluhy. Lze je proto popsat pouze přibližně pomocí známých rozdělení pravděpodobnosti. Při popisu ostatních prvků systému se zpravidla problémy tohoto druhu nevyskytují. Základní charakteristiky, které sledujeme při modelování SHO, zpravidla zahrnují průměrný počet požadavků v systému a ve frontě, průměrnou dobu strávenou v systému, průměrnou dobu čekání ve frontě či pravděpodobnost blokování. Matematické vztahy použité ke zjištění těchto charakteristik jsou pro jednotlivé modely uvedeny v kapitole 4.

1.1.1 Vstupní proud požadavků

Vlastnosti a intenzita vstupního proudu mají rozhodující vliv na činnost systému hromadné obsluhy a na jeho základní charakteristiky. Požadavky přicházejí do SHO nejčastěji v náhodných okamžicích, takže vstupní proud požadavků lze obecně charakterizovat jako stochastický proces. Délky intervalů mezi jejich příchody představují hodnoty spojitých náhodných veličin. Údaje získané ze SHO velmi často vyhovují předpokladu Poissonova rozdělení počtu vstupujících požadavků za určitý interval. Tomuto rozložení počtu požadavků odpovídá exponenciální rozdělení intervalů mezi jejich příchody, takže systémy s Poissonovým vstupním proudem požadavků se také nazývají exponenciální systémy.

V praxi je však možno se setkat i s jinými zákony rozdělení vstupního proudu, jako jsou např. Erlangovo rozdělení, obecné rozdělení a jiná. Řešení modelů SHO s takovým vstupním proudem požadavků je však analyticky obtížnější.

Pokud je zdroj, ze kterého přicházejí požadavky do systému obsluhy omezený (konečný), jde o uzavřený systém, v opačném případě se jedná o otevřený systém. Zdroj požadavků je považován za neomezený i v případě, že nelze s jistotou určit, kolik požadavků si může vyžádat obsluhu.

1.1.2 Doba obsluhy

Stejně jako v případě vstupního proudu, je při modelování systému obsluhy potřeba popsat i dobu obsluhy jednotlivých požadavků. Dobu obsluhy jednotlivých požadavků lze také pokládat za náhodnou veličinu, která se též obvykle řídí exponenciálním rozdělením [13].

1.1.3 Čekání ve frontě

Podle „trpělivosti“ požadavků na obsluhu se SHO dělí na systémy bez čekání a systémy s čekáním. První typ je charakterizován tak, že pokud jsou všechna zařízení obsluhy obsazena, požadavek do systému nevstoupí a rezignuje na obsluhu. Ve druhém typu systému každý požadavek, který vstoupí, buď čeká tak dlouho, dokud neopustí systém po úspěšném ukončení obsluhy, nebo čeká alespoň určitou dobu a teprve pak, není-li během této doby obsloužen, opouští systém.

Mezi systémy obsluhy s čekáním patří jak systémy s neomezenou délkou fronty, tak systémy s omezenou kapacitou fronty (popř. systému). Je-li počet míst ve frontě a tím v systému omezen, pak při jejich plném obsazení není dalším požadavkům přicházejícím do systému obsluha poskytnuta. Při neomezené délce fronty naopak všechny požadavky, které přicházejí do systému a jsou ochotny čekat, obslouženy.

Systémy hromadné obsluhy, ve kterých nejsou z různých důvodů obslouženy všechny přicházející požadavky, se nazývají systémy se ztrátami. Jedná se tedy jak o systémy bez čekání, tak o systémy s čekáním, které mají buď omezenou kapacitu, nebo požadavky, které nebyly uspokojeny do určitého časového limitu během čekání ve frontě.

1.1.4 Režim výstupu z fronty

Způsob, jakým se požadavky vstupující do systému řadí do fronty, popř. pořadí, v němž jsou čekající požadavky obsluhovány, charakterizují režim výstupu z fronty. Správně zvolený režim fronty je také významným faktorem ovlivňujícím efektivnost obsluhy. Většinou jsou požadavky obsluhovány v pevně určeném pořadí, kam patří dva základní režimy:

- a) FIFO (fronta) – požadavky jsou obsluhovány podle pořadí vstupu do systému
- b) LIFO (zásobník) – požadavky jsou obsluhovány v obráceném pořadí, než v jakém vstoupily do systému

Systémy obsluhy s pevně stanoveným pořadím se označují jako systémy s uspořádanou frontou. Není-li stanoveno pevné pořadí vstupu požadavků do obsluhy, může režim fronty

bud' zohledňovat důležitost požadavků či naléhavost jejich obsluhy vyjádřenou různým stupněm priority (PRI), anebo jsou požadavky k obsluze přiřazovány zcela náhodně, bez jakýchkoliv předem daných pravidel (SIRO). V prvním případě se jedná o systémy s prioritami, v druhém o systémy s neuspořádanou frontou.

Pokud jsou druhy priorit známy před vstupem požadavků do systému, jde o apriorní priority. Přiřazuje-li priority samotný systém v průběhu čekání požadavků na obsluhu, jde o aposteriorní priority. V systémech s prioritami rozlišujeme dva druhy priorit: preemptivní (absolutní) a nepreemptivní (relativní). Nepreemptivní priorita zajišťuje zvýhodněným požadavkům přednostní obsluhu před požadavky s nižší prioritou. Požadavek s určitým stupněm nepreemptivní priority je při všech obsazených zařízeních obslužen, jakmile se uvolní některé místo v obsluze. Preemptivní priorita znamená, že preferovaný požadavek je okamžitě obslužen i za cenu přerušování již probíhající obsluhy jiného požadavku s nižší prioritou.

Z hlediska počtu front, které se v systému vytvářejí, rozlišujeme ještě systémy s jednou frontou pro všechna zařízení obsluhy nebo s více frontami, kdy pro každé zařízení existuje zpravidla zvláštní fronta. V takových případech je obvykle možno během čekání na obsluhu přecházet z jedné fronty do jiné fronty.

1.1.5 Režim obsluhy

Počet a uspořádání obslužných linek představuje další faktor, na kterém závisí efektivnost provozu systému hromadné obsluhy. Probíhá-li obsluha pouze na jedné obslužné lince, jedná se o systém s jednoduchou obsluhou (jednokanálový systém obsluhy). Lze-li obsluhovat současně několik požadavků, jde o systém s vícenásobnou obsluhou neboli o vícekanálový systém obsluhy. Speciálním případem je adaptabilní systém, jehož počet obslužných linek se může se měnit (např. úměrně délce fronty). V systémech s více obslužnými linkami se zpravidla používá paralelní uspořádání, existuje ovšem i řazení sériové. Za sebou řazená zařízení obsluhy se vyskytují v systémech s vícefázovým neboli s Erlangovým režimem obsluhy, kdy požadavky procházejí postupně několika různorodými zařízeními obsluhy.

V praktických úlohách se také rozlišuje, zda je zkoumaný systém obsluhy již v provozu nebo zda se teprve projektuje. V prvním případě lze analýzou zkvalitnit fungování SHO nebo zlepšit jeho organizaci a řízení, čímž lze například dosáhnout zkrácení doby obsluhy. V druhém případě jde o odhad kvalitativní charakteristiky systému, který vychází z různých

variant jeho budoucí kapacity a z podmínek, ve kterých bude fungovat. Zároveň se zkoumají pravděpodobnosti vzniku nežádoucích situací v SHO jako jsou například růst fronty nebo čekací doby ve frontě nad předem danou hranici.

1.2 Klasifikace systémů hromadné obsluhy

SHO lze klasifikovat z různých hledisek. Nejčastěji používanými kritérii jsou charakter a typ rozdělení vstupního proudu požadavků, charakter a typ rozdělení doby obsluhy, režim obsluhy, čekání ve frontě, počet míst ve frontě a režim výstupu z fronty. Standardní klasifikace základních typů systémů hromadné obsluhy byla navržena D. G. Kendallem. Používá třímístného kódu, založeného na kombinaci dvou písmen a jedné číslice ve formě $A/B/s$, kde první písmeno A označuje rozdělení intervalů mezi příchody požadavků, druhé písmeno B slouží k označení rozdělení dob trvání obsluhy a s udává počet paralelních zařízení obsluhy v systému [17]. Nejčastěji používanými symboly jsou:

M – exponenciální rozdělení (Markovova typu),

E_k – Erlangovo rozdělení k -tého řádu,

D – deterministické rozdělení,

U – rovnoměrné rozdělení,

H_k – hyperexponenciální rozdělení k -tého řádu,

GI – obecné nezávislé rozdělení,

G – obecné rozdělení.

Například $M/M/1$ označuje SHO s exponenciálním rozdělením intervalů mezi příchody požadavků i dob obsluhy s jednou obslužnou linkou. Obdobně systém typu $D/H_2/1$ je jednodánalový s pravidelným vstupním proudem požadavků a s hyperexponenciálním rozdělením druhého řádu dob trvání obsluhy. Kromě třímístného kódu se někdy používá i rozšířený pětímístný kód typu $A/B/s/x/y$, kde x označuje maximální počet možných požadavků v systému (obsluhovaných + čekajících na obsluhu) a y režim fronty. Méně často se vyskytuje pětímístný kód, v němž x znamená počet míst ve frontě a y velikost zdroje

požadavků. Je-li některý z deskriptorů x nebo y vynechán, předpokládá se, že jeho hodnota je nekonečně velká.

2 STOCHASTICKÉ PROCESY

Stochastické modely jsou založeny na aplikaci počtu pravděpodobnosti. Vychází se u nich z předpokladu, že k určitým změnám dochází s určitými pravděpodobnostmi – pracuje se s náhodnými veličinami.

Hlavním nástrojem modelování stochastických procesů jsou náhodné procesy. Popisují změny procesů v čase, které souvisejí s působením náhodných faktorů. Náhodný proces lze obecně definovat jako množinu náhodných veličin, závislých na určitém počtu parametrů, z kterých je každý definován na množině reálných čísel. Náhodné procesy závislé na jednom parametru (v tomto případě závislé na čase) se nazývají stochastické procesy. Tento proces je definován jako množina náhodných veličin $\{X(t), t \in T\}$, kde T představuje množinu časových okamžiků. V závislosti na charakteru množiny T se stochastické procesy rozdělují na procesy s diskretním časem, jestliže T je spočetná množina a na procesy se spojitým časem, je-li T interval. Pokud ke změnám systému může docházet pouze v diskretních časových okamžicích $t = kT_{vz}$ a mezi sousedními okamžiky je systém neměnný, pak se hovoří o Markovově řetězci, jinak se jedná o Markovův proces se spojitým časem. Rozdělení náhodných veličin $X(t)$ souvisí se stavy procesu, kterých může modelovaný systém nabývat. Stavem stochastického procesu $\{X(t), t \in T\}$ rozumíme určitou zjednodušenou situaci, vymezenou za účelem studia modelu zkoumaného procesu. V určitém okamžiku se vyskytuje situace z určitého souboru možností. Hovoříme o výskytu stavu i v okamžiku n , přičemž je-li množina stavů sledovaného procesu spočetná, jedná se o stavově diskretní proces, v opačném případě o stavově spojitý proces. V diskretním případě se množina stavů často zobrazuje do množiny přirozených čísel, takže stavy procesu se značí čísla $1, 2, \dots, N$.

3 MARKOVOVY MODELY

Markovovy modely jsou funkce dvou náhodných proměnných – stavu soustavy a času nebo jiné veličiny, na jejíž závislosti stav soustavy sledujeme. Obě veličiny mohou být spojité nebo diskrétní, čemuž odpovídají čtyři druhy modelů. V dalším textu se budeme věnovat dvěma modelům: Markovskému řetězci (proces s diskrétními stavy a diskrétním časem) a Markovskému procesu (proces s diskrétními stavy a spojitým časem).

3.1 Markovovy řetězce

Markovovy řetězce jsou nejjednodušším typem procesů, které mají Markovskou vlastnost. Ta říká, že v každém stavu procesu je pravděpodobnost navštívení dalších stavů nezávislá na dříve navštívených stavech. To znamená, že chování v Markovových řetězcích je „bezpaměťové“. U Markovových řetězců se předpokládá diskrétnost času i stavů. Předpoklad, že v Markovském řetězci je čas diskrétní, znamená, že nás stavy procesu zajímají pouze v určitých okamžicích, které tvoří rostoucí posloupnost. Mezi těmito okamžiky se v Markovském řetězci nic neděje, čas mezi těmito okamžiky v modelu neexistuje. Markovovy řetězce se uplatňují při popisu systémů, které se mohou nacházet v jednom z konečného, popřípadě v jednom z nekonečného, ale spočetného počtu stavů. Konkrétním příkladem Markovského řetězce může být např. model předpovědi počasí.

Předpokládáme, že máme celkem M stavů (počet může být konečný nebo spočetný) a že diskrétní čas uvažujeme od okamžiku „0“. Chování systému lze popsat:

- a) vektorem absolutních (tj. nepodmíněných) pravděpodobností v určitém okamžiku $\mathbf{p}(n) = [p_1(n), p_2(n), \dots, p_M(n)]$ pro $n = 0, 1, 2, \dots$, kde $p_i(n)$ značí pravděpodobnost, že proces je v okamžiku nT_{vz} ve stavu i .
- b) maticí pravděpodobností přechodu (z jednoho stavu do jiného, popř. pravděpodobnosti setrvání v témže stavu) po n krocích $\mathbf{P}(n) = \{p_{ij}(n)\}$ pro $n = 0, 1, 2, \dots$, $i, j = 1, 2, \dots, M$

Pravděpodobnost $p_{ij}(n)$ se nazývá podmíněnou pravděpodobností přechodu ze stavu i do stavu j , k němuž dojde mezi okamžiky $n-1$ a n . Jestliže tyto pravděpodobnosti $p_{ij}(n)$ nezávisí na tom, mezi kterými okamžiky k přechodu dojde, tedy pokud platí $p_{ij}(n) = p_{ij}$, pak se příslušný Markovův řetězec nazývá homogenní.

Vzhledem k tomu, že se jedná o pravděpodobnosti, musí $p_{ij}(n)$ splňovat podmínky:

$$p_{ij}(n) \geq 0, \text{ pro všechna } i, j \text{ a} \quad (3.1)$$

$$\sum_{j=1}^M p_{ij}(n) = 1, \text{ pro všechna } i, j. \quad (3.2)$$

Tedy matice pravděpodobností přechodu $\mathbf{P}(n)$ má nezáporné prvky a jedničkové řádkové součty.

Přechod systému ve dvou po sobě následujících okamžicích lze popsat tímto základním schématem

$$\mathbf{p}(n+1) = \mathbf{p}(n) \cdot \mathbf{P}, \quad (3.3)$$

resp. přechod mezi dvěma (počátečním a aktuálním) časovými okamžiky:

$$\mathbf{p}(n+1) = \mathbf{p}(n) \cdot \mathbf{P} = \mathbf{p}(n-1) \cdot \mathbf{P}^2 = \dots = \mathbf{p}(0) \cdot \mathbf{P}^{n+1}. \quad (3.4)$$

Pro přechod od prvního („0“) ke druhému („1“) časovému okamžiku to znamená vyjádření

$$\mathbf{p}(1) = \mathbf{p}(0) \cdot \mathbf{P}. \quad (3.5)$$

Podobně lze odvodit vztah pro absolutní pravděpodobnosti $\mathbf{p}(2)$, $\mathbf{p}(3)$, ..., $\mathbf{p}(n)$, takže postupným dosazováním (viz lit. [13]) dojdeme ke vztahu

$$\mathbf{p}(n) = \mathbf{p}(0) \cdot \mathbf{P}^n. \quad (3.6)$$

Chování homogenních Markovových řetězců po n časových okamžicích je tedy popsáno mocninami matice pravděpodobností přechodu a vektorem výchozích absolutních pravděpodobností.

Vhodnou souhrnnou reprezentací všech pravděpodobností přechodu po 1 kroku $p_{ij} = p_{ij}(1)$ je maticová forma, při níž matici \mathbf{P} zapíšeme jako (i, j udává číslo stavu)

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1M} \\ p_{21} & p_{22} & \dots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M1} & p_{M2} & \dots & p_{MM} \end{pmatrix}, \text{ pro } i, j = 1, 2, \dots, M. \quad (3.7)$$

Protože p_{ij} jsou pravděpodobnosti, musí být splněny vlastnosti (3.1) a (3.2).

Na diagonále matice \mathbf{P}^n dostáváme pravděpodobnosti návratu do téhož stavu po n časových periodách. Jestliže p_{ij}^n je různé od nuly pro libovolné n , hovoříme o stavech rekurentních, v opačném případě o stavech transientních. Pravděpodobnosti přechodu vyšších

stupňů p_{ij}^n dovolují rozlišit dále stavy dosažitelné a nedosažitelné z určitého stavu. Stav j je dosažitelný ze stavu i , platí-li $p_{ij}^n > 0$, tj. existují-li nenulové pravděpodobnosti přechodu ze stavu i do stavu j po n časových okamžicích. V opačném případě je stav nedosažitelný. Stavy vzájemně dosažitelné se nazývají sousledné. Skupina vzájemně sousledných stavů představuje uzavřenou třídu. V případě, že je v řetězci taková třída jedna, nazýváme jej nerozložitelným. Tvoří-li všechny stavy řetězce uzavřenou třídu a jsou ergodické, nazýváme řetězec regulárním. Jestliže pro jeden nebo několik stavů platí, $p_{ij} = 1$ (tedy setrvání ve stavu j je jev jistý) a do těchto stavů existuje vstup, pak se jedná o stavy absorpční a řetězec obsahující tyto stavy nazýváme absorpčními. Definice jednotlivých stavů lze nelézt v literatuře [16].

3.2 Markovovy procesy se spojitým časem

Markovův proces se spojitým časem předpokládá, že přechody mezi jednotlivými stavy se mohou uskutečnit v libovolně blízkých časových okamžicích a tedy případy změn lze vystihnout ve spojitém čase. Na rozdíl od Markovových řetězců nás zde zajímají nejen změny stavů, ale také doby, za jak dlouho ke změně stavu dochází. Konkrétním příkladem může být např.: model sledující dobu poruchy nějakého zařízení.

Náhodné proměnné $X(t)$ nabývají hodnoty, které jsou přiřazené určitým stavům (stejně jako u Markovových řetězců). V okamžiku t_i se může vyskytnout jeden ze stavů i_1, i_2, \dots, i_M ; okamžiky t_i, t_{i+1}, \dots se liší o veličinu Δt , blížící se k nule.

Intenzita pravděpodobnosti přechodu ze stavu i do stavu j je vyjádřena pomocí limity

$$\lim_{\Delta t \rightarrow 0} \frac{p_{ij}(t, t + \Delta t)}{\Delta t} = a_{ij}(t) \geq 0. \quad (3.8)$$

Intenzita výstupu ze stavu j představuje doplněk pravděpodobnosti setrvání $p_{jj}(t, t + \Delta t)$ (v limitně malé době by se totiž pravděpodobnost setrvání blížila jedné)

$$\lim_{\Delta t \rightarrow 0} \frac{1 - p_{jj}(t, t + \Delta t)}{\Delta t} = a_{jj}(t) \geq 0. \quad (3.9)$$

Matici pravděpodobností přechodu, zachycující podmíněné pravděpodobnosti výskytu určitých stavů v době $t + \Delta t$, podmíněné výskytem určitých stavů v době t , lze zapsat

$$\mathbf{P}(t) = \begin{pmatrix} 1 - a_{11}(t)\Delta t & a_{12}(t)\Delta t & \cdots & a_{1n}(t)\Delta t \\ a_{21}(t)\Delta t & 1 - a_{22}(t)\Delta t & \cdots & a_{2n}(t)\Delta t \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}(t)\Delta t & a_{n2}(t)\Delta t & \cdots & 1 - a_{nn}(t)\Delta t \end{pmatrix} \quad (3.10)$$

A matice intenzit pravděpodobností přechodu má tvar

$$\mathbf{A}(t) = \begin{pmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t) \\ a_{21}(t) & a_{22}(t) & \cdots & a_{2n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}(t) & a_{n2}(t) & \cdots & a_{nn}(t) \end{pmatrix}. \quad (3.11)$$

Jestliže proces nezávisí na době t , která uplynula od počátku procesu (závisí pouze na hodnotě Δt), jedná se opět o homogenní proces.

Vektor pravděpodobností lze vyjádřit pomocí matice intenzit pravděpodobností přechodu.

Pro homogenní proces platí vztah (odvození viz lit. [13])

$$\mathbf{p}'(t) = \mathbf{p}(t) \cdot \mathbf{A}, \quad (3.12)$$

jenž má řešení ve tvaru:

$$\mathbf{p}(t) = \mathbf{p}(0) \cdot e^{\mathbf{A}t}. \quad (3.13)$$

Z tohoto vztahu pak lze odvodit rovnici pro stacionární pravděpodobnosti procesu ve tvaru

$$\mathbf{p} \cdot \mathbf{A} = \mathbf{0}, \quad (3.14)$$

kde $\mathbf{0}$ značí nulový vektor.

Z aplikačního hlediska jsou nejvýznamnější jednoduché Markovské procesy, u kterých jsou přechody mezi stavy velmi omezené. Mezi tyto procesy patří Poissonův proces a procesy množení a zániku, popsané v následující části.

3.2.1 Poissonův proces

Poissonův proces je asi nejdůležitější a nejznámějším zástupcem skupiny procesů nazývaných bodové procesy. Poissonovo rozložení příchodu požadavků poprvé popsal A.K. Erlang v [11]. Pokud uvažujeme příchod požadavků jako Poissonův proces, dojde ke zjednodušení analýzy jinak komplexních problémů. Ve většině případů je popis pomocí Poissonova rozložení vhodný – čím komplexnější proces je, tím lépe bude vystižen pomocí Poissonova procesu, protože Poissonův proces je „nejvíce náhodný“ – je charakterizován pouze intenzitou vyskytujících se jevů, resp. střední hodnotou těchto jevů, viz dále.

Bodový proces lze definovat jako výskyt jevů $X(t)$ v čase $(0, t)$, které se vzájemně liší jen různým umístěním v čase. Jde tedy o posloupnost jevů, které se vyskytují za sebou v určitých náhodných časových okamžicích. Předpokládáme přitom, že počet výskytů jevu $X(t)$ může nabývat jen nezáporné celočíselné hodnoty $i = 0, 1, 2, 3, \dots$ a jeho přírůstky $X(t_2) - X(t_1)$ pro libovolné $t_1, t_2, (t_1 < t_2)$ mohou také nabývat pouze nezáporných celočíselných hodnot.

Poissonův proces je charakteristický těmito vlastnostmi [13]:

- Proces $\{X(t)\}$ má nezávislé přírůstky – jevy, které se vyskytnou v nepřekrývajících se časových okamžicích, jsou navzájem nezávislé náhodné veličiny. Tedy počet jevů připadajících na určitý interval nezávisí na počtu jevů v jakémkoliv jiném intervalu.
- Proces $\{X(t)\}$ má homogenní přírůstky – intenzita vyskytujících se jevů, tj. střední hodnota počtu těchto jevů za časovou jednotku (značíme λ) je konstantní. Tato vlastnost se označuje jako stacionarita.
- Pro Δt dostatečně malé a při neměnné hodnotě λ jsou pravděpodobnosti přechodu ze stavu n do stavu $n + 1$ během intervalu $(t, t + \Delta t)$ rovny:

$$p_{n,n+1}(t, t + \Delta t) = \lambda \cdot \Delta t + o(\Delta t), \quad (3.15)$$

kde $o(\Delta t)$ představuje výrazy $(\Delta t)^2$ a vyšších řádů, pro které platí $\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0$.

Pro pravděpodobnosti setrvání ve stejném stavu během časového intervalu $(t, t + \Delta t)$ platí

$$p_{n,n}(t, t + \Delta t) = 1 - \lambda \cdot \Delta t + o(\Delta t). \quad (3.16)$$

Pravděpodobnosti jiných přechodů jsou v porovnání s předchozími zanedbatelné, tedy s vyjádřením

$$\sum_{j=i+2}^{\infty} p_{i,j}(t, t + \Delta t) = o(\Delta t). \quad (3.17)$$

Z přijatých předpokladů vyplývá, že matice intenzit přechodu Poissonova procesu má tvar:

$$\mathbf{A} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ 0 & -\lambda & \lambda & 0 & \dots \\ 0 & 0 & -\lambda & \lambda & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (3.18)$$

Pro pravděpodobnosti přechodu p_n tj. pravděpodobnosti, že systém bude v období $(t, t + \Delta t)$ ve stavu n , budou dány vztahy:

$$p_0(t + \Delta t) = [1 - \lambda \cdot \Delta t + o(\Delta t)] \cdot p_0(t), \text{ pro } n = 0 \quad (3.19)$$

$$p_n(t + \Delta t) = [1 - \lambda \cdot \Delta t + o(\Delta t)] \cdot p_n(t) + [\lambda \cdot \Delta t + o(\Delta t)] \cdot p_{n-1}(t) + o(\Delta t),$$

pro $n > 0$. (3.20)

Z vlastností Poissonova procesu plynou počáteční podmínky:

- během intervalu délky 0 nepříjde žádný požadavek (jev jistý), tedy: $p_n(0) = 1$, pro $n = 0$
- během nekonečně dlouhého intervalu nedojde k příchodu žádného požadavku (jev nemožný): $p_n(0) = 0$ pro $n > 0$.

Ze vztahů (3.19) a (3.20) pak dostáváme diferenciální rovnice:

$$\frac{p_0(t)}{dt} = -\lambda \cdot p_0(t) \text{ a} \quad (3.21)$$

$$\frac{p_n(t)}{dt} = -\lambda \cdot p_n(t) + \lambda \cdot p_{n-1}(t), \quad (3.22)$$

jejichž řešením dostáváme

$$p_n(t) = \frac{(\lambda \cdot t)^n}{n!} e^{-\lambda t}, \quad (3.23)$$

což je tvar hustoty Poissonova rozdělení ve vztahu k počtu změn za časový interval t . Člen $p_0(t) = e^{-\lambda t}$ udává pravděpodobnost toho, že za interval délky t nedojde k žádné změně. Je-li rozdělení počtu změn systému za určitou dobu Poissonovo, pak je pro tentýž proces rozdělení dob mezi změnami exponenciální.

Matice pravděpodobností přechodu Poissonova procesu má tvar

$$\mathbf{P} = \begin{pmatrix} 1 - \lambda \cdot \Delta t & \lambda \cdot \Delta t & 0 & 0 & \dots \\ 0 & 1 - \lambda \cdot \Delta t & \lambda \cdot \Delta t & 0 & \dots \\ 0 & 0 & 1 - \lambda \cdot \Delta t & \lambda \cdot \Delta t & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (3.24)$$

3.2.2 Procesy množení a zániku

Speciální proces, kterým lze efektivně analyzovat značné množství problémů řešených v modelech hromadné obsluhy, je proces množení a zániku. Stavby procesu představují

aktuální velikost populace a existují u nich pouze dva druhy přechodů: narození představuje příchod požadavku, zánik pak představuje opuštění systému obsluhy.

Zavedeme předpoklad, že narození a zánik v populaci jsou nezávislé jevy. Pravděpodobnost, že v k -členné populaci dojde během intervalu $(t, t + \Delta t)$ k jednomu narození je $\lambda_k \cdot \Delta t + o(\Delta t)$, tedy:

$$P\{1 \text{ narození během doby } (t, t + \Delta t), \text{ pokud populace má } k \text{ členů}\} = \lambda_k \cdot \Delta t + o(\Delta t).$$

Z prvního tvrzení pak můžeme vyvodit:

$$P\{0 \text{ narození během doby } (t, t + \Delta t), \text{ pokud populace má } k \text{ členů}\} = 1 - \lambda_k \cdot \Delta t + o(\Delta t),$$

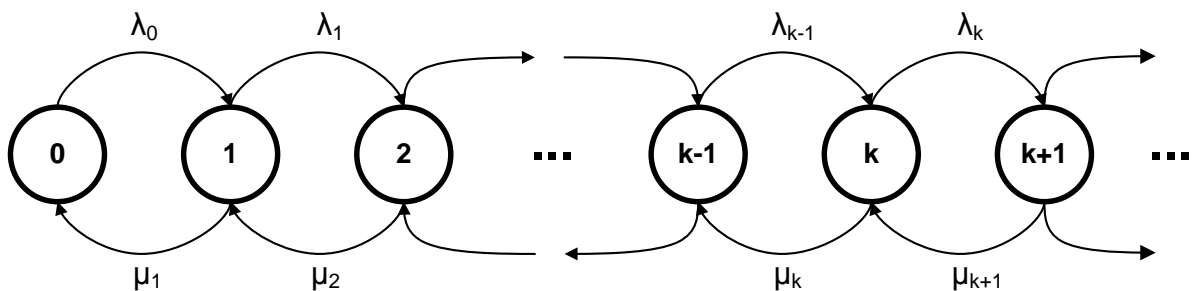
$$P\{1 \text{ zániku během doby } (t, t + \Delta t), \text{ pokud populace má } k \text{ členů}\} = \mu_k \cdot \Delta t + o(\Delta t),$$

$$P\{0 \text{ zániků během doby } (t, t + \Delta t), \text{ pokud populace má } k \text{ členů}\} = 1 - \mu_k \cdot \Delta t + o(\Delta t),$$

μ_k představuje intenzitu úmrtí. Protože v populaci, která má 0 členů, nemůže dojít k úmrtí, položíme $\mu_0 = 0$. Na druhé straně ale předpokládáme, že v nulové populaci může dojít ke zrození, takže $\lambda_0 \geq 0$. Jak vyplývá z předchozího zápisu, pravděpodobnosti narození a zániku více jedinců populace jsou během intervalu $(t, t + \Delta t)$ při $\Delta t \rightarrow 0$ zanedbatelné, protože jsou zahrnuty do parametru $o(\Delta t)$.

Pokud označíme $p_k(t)$ jako nepodmíněnou pravděpodobnost, že populace má v okamžiku t právě k členů $k = 0, 1, 2, 3, \dots$ pak musí platit vztah $\sum_{k=0}^{\infty} p_k(t) = 1$.

Při odvození vztahů pro $p_k(t + \Delta t)$ budeme vycházet z následujících úvah (pro názornost viz obr. 2):



Obr. 2: Proces množení a zániku

- a) Populace bude mít v okamžiku $t + \Delta t$ velikost $k, k \geq 1$ právě tehdy, když nastane jeden z těchto disjunktních jevů:
 - populace má v okamžiku t velikost k a během Δt nedojde k žádné změně.

- populace má v okamžiku t velikost k a během Δt dojde k jednomu narození a jednomu zániku.
- populace má v okamžiku t velikost $k - 1$ a během Δt dojde k jednomu narození a žádnému zániku.
- populace má v okamžiku t velikost $k + 1$ a během Δt nedojde k žádnému narození a k jednomu zániku.

b) Pro případ $k = 0$ budeme uvažovat dvě možnosti:

- populace má v okamžiku t velikost 0 a během Δt nedojde k žádnému narození.
- populace má v okamžiku t velikost 1 a během Δt nedojde k žádnému narození, ale dojde k jednomu zániku.

Matice intenzit přechodu \mathbf{A} je

$$\mathbf{A} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \dots \\ \mu_1 & -\lambda_1 - \mu_1 & \lambda_1 & 0 & \dots \\ 0 & \mu_2 & -\lambda_2 - \mu_2 & \lambda_2 & \dots \\ 0 & 0 & \mu_3 & -\lambda_3 - \mu_3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (3.25)$$

Matice pravděpodobností přechodu \mathbf{P} takto bude mít tvar:

$$\mathbf{P} = \begin{pmatrix} 1 - \lambda_0 \cdot \Delta t & \lambda_0 \cdot \Delta t & 0 & 0 & \dots \\ \mu_1 \cdot \Delta t & 1 - (\lambda_1 + \mu_1) \cdot \Delta t & \lambda_1 \cdot \Delta t & 0 & \dots \\ 0 & \mu_2 \cdot \Delta t & 1 - (\lambda_2 + \mu_2) \cdot \Delta t & \lambda_2 \cdot \Delta t & \dots \\ 0 & 0 & \mu_3 \cdot \Delta t & 1 - (\lambda_3 + \mu_3) \cdot \Delta t & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (3.26)$$

Pak lze vyjádřit hodnoty $p_k(t + \Delta t)$ a $p_0(t + \Delta t)$ ve formě součtů pravděpodobností příslušných jevů takto:

$$p_k(t + \Delta t) = [1 - \lambda_k \cdot \Delta t - \mu_k \cdot \Delta t] \cdot p_k(t) + [\lambda_{k-1} \cdot \Delta t] \cdot p_{k-1}(t) + [\mu_{k+1} \cdot \Delta t] \cdot p_{k+1}(t), \text{ pro } k \geq 1 \text{ a} \quad (3.27)$$

$$p_0(t + \Delta t) = [1 - \lambda_0 \cdot \Delta t] \cdot p_0(t) + [\mu_1 \cdot \Delta t] \cdot p_1(t), \text{ pro } k = 0. \quad (3.28)$$

Řešením soustavy rovnic (3.27) a (3.28) pak získáme obecný tvar (viz lit. [13]):

$$p_k = \frac{\lambda_{k-1} \cdot \lambda_{k-2} \cdot \dots \cdot \lambda_2 \cdot \lambda_1 \cdot \lambda_0}{\mu_k \cdot \mu_{k-1} \cdot \dots \cdot \mu_3 \cdot \mu_2 \cdot \mu_1} \cdot p_0. \quad (3.29)$$

Kromě vztahu (3.29) platí také výše zmíněný vztah $\sum_{k=1}^{\infty} p_k = 1$ a pravděpodobnost p_0 pak můžeme vyjádřit

$$p_0 = \left[1 + \sum_{k=1}^{\infty} \frac{\lambda_{k-1} \cdot \lambda_{k-2} \cdot \dots \cdot \lambda_2 \cdot \lambda_1 \cdot \lambda_0}{\mu_k \cdot \mu_{k-1} \cdot \dots \cdot \mu_3 \cdot \mu_2 \cdot \mu_1} \right]^{-1}. \quad (3.30)$$

4 MODELOVÁNÍ SÍŤOVÝCH PRVKŮ A SÍŤOVÉHO PROVOZU

Jelikož cílem této práce je určit zpoždění, proměnlivost a ztrátu dat v síti, je potřeba vytvořit patřičné modely jednotlivých síťových prvků a služeb, s jejichž pomocí budou provedeny simulace k získání zmíněných dat.

Zkoumaný provoz bude rozdělen na dva typy – obecný TCP provoz, který bude představovat best-effort službu a VoIP službu, jejíž kvalita závisí na zpoždění a pro jejíž kvalitní provoz je potřeba využít prioritizace paketů.

Tradiční internetové aplikace, jako třeba FTP využívají TCP jako transportní protokol, jsou v případě zahlcení sítě závislé na ochraně proti zahlcení tohoto protokolu. V TCP odesílatel postupně zvyšuje rychlost odesílání paketů, dokud nezačne zaznamenávat jejich ztrátu, což se pokládá za indikaci zahlcení sítě. Odesílatel pak sníží rychlost odesílání, čímž rychle zareaguje na vzniklou kolizi. Z tohoto chování vyplývá, že intenzita odesílání TCP paketů je určena stupněm zahlcení v síti. Proměnlivost zpoždění ani jeho velikost pro aplikace typu FTP a částečně i HTTP nehraje výraznou roli, klíčové je doručení všech paketů (ztracené pakety musí být znovu odeslány) a správnost dat. Naproti tomu UDP/RTP přenosy, které pro svou funkci využívá například VoIP nevyžadují 100% přenos všech paketů k cíli, zato jsou náchylné na zpoždění i jeho proměnlivost (kolísání zpoždění je způsobeno právě rozdílnou dobou čekání ve frontách v jednotlivých mezilehlých zařízeních, a to se musí v přijímači odstraňovat bufferováním). Vzhledem k tomu, že z principu tohoto způsobu přenosu nejde uplatnit výše popsany mechanismus pro řízení intenzity odesílání dat, jsou data v UDP odesílána s konstantní intenzitou a k zajištění jejich správné funkčnosti je potřeba zajistit prioritu paketů závislých na zpoždění na jednotlivých mezilehlých prvcích sítě pomocí QoS.

Analýza moderních telekomunikačních systémů je velmi komplexní záležitostí, jelikož většina standardních postupů pro modelování analyzuje každý prvek zvlášť a ne vždy bere v potaz vazby, které mezi jednotlivými komponentami systému existují. Simulace je nástroj, který se dnes běžně používá ke stanovení předpokládaného chování a výkonnosti velkých komplexních stochastických systémů [10]. Vytvoření přesného simulačního modelu však vyžaduje rozsáhlé zdroje. Pokud model není přesný, může se stát, že na základě výsledků simulace jsou vyvozeny chybné závěry. Základním problémem je, že každý model je už v základu do určité míry chybný. V důsledku toho je výsledek simulace vždy pouze tak

dobrý, jak dobře je navržený model simulovaného systému. Stále ovšem půjde pouze o pravděpodobný odhad výsledku.

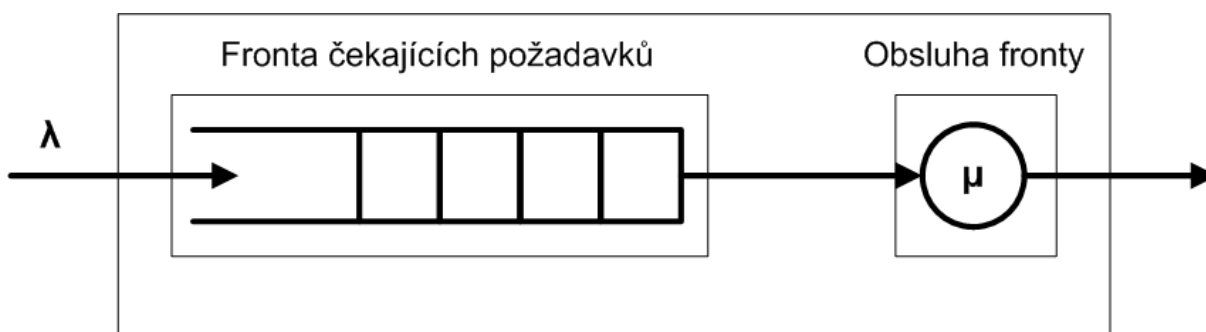
Simulační program, vytvořený pro účely této práce využívá dva druhy modelů – konkrétně $M/M/1$ a $M/G/1$ dle Kendallovy klasifikace. U obou druhů modelů je možné nastavit dvě třídy provozu a lze tedy simulovat aplikaci priorit při zpracovávání paketů (QoS) v uzlech sítě. U obou druhů modelů je také možno nastavit konečnou délku fronty (označení $M/M/1/K$ a $M/M/1/K$ dle Kendallovy klasifikace) a tím pádem monitorovat ztrátovost paketů.

Většinu teoretických poznatků, uvedených u popisu modelů $M/M/1$, lze aplikovat i na modely $M/G/1$. U modelů $M/G/1$ jsou proto popsány především odlišnosti od modelu $M/M/1$, aby se předešlo opakování stejných informací.

4.1 Model $M/M/1$

Systém $M/M/1$ je základní a nejdůležitější model systému front, který je využíván i jako porovnávací případ pro ostatní modely. Z pohledu složitosti analýzy se jedná o nejjednodušší a nejobecnější model hromadné obsluhy, kde má rozdělení dob mezi příchody a dob obsluhy charakter exponenciálního rozdělení a jde o otevřený systém s jedním obslužným zařízením. Režim fronty u tohoto modelu je jednoduchý a v případě nedostatečné kapacity obslužného zařízení požadavky trpělivě vyčkávají ve frontě na obsluhu, kam vstupují v pořadí, v jakém přišly (FIFO).

Průměrný počet požadavků, které vstupují do systému za určitý časový interval (tj. intenzita příchodů) je označen λ a průměrný počet požadavků obslužených za časový interval (intenzita obsluhy) je označen μ .



Obr. 3: Schéma modelu $M/M/1$

U analýzy vstupního toku nás bude zajímat především rozdělení pravděpodobnosti pro interval mezi dvěma po sobě jdoucími příchody. Náhodnou veličinou je počet jednotek, které do systému vstoupily – což je diskrétní veličina. Vstupní tok tedy popisuje počet jednotek, které v určitém časovém intervalu vstoupí do systému a na základě předpokladů dle [14] je možno pravděpodobnost vstupu n jednotek v intervalu $T = (0, t)$ vyjádřit ve tvaru

$$p_n(T) = \frac{(\lambda T)^n}{n!} \cdot e^{-\lambda T}. \quad (4.1)$$

Druhou veličinou, kterou se u analýzy vstupního toku budeme zabývat je rozdělení pravděpodobnosti pro interval mezi dvěma po sobě jdoucími příchody. Tento interval představuje spojitou veličinu. Odpovídá-li rozdělení dob mezi příchody t exponenciálnímu rozdělení, pak distribuční funkci a hustotu rozdělení popisují vztahy

$$F(t) = 1 - e^{-\lambda t}, \quad (4.2)$$

$$f(t) = \lambda e^{-\lambda t}. \quad (4.3)$$

Dále lze předpokládat, že stav systému v libovolném časovém okamžiku t , bude jednoznačně určen číslem n , které udává počet jednotek v systému a nezávisí kromě předcházejícího stavu na žádném jiném předešlém stavu – tedy proces hromadné obsluhy má charakter Markovova procesu.

Pak lze uvažovat pouze možnost setrvání ve stavu nebo přechod mezi sousedními stavy. To je důsledek toho, že pravděpodobnosti přechodů (za předpokladu exponenciálního rozdělení) mezi nesousedními stavy budou nekonečně malé, neboť pravděpodobnosti vstupu (popř. obsluhy) více než jedné jednotky v časovém intervalu Δt jsou zanedbatelné.

Chování systému odpovídá procesu množení a zániku popsaného v kapitole 3.2.2 s tím rozdílem, že pro systém $M/M/1$ je intenzita narození i intenzita zániku nezávislá na počtu jednotek v systému, tj. $\lambda_k = \lambda$ a $\mu_k = \mu$ pro všechna k . Soustavu pravděpodobností přechodů mezi jednotlivými stavy za dobu Δt pak lze vyjádřit přehledně pomocí matice přechodu \mathbf{P} , která má tvar viz matice (3.26). Intenzitu pravděpodobností přechodu tedy udávají veličiny λ a μ . Na úhlopříčce matice jsou výrazy, odpovídající pravděpodobnostem setrvání ve stejném stavu, vpravo od ní jsou pravděpodobnosti příchodu požadavku a vlevo pravděpodobnosti ukončení obsluhy požadavku. Ostatní prvky jsou nulové, protože prakticky neexistuje jiná možnost přechodu mezi nesousedními stavy.

Vzhledem k tomu, že další analýza takového systému by vyžadovala řešení soustavy lineárních homogenních diferenciálních rovnic (tzv. Erlangovy soustavy), budeme zkoumat, zda je uvažovaný systém hromadné obsluhy stabilní, a tedy nezávislý na čase t a vstupních podmínkách. Výše jsme došli k závěru, že proces je Markovským procesem a aby byl stabilní, musí být splněna podmínka stability $\rho < 1$, kde ρ označuje intenzitu provozu (provozní zatížení), kterou lze při znalosti intenzity vstupního toku a intenzity obsluhy definovat jako $\rho = \frac{\lambda}{\mu}$. Z uvedeného vztahu plyne, že pro stabilitu musí platit $\lambda < \mu$, tedy obsluha musí být rychlejší než příchody [19]. Z hlediska zákazníka systém bude fungovat lépe, když intenzita provozu bude menší, protože se nebude tvořit tak dlouhá fronta a tak i doba čekání bude kratší.

Známe-li rozdělení stacionárních pravděpodobností vyjadřujících pravděpodobnosti počtu jednotek nacházejících se v systému hromadné obsluhy, můžeme určit základní charakteristiky používané k posouzení efektivnosti obslužných zařízení. Jedná se o (odvození vzorců viz lit. [13]):

- Průměrný počet požadavků v systému:

$$E(L) = \frac{\lambda}{\mu - \lambda} \quad (4.4)$$

- Průměrný počet požadavků ve frontě:

$$E(L_q) = \frac{\lambda^2}{\mu(\mu - \lambda)} \quad (4.5)$$

- Průměrná doba strávená v systému:

$$E(W) = \frac{1}{\mu - \lambda} = \frac{1}{\mu(1 - \rho)} \quad (4.6)$$

- Průměrná doba čekání ve frontě:

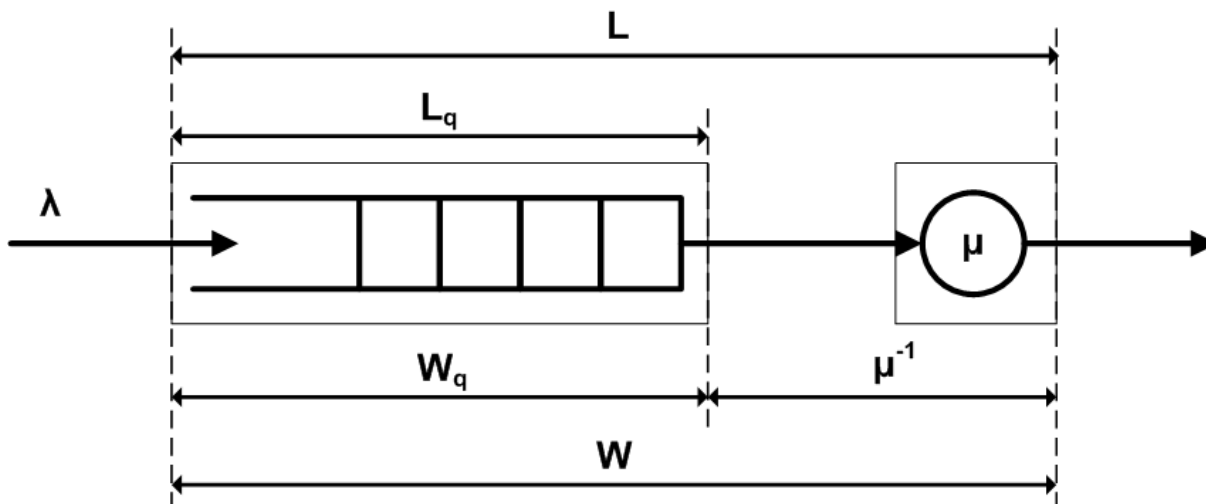
$$E(W_q) = \frac{\lambda}{\mu(\mu - \lambda)} \quad (4.7)$$

Mezi těmito charakteristikami platí tyto dva vztahy:

$$E(L) = E(L_q) + \rho \quad (4.8)$$

$$E(W) = E(W_q) + E(S) \quad (4.9)$$

Výše uvedené charakteristiky SHO znázorňuje obr. 4.



Obr. 4: Charakteristiky modelu $M/M/1$

Vztah (4.6) byl odvozen ze vztahu (4.4) na základě rovnice

$$E(L) = \lambda \cdot E(W), \quad (4.10)$$

známé pod názvem Littleova formule. Obecné znění Littleova teorému říká, že očekávaný počet požadavků v systému se rovná součinu intenzity vstupního toku a očekávaného času, který požadavky v systému stráví. Littleův teorém neplatí jen pro model $M/M/1$, ale i pro většinu ostatních komplexních modelů SHO [9].

4.1.1 Model $M/M/1$ s prioritami (2 prioritní třídy)

Prioritní mechanismy jsou do síťových uzlů implementovány za účelem minimalizace vlivu zatížení sítě na zpoždění paketů multimediálních služeb. Tyto mechanismy tedy zmenšují zpoždění paketů, které vyžadují přednostní obsluhu. Při implementaci priorit je nutné rozdělit provoz v síti do tzv. prioritních tříd (neboli tříd provozu). Rozdělení se provádí s ohledem na citlivost služby na zpoždění. Jednotlivým třídám je přiřazeno číslo, které označuje přednost při obsluze. Platí konvence, že paket z třídy s nižším číslem je obslužen dříve než paket z třídy s vyšším číslem (tedy paket s nižším prioritním číslem má vyšší prioritu).

V modelu s prioritami uvažujeme, že příchody požadavků obou priorit do systému odpovídají Poissonovu rozložení (navzájem nezávislých), intenzity příchodů označíme jako λ_1 a λ_2 , doba obsluhy odpovídá exponenciálnímu rozložení s intenzitou μ a je tedy stejná pro

všechny požadavky. Aby byla splněna podmínka stability, uvažujeme $\rho_1 + \rho_2 < 1$, kde ρ_1 představuje provozní zatížení tvořené požadavky první prioritní třídy a ρ_2 je provozní zatížení tvořené požadavky druhé prioritní třídy.

Prioritní mechanismy je možné rozdělit podle způsobu, jakým nakládají s pakety, které jsou právě obsluhované. Rozlišujeme preemptivní prioritu a nepreemptivní prioritu.

Preemptivní priorita zajišťuje, že pokud přijde do systému s obsazenou linkou obsluhy paket, který má vyšší prioritu než momentálně obsluhovaný paket, je obsluha paketu s nižší prioritou přerušena nebo ukončena (preemptivní priorita lze dále dělit na mechanismy s dokončením obsluhy a bez dokončení obsluhy). Přijde-li do systému s obsazenou obsluhovou linkou paket se stejnou nebo nižší prioritou než má paket v obsluze, pak je příchozí paket zařazen do fronty a jeho další zpracování se řídí prioritou a režimem fronty.

Nepreemptivní priorita zajišťuje, že příchozí pakety jsou v případě obsazené linky obsluhy řazeny do fronty bez ohledu na prioritu paketu či paketů, které jsou momentálně obsluhovány. Po uvolnění linky obsluhy je následně obslužen paket s nejvyšší prioritou, který je podle režimu fronty další v pořadí (každá prioritní třída má vlastní logickou frontu a jak se uvolní linka obsluhy, paket na začátku fronty s nejvyšší prioritou vstoupí do obsluhové linky). V dnešních zařízeních, využívaných jako uzlové body sítě, se používá výhradně nepreemptivní priorita [18]. Uvedené vztahy platí právě pro tento typ priority. Střední dobu, kterou stráví požadavek 1. prioritní třídy v systému, pak určíme pomocí tohoto vztahu [15]:

$$E(W_1) = \frac{1 + \rho_2}{\mu(1 - \rho_1)}. \quad (4.11)$$

Střední doba, kterou stráví v systému požadavek 2. prioritní třídy je dána [15]

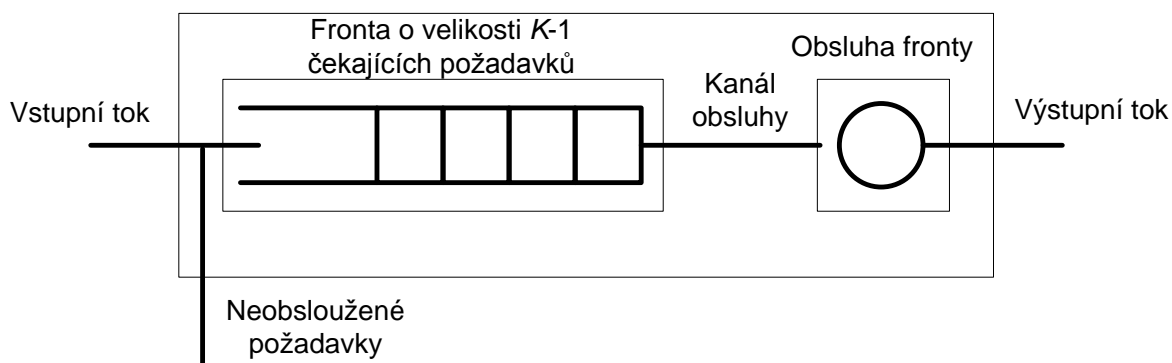
$$E(W_2) = \frac{(1 - \rho_1(1 - \rho_1 - \rho_2))}{\mu(1 - \rho_1)(1 - \rho_1 - \rho_2)}. \quad (4.12)$$

Analogicky by bylo možné odvodit vzorce pro více prioritních tříd. Hodnoty ostatních parametrů lze získat s využitím Littleovy formule.

4.1.2 Model M/M/1 s konečnou délkou fronty

Vzhledem k tomu, že požadavek nekonečné fronty lze u reálného systému jen obtížně splnit, model typu $M/M/1/K$ reálný systém vystihuje lépe. Dle Kendallovy notace K označuje

maximální počet požadavků, které se mohou nacházet v systému – tj. jeden požadavek je obsluhován a $K-1$ požadavků čeká ve frontě na obsluhu, viz obr. 5. Protože kapacita fronty je konečná, požadavky, které přijdou v době, kdy je systém plně obsazen, jsou ztraceny a opouští systém bez obslužení. O takových požadavcích říkáme, že byly zablokovány a takový systém je označován jako systém se ztrátami. Pravděpodobnost ztracených požadavků (označovaná jako pravděpodobnost blokování – P_B) je u systémů se ztrátami jedním z nejdůležitějších parametrů.



Obr. 5: Schéma modelu $M/G/1/K$

Pravděpodobnost, že požadavek nebude obslužen v důsledku plné fronty (tj. pravděpodobnost, že v systému je právě K požadavků), je dána vztahem (odvození viz lit. [13]):

$$P_B = p_K = \frac{(1 - \rho)\rho^K}{1 - \rho^{K+1}}. \quad (4.13)$$

4.2 Model $M/G/1$

Předchozí část popisovala systém, kde rozdělení dob mezi příchody odpovídá Poissonovu rozložení (což se dá považovat v mnoha případech za realistický model procesu příchodů) a rozdělení dob obsluhy má charakter exponenciálního rozdělení. V praxi není exponenciálně daná doba obsluhy příliš běžná. Ve většině systémů bude koeficient variace menší než 1 a proto je potřeba rozšířit základní teorii o jiné případy rozložení dob obsluhy. Nyní se tedy budeme zabývat modelem, kde příchody odpovídají Poissonovu rozdělení a mají obecné, i když nezávislé, rozložení doby obsluhy. Typ fronty je opět FIFO.

Charakteristiky modelu můžeme získat podobně jako u modelu $M/M/1$ pomocí středních hodnot. Doba mezi příchody je popsána totožně jako u předchozího modelu, doby obsluhy je pak popsána intenzitou obsluhy μ a koeficientem variace c_v . Pravděpodobnost přechodu mezi stavy nyní závisí také na době, kterou již požadavek v systému strávil. Nově přichodící požadavek musí nejprve čekat, a to, než bude dokončena obsluha právě zpracovávaného požadavku (pokud takový požadavek existuje) a dále pak čeká ve frontě, než budou zpracovány všechny požadavky před ním. Vlastnost PASTA (Poisson Arrivals See Time Averages) udává, že procento požadavků, které při příchodu naleznou systém ve stavu, kdy právě obsluhuje požadavek, je stejné, jako procento času, po který se systém ve stavu obsluhy požadavku nachází – systém tedy při příchodu požadavku právě vykonává obsluhu s pravděpodobností ρ [23]. Označíme-li průměrnou dobu obsluhy jako $E(S)$, $E(R)$ bude označovat zbývající dobu obsluhy požadavku a $E(L_q)$ značí střední počet požadavků ve frontě, dostáváme:

$$E(W_q) = E(L_q)E(S) + \rho E(R). \quad (4.14)$$

Aplikací Littleovy formule: $E(L_q) = \lambda \cdot E(W_q)$ a vztahu $\rho = \lambda \cdot E(S)$ dostáváme:

$$E(W_q) = \frac{\rho E(R)}{1 - \rho}. \quad (4.15)$$

Rovnice (4.15) se často označuje jako Pollaczkův-Khintchineův vzorec. Pro průměrnou zbývající dobu obsluhy požadavku platí vztah [9,22]

$$E(R) = \lambda \cdot \frac{1}{2} E(S^2), \quad (4.16)$$

a pro střední dobu obsluhy platí

$$E(S^2) = (1 + c_v^2) \cdot E(S)^2. \quad (4.17)$$

S využitím (4.15), (4.16) a (4.17) lze rovnici (4.14) pro průměrnou dobu čekání ve frontě zapsat jako

$$E(W_q) = \frac{1 + c_v^2}{2} \cdot \frac{\rho}{1 - \rho} \cdot E(S). \quad (4.18)$$

Další vztahy jsou pak následující [9,22]:

- Průměrná doba strávená v systému

$$E(W) = \left(1 + \frac{1 + c_v^2}{2} \cdot \frac{\rho}{1 - \rho}\right) \cdot E(S) \quad (4.19)$$

- Průměrný počet požadavků v systému

$$E(L) = \rho + \frac{1 + c_v^2}{2} \cdot \frac{\rho^2}{1 - \rho} \quad (4.20)$$

- Průměrný počet požadavků ve frontě

$$E(L_q) = \frac{1 + c_v^2}{2} \cdot \frac{\rho^2}{1 - \rho} \quad (4.21)$$

4.2.1 Model M/G/1 s prioritami (2 prioritní třídy)

Pro určení střední doby strávené v systému s prioritami vyjdeme ze vztahu (4.14) a aplikace Littleovy formule. Střední doba, kterou stráví v systému požadavek 1. prioritní třídy je pak dána vztahem [15]

$$E(W_1) = \frac{\rho_1 \cdot E(R_1) + \rho_2 \cdot E(R_2)}{1 - \rho_1} + E(S_1), \quad (4.22)$$

a střední dobu, kterou stráví v systému požadavek 2. prioritní třídy určuje vztah [15]

$$E(W_2) = \frac{\rho_1 \cdot E(R_1) + \rho_2 \cdot E(R_2)}{(1 - (\rho_1 + \rho_2))(1 - \rho_1)} + E(S_2). \quad (4.23)$$

Uvedený vzorec je opět možné zobecnit pro více prioritních tříd.

4.2.2 M/G/1/K s konečnou délkou fronty

Obdobně jako u systému $M/M/1/K$ základní charakteristikou, která nás bude zajímat je pravděpodobnost blokování (ztrát), daná vztahem [21]:

$$P_B = \frac{\rho^K}{K! \sum_{k=0}^K \frac{\rho^k}{k!}}, \quad (4.24)$$

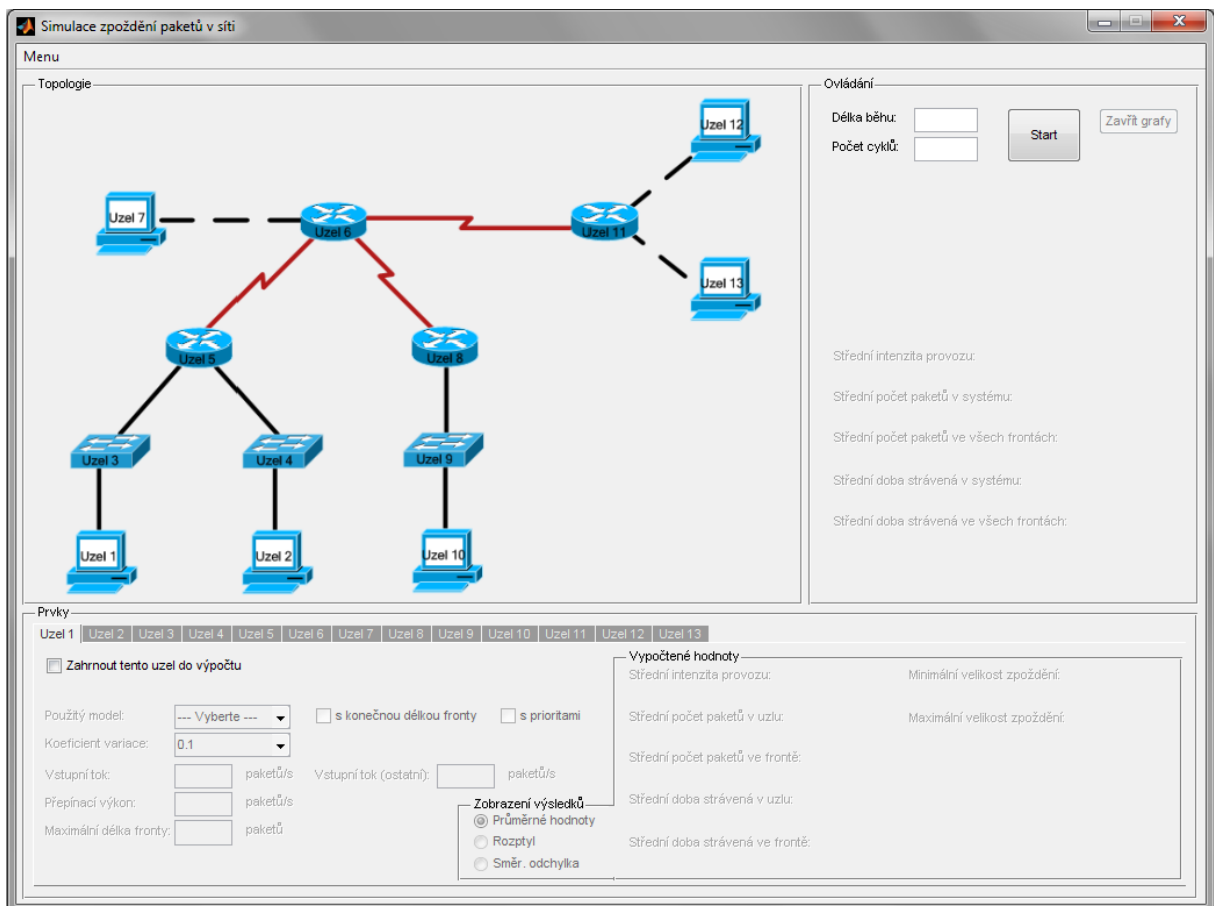
kde k představuje aktuální počet požadavků ve frontě.

5 PROGRAMOVÁ REALIZACE

5.1 Popis programu

5.1.1 Úvodní popis

Tento program využívá metod a poznatků z teorie hromadné obsluhy k modelování síťových prvků. Jednotlivé uzly sítě vycházejí ze dvou modelů dle Kendallovy klasifikace obsluhových systémů a jsou doplněny o několik dalších nastavení. Modelovaná síť má pevně danou topologii, uživatel si může zvolit modelování konkrétního uzlu či „cesty“ přes více uzlů. Výsledkem simulace je množina parametrů charakterizující daný uzel a také grafy zobrazující průběh zpoždění paketů, proměnlivost tohoto zpoždění a úroveň zaplnění front (vyrovnávacích pamětí) uzlů.



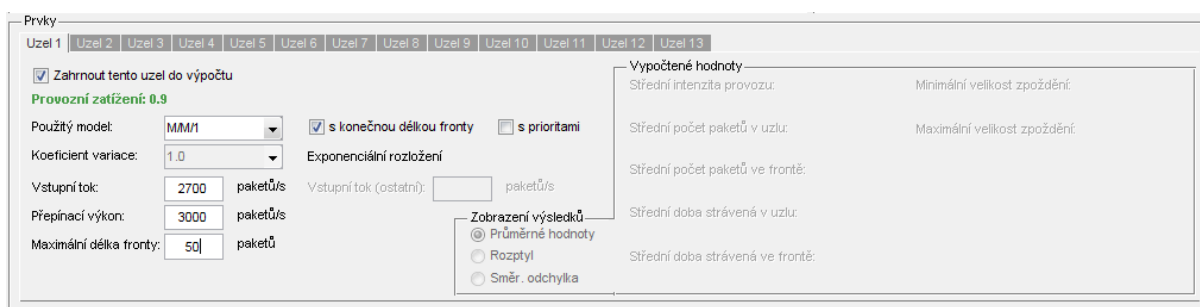
Obr. 6: Hlavní okno programu

Hlavní okno programu, zobrazené na obr. 6, se skládá ze tří hlavních částí:

- v levé horní části je zobrazena topologie sítě umožňující přesnou orientaci při určování uzlů, které budou simulovány. Tato topologie je pevně daná a neměnná.
- Spodní část hlavního okna programu zabírá box s kartami jednotlivých uzlů sítě a spojovacích linek. V této části se nastavují vstupní parametry pro jednotlivé uzly a zároveň jsou zde také zobrazeny výsledky simulace pro konkrétní uzly.
- Poslední částí je panel ovládání, který se nachází v pravé horní části. Obsahuje jednak tlačítka pro ovládání programu a také zobrazuje celkové výsledky simulace ze všech prvků, které byly do simulace zahrnuty.

5.1.2 Práce s programem

Po spuštění programu uživatel nejprve s pomocí obrázku topologie určí, které uzly budou zahrnuty do simulace. Následně těmto uzlům nastaví potřebné vstupní parametry.



Obr. 7: Nastavení vstupních parametrů uzlu

Každý uzel, který má být součástí simulace, je nejprve nutné zahrnout do simulace pomocí stejnojmenného zaškrtačacího políčka v levé polovině záložky každého uzlu, viz obr. 7. Tím se aktivuje další položka pro výběr modelu, pomocí kterého bude daný uzel simulován. K dispozici jsou dva modely – $M/M/1$ a $M/G/1$. U modelu $M/G/1$ může uživatel zvolit doplňující charakteristiku, tzv. koeficient variace, který určuje, jaké rozložení bude mít doba obsluhy paketu. Spolu s modelem $M/M/1$ jsou tak k dispozici tři druhy rozložení doby obsluhy: Erlangovo, exponenciální a hyperexponenciální. Rozložení příchodu paketů má u obou modelů shodně exponenciální charakter.

Oba typy modelu mají dva doplňující parametry: prvním je možnost zvolit konečnou délku fronty (vyrovnávací paměti), tj. při dosažení stanoveného počtu paketů ve frontě budou případné další přicházející pakety zahazovány. Druhým doplňujícím parametrem je režim priorit – v tom případě je vstupní tok rozdělen na dva toky a jeden z nich je upřednostňován před druhým. Je zde uplatněna časová nepreemptivní priorita. Vstupní parametry uživatel

zadáva do příslušných polí. Vstupní tok představuje počet paketů, které do uzlu přijdou během jedné vteřiny. Přepínací výkon udává počet paketů, které je uzel schopen během jedné vteřiny obsloužit. Orientační tabulku přepínacích výkonů některých reálných síťových prvků lze zobrazit pomocí stejnojmenné volby v menu programu. Obecně platí, že vstupní tok(y) musí být minimálně 2 pakety. Přepínací výkon musí být větší jak vstupní tok (či součet prioritního a ostatního vstupního toku) – to proto, aby byla zachována podmínka stability. Minimální délka fronty uzlu je 1 paket.

Pravá polovina záložky každého uzlu obsahuje výstupní parametry každého uzlu. Tyto hodnoty se zobrazí po provedení simulace. Uživatel má možnost pomocí přepínače zobrazit buď průměrné hodnoty, jejich rozptyl nebo směrodatnou odchylku. Pro každý uzel jsou sledovány následující hodnoty:

- **Intenzita provozu** – podíl počtu čekajících paketů a celkového počtu paketů
- **Počet paketů v uzlu** – počet paketů, které se nachází v zařízení (čekají ve frontě na obsluhu a jsou právě obsluhovány)
- **Počet paketů ve frontě** – počet paketů, které čekají ve frontě na obsluhu
- **Doba strávená v uzlu** – doba paketu v uzlu (od příchodu paketu až po jeho obslužení a opuštění uzlu)
- **Doba strávená ve frontě** – doba paketu ve frontě – od příchodu paketu po začátek obsluhy paketu
- **Minimální velikost zpoždění** – udává nejmenší dobu, kterou některý z odeslaných paketů stráví v uzlu
- **Maximální velikost zpoždění** – udává nejdelší dobu, kterou některý z odeslaných paketů stráví v uzlu

Vypočtené hodnoty	
Střední intenzita provozu: 0.74717	Minimální velikost zpoždění: 0.002839
Střední počet paketů v uzlu: 2.5073	Maximální velikost zpoždění: 0.012585
Střední počet paketů ve frontě: 1.7682	Počet zahozených paketů v uzlu: 4.18
Střední doba strávená v uzlu: 0.0097584	Ztrátovost paketů v uzlu: 1.3933 %
Střední doba strávená ve frontě: 0.0072294	

Obr. 8: Vypočtené hodnoty (bez priorit)

Pokud uživatel zvolil model s konečnou délkou fronty, je navíc zobrazen také **počet zahozených paketů v uzlu** a **ztrátovost paketů** (poměr zahozených paketů a celkem vyslaných paketů) viz obr. 8.

Vypočtené hodnoty	
Střední intenzita provozu: 0.87429	Minimální velikost zpoždění (PRI,OST,BEZ):* 0.0025015 0.012621 0.0024491
Střední počet paketů v uzlu: 5.3667	Maximální velikost zpoždění (PRI,OST,BEZ):* 0.013726 0.071118 0.021521
Střední počet paketů ve frontě: 4.4851	
Střední doba strávená v uzlu (PRI,OST,BEZ):* 0.010704 0.048794 0.016413	
Střední doba strávená ve frontě (PRI,OST,BEZ):* 0.0082189 0.046309 0.01393	* PRI = prioritní tok, OST = ostatní tok, BEZ = tok bez aplikace priorit

Obr. 9: Vypočtené hodnoty (s prioritami)

Pokud uživatel zvolil model s prioritami, jsou u položek doby strávené v uzlu, doby strávené ve frontě, minimální velikost zpoždění a maximální velikost zpoždění uvedeny hodnoty pro prioritní tok – označené zeleně, pro ostatní tok – označené červeně a pro porovnání i hodnoty pro případ, že by nebyl použit model s prioritami – modře označené hodnoty. Tato situace je znázorněna na obr. 9.

Před samotným spuštěním simulace je potřeba zadat délku simulace (zadáva se v sekundách) a počet cyklů – tj. počet opakování náhodného pokusu. Počet opakování má vliv na dobu běhu simulace a v některých případech může být omezen velikostí dostupné operační paměti počítače, na kterém je program provozován. Velmi vysoký počet opakování také může vést k překročení maximální velikostí polí, se kterými dokáže program zpracovat (závisí

na systému, na kterém je program provozován a na velikosti operační paměti). Na tyto skutečnosti je uživatel před zahájením simulace upozorněn.

Tlačítko start slouží pro začátek simulace. Dále se zde nachází tlačítko pro skrytí výsledných grafů.

Po provedení simulace jsou zobrazeny hodnoty výstupních parametrů pro celý systém, viz obr. 10, jejich popis odpovídá popisu parametrů pro jednotlivé uzly uvedenému výše.

Ovládání

Délka běhu: Start

Počet cyklů:

Střední intenzita provozu:
0.9137

Střední počet paketů v systému:
7.7927

Střední počet paketů ve všech frontách:
6.8743

Střední doba strávená v systému (PRI,OST,BEZ):*
0.0077769 0.048711 0.032515

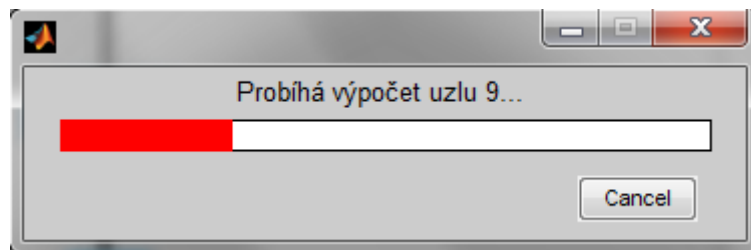
Střední doba strávená ve všech frontách (PRI,OST,BEZ):*
0.0044413 0.045376 0.029177

* PRI = prioritní tok, OST = ostatní tok,
BEZ = tok bez aplikace priorit

Obr. 10: Panel ovládání a celkových výsledků

5.1.3 Spuštění simulace

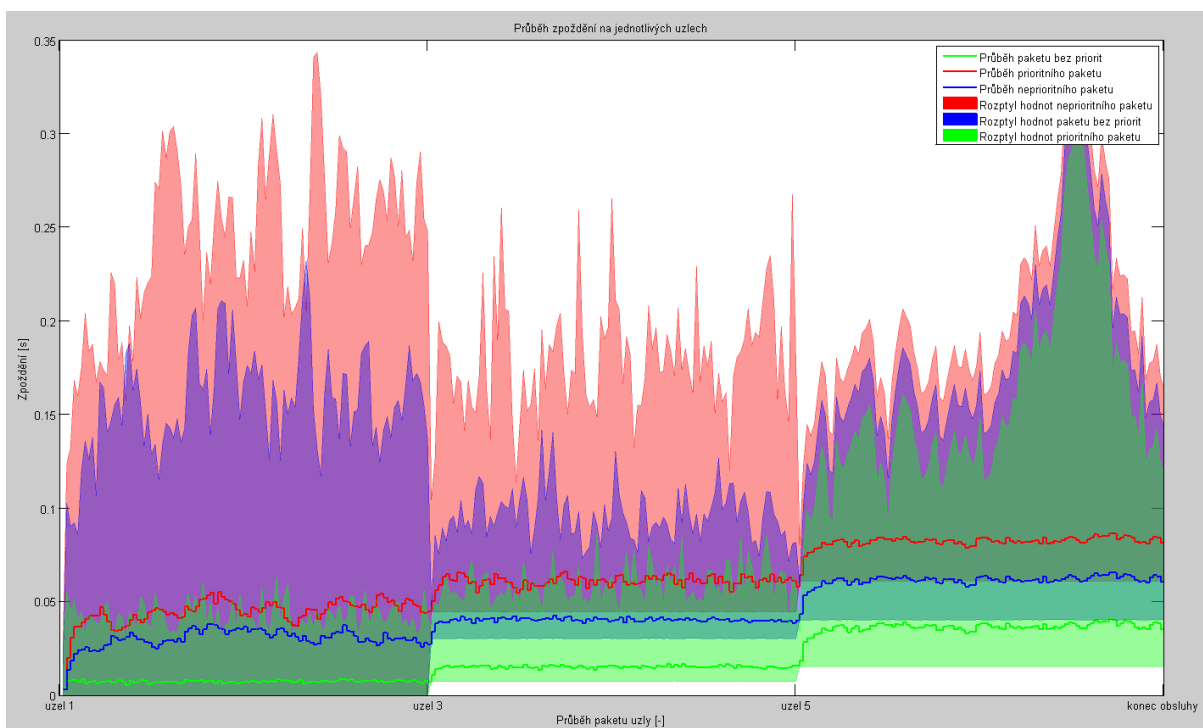
Po kliknutí na tlačítko start je spuštěna simulace dle zadaných parametrů. Průběh výpočtu jednotlivých uzlů je znázorněn pomocí dialogových oken, viz obr. 11. Kliknutím na tlačítko Cancel v dialogovém okně zobrazující průběh výpočtu lze simulaci kdykoliv zastavit.



Obr. 11: Zobrazení probíhajícího výpočtu

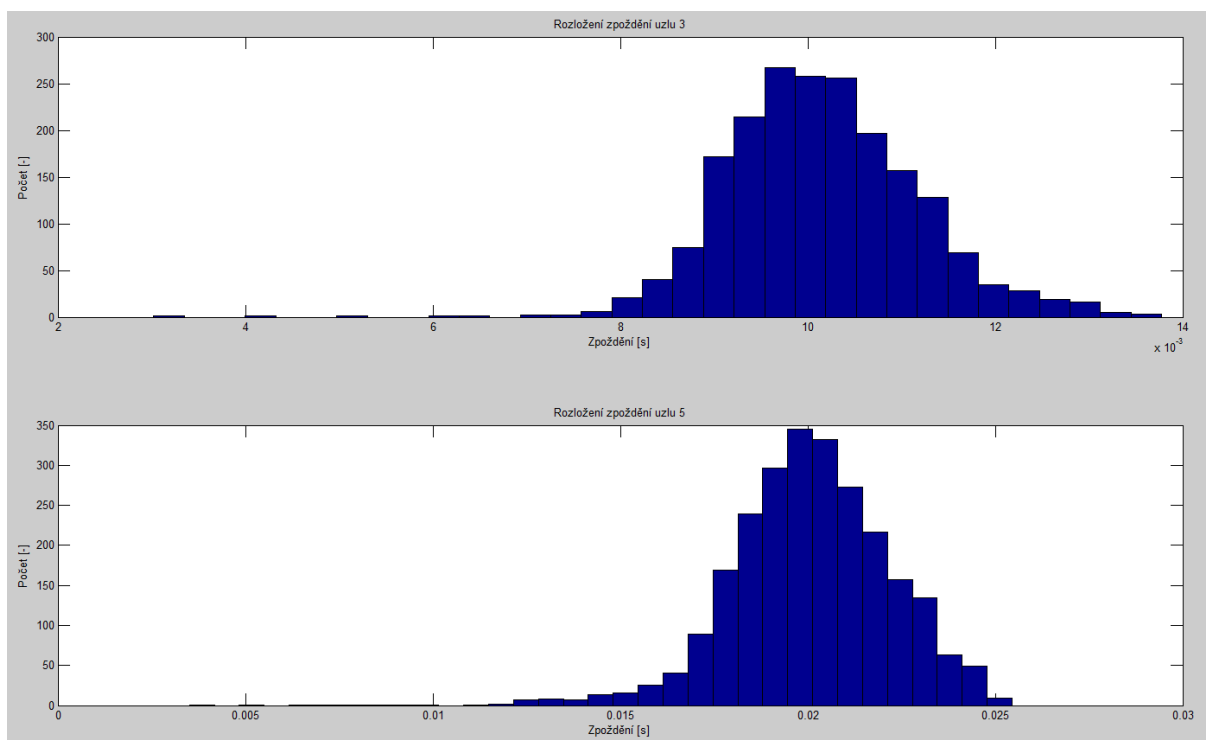
Po dokončení simulace je zobrazena trojice grafů:

První graf zobrazuje vývoj zpoždění jednotlivých paketů procházejících uzlem (uzly). Na vodorovné je pro lepší přehlednost místo hodnot označujících čas zvoleno označení jednotlivých uzlů, přes které paket putuje. Pokud uživatel u některého ze zvolených uzlů zvolil model s prioritami, pak budou v grafu zobrazeny tři křivky – pro prioritní pakety, pro ostatní pakety a pro všechny pakety bez priorit. Jednotlivé křivky jsou barevně označeny a toto barevné označení odpovídá i barvám ve výsledkové části programu. Světlejší plochy vymezují rozptyl hodnot ze všech realizací. Při velkém množství paketů vstupujících do uzlu nejsou v grafu znázorněny všechny pakety z důvodu zachování přehlednosti a lepší vypovídající hodnotě grafu. Příklad tohoto grafu uvádí obr. 12.



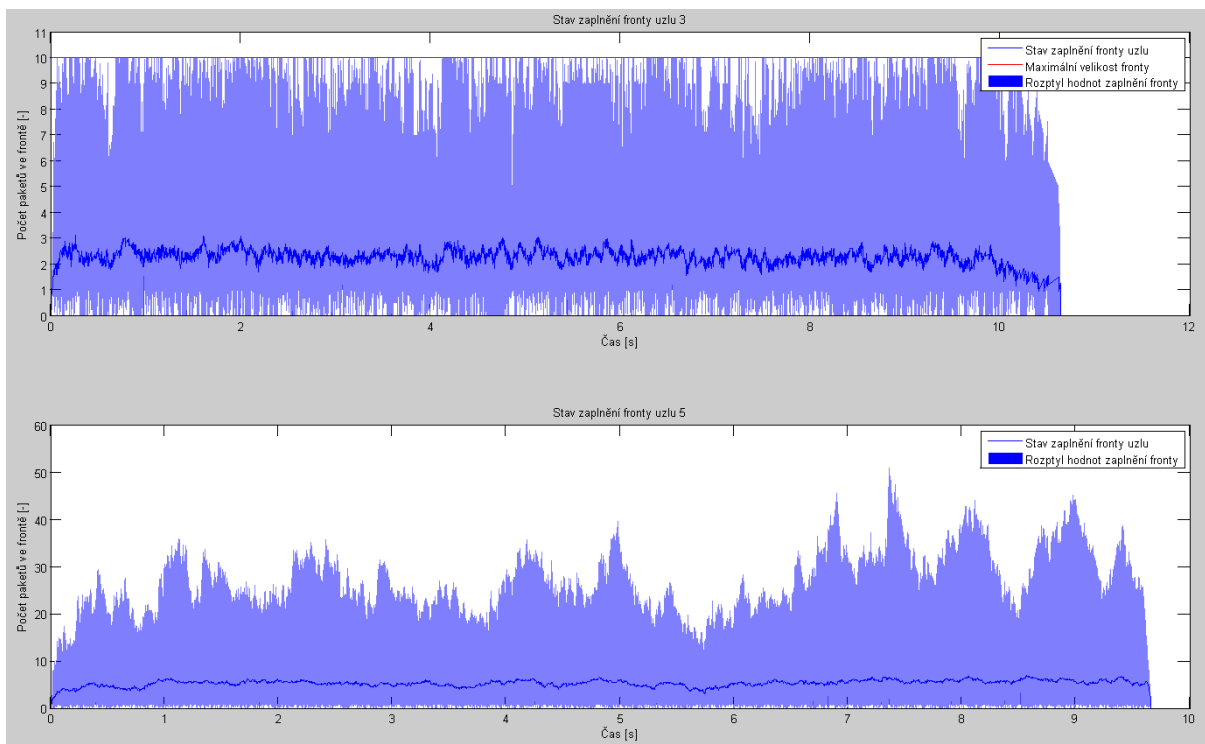
Obr. 12: Graf zpoždění paketů

Druhý graf zobrazuje histogram rozložení zpoždění paketů na jednotlivých uzlech. Rozložení zpoždění je zobrazeno pro každý uzel zvlášť, tedy s větším počtem zvolených uzlů je zobrazeno větší množství grafů. Příklad je uveden na obr. 13.



Obr. 13: Graf rozložení zpoždění paketů

Třetí graf znázorňuje stav zaplnění fronty jednotlivých uzlů. Je zde zachyceno průměrné zaplnění fronty v průběhu času, světlejší barva ohraničuje oblast hodnot ze všech realizací. Zaplnění fronty je opět zobrazeno pro každý uzel zvlášť. Pokud uživatel zvolil model s konečnou délkou fronty, je graf doplněn o barevně odlišenou přímku představující maximální velikost fronty. Příklad tohoto grafu uvádí obr. 14.



Obr. 14: Graf zaplnění vyrovnávací paměti uzlů

5.1.4 Doplnující poznámky

- Program lze ukončit vybráním položky z menu či uzavřením hlavního okna programu
- Grafy je možné zobrazovat a skrývat pomocí tlačítka v ovládacím panelu.

5.2 Popis řešení funkcí, které realizují výpočty

Základním principem simulace je opakované provádění experimentů za účelem pochopení chování studovaného systému – čím větší počet opakování, tím přesnější bude výsledek experimentu. V programu to znamená opakované provádění výpočtu a následné statistické zpracování získaných hodnot.

Pro všechny funkce provádějící výpočty je třeba vytvořit odpovídající posloupnosti hodnot náhodné veličiny, které budou odpovídat danému rozdělení pravděpodobnosti. V případě tohoto programu jsou náhodnými veličinami příchod paketů a doba trvání obsluhy. K tvorbě posloupnosti hodnot náhodné veličiny je využito pseudonáhodných čísel, která generuje MATLAB. Tyto pseudonáhodná čísla mají rovnoměrné rozdělení a jsou z intervalu $<0; 1)$.

5.2.1 Výpočet modelu M/M/1

Abychom získali posloupnost hodnot náhodné veličiny s odpovídajícím rozdělením, je využito metody inverzní transformace. Podrobnější popis metody viz [20]. Hodnoty náhodné veličiny X s exponenciálním rozdělením pak získáme transformací podle vztahu:

$$X = -\frac{1}{\mu} \ln(1 - r), \quad (5.1)$$

kde μ je parametr exponenciálního rozdělení a r je vygenerované náhodné číslo z intervalu $(0; 1)$.

Okamžiky příchodu paketů jsou stanoveny kumulativním načítáním intervalů mezi příchody.

K určení času ukončení obsluhy je potřeba porovnat okamžik ukončení obsluhy předchozího paketu a okamžik příchodu aktuálního paketu. Pokud aktuální paket přišel až po dokončení obsluhy předchozího paketu, je doba ukončení obsluhy dána dobou příchodu aktuálního paketu, ke které je přičtena doba obsluhy aktuálního paketu. V případě, že aktuální paket přišel před ukončením obsluhy předchozího paketu, je doba ukončení jeho obsluhy dána dobou příchodu aktuálního paketu, ke které je přičtena doba čekání (rozdíl mezi dobou příchodu nového paketu a dobou ukončení obsluhy předchozího paketu) a doba obsluhy aktuálního paketu.

Při znalosti těchto hodnot již můžeme získat některé ze základních charakteristik – průměrnou dobu obsluhy paketu, průměrnou dobu čekání ve frontě na obsluhu a průměrnou dobu strávenou v uzlu. Průměrnou intenzitu provozu pak získáme jako podíl počtu čekajících paketů a celkového počtu paketů.

Pro stanovení průměrného počtu paketů ve frontě a v uzlu je potřeba určit počty paketů v jednotlivých časových úsecích – tedy určíme dobu, po kterou se v uzlu nacházel 1 paket, 2 pakety, atd. V této fázi se také sleduje stav zaplnění fronty. V případě, že je počítán model s konečnou délkou fronty a následující paket by překročil nastavenou hodnotu délky fronty, zvýší se počítadlo zahozených paketů a zahozený paket není zahrnut do výpočtů. Ze znalosti doby, kterou určitý počet paketů strávil v uzlu a ze znalosti celkové sledované doby pak podílem těchto dvou hodnot dostaneme pravděpodobnosti, s jakými se v uzlu nachází právě n paketů. Hledané hodnoty – průměrný počet paketů v systému a ve frontě jsou pak dány sumou součinů n a příslušných pravděpodobností [8].

5.2.2 Výpočet modelu M/M/1 s prioritami

Tento výpočet je rozšířením výše popsaného výpočtu $M/M/1$. Jsou zde tedy dvě posloupnosti intervalů mezi příchody (pro prioritní a neprioritní vstupní tok) a je přidán mechanismus pro upřednostnění prioritních paketů. Program pracuje s nepreemptivní prioritou, pokud přijde prioritní paket během obsluhování paketu s nižší prioritou, obsluha paketu s nižší prioritou není přerušena, ale je dokončena. V programu to tedy znamená zajistit, aby po dokončení obsluhy paketu byla prozkoumána fronta a následně byl k obsluze vybrán nejdéle čekající prioritní paket. V případě, že se ve frontě žádný prioritní paket nenachází, je obslužen nejdéle čekající neprioritní paket. Pokud se ve frontě nenachází žádný paket, bude systém po určitý čas v nečinnosti.

5.2.3 Výpočet modelu M/G/1

V tomto modelu je možné nastavovat rozložení doby obsluhy, které se provádí pomocí volby koeficientu variace c . Pokud je $c < 1$, bude mít doba obsluhy Erlangovo rozdělení, pokud je $c > 1$, bude se jednat o hyperexponenciální rozdělení a v případě, kdy se $c = 1$, jde o již popsané exponenciální rozdělení. Pro získání hodnot Erlangova a hyperexponenciálního rozložení využijeme faktu, že všechny tyto tři typy rozložení jsou speciálními případy rozdělení fázového typu (PH). Exponenciální rozdělení tvoří jedna fáze. Erlangovo rozdělení je tvořeno sériovým řazením dvou nebo více fází – proto lze Erlangovo rozdělení získat součtem dvou nebo více exponenciálních rozdělení. Počet fází je určen převrácenou hodnotou koeficientu variace. Hyperexponenciální rozdělení je pak tvořeno dvěma nebo více fázemi, z nichž každá má určitou pravděpodobnost výskytu. Výskyt jednotlivých fází je vzájemně vylučný, což odpovídá paralelnímu řazení fází [24]. Vzhledem k náročnějšímu získávání hodnot při paralelním řazení více fází je hyperexponenciální rozdělení v programu tvořeno dvěma fázemi se shodnou pravděpodobností výskytu. Koeficient variace pro hyperexponenciální rozdělení uvedený v programu byl pak získán empiricky.

5.2.4 Výpočet modelu M/G/1 s prioritami

Tento výpočet využívá postupů popsaných u výpočtu modelu $M/G/1$ doplněný o mechanismus vyřizování priorit popsaný u výpočtu modelu $M/M/1$ s prioritami.

5.3 Orientační porovnání simulace a výpočtu pomocí vzorců

V této části uvádím porovnání výsledků získaných ze simulace pomocí popsaného programu a pomocí vzorců uvedených v předchozích kapitolách. Vstupní parametry jsou následující: 1 uzel, model $M/M/1$ s prioritami, $\lambda_1 = 1000$, $\lambda_2 = 1700$, $\mu = 3000$, délka běhu 10 sekund.

Tabulka 1: Vypočtené a nasimulované hodnoty

	Výpočet	Simulace (počet běhů)				Rozdíl výpočtu a simulace (500 běhů)
Počet běhů	-	2	10	100	500	
Provozní zatížení	0,90000000	0,89793000	0,89693000	0,89919000	0,90444000	-0,04440000
Střední počet paketů v uzlu	9,00000000	9,53630000	9,08130000	8,77420000	9,01580000	-0,01580000
Střední počet paketů ve frontě	8,10000000	8,64130000	8,19200000	7,87680000	8,11610000	-0,01610000
Střední doba strávená v uzlu						
prioritní	0,00078333	0,00079011	0,00078286	0,00078446	0,00078152	0,00000100
neprioritní	0,00483100	0,00502450	0,00480760	0,00481140	0,00479890	0,00003200
bez priorit	0,00333000	0,00310300	0,00343190	0,00331060	0,00332400	0,00000600
Střední doba strávená ve frontě						
prioritní	0,00045300	0,00045499	0,00044920	0,00045083	0,00044821	0,00000500
neprioritní	0,00451000	0,00468930	0,00447390	0,00447780	0,00446560	0,00005000
bez priorit	0,00300000	0,00277100	0,00309810	0,00297710	0,00299080	0,00000900

Z výsledku je zřejmé, že pro danou délku běhu a počet opakování je rozdíl mezi vypočtenými a simulovanými hodnotami nižší než 5%. Z tabulky je také vidět, že simulace dává pro tento případ poměrně přesné výsledky už při deseti opakováních.

6 NÁVRH LABORATORNÍ ÚLOHY

Navržená laboratorní úloha je zaměřena na základy teorie hromadné obsluhy a její aplikaci v praxi při návrhu počítačové sítě. Úloha je určena studentům, kteří se s problematikou teorie hromadné obsluhy setkávají poprvé – tomu odpovídá i koncepce úlohy. K jejímu provedení studenti nebudou potřebovat žádné znalosti z této oblasti, předpokládá se pouze elementární znalost teorie pravděpodobnosti, se kterou se student na této úrovni již musel dříve setkat. Potřebné teoretické základy a vysvětlení principů modelů hromadné obsluhy je v dostatečné míře rozebráno v teoretickém úvodu práce. Cílem úlohy je kromě získání základního povědomí o systémech hromadné obsluhy a jejich využití při návrhu či optimalizaci paketových sítí také praktické stanovení parametrů jednoduchého obsluhového systému. Praktická část úlohy je založena na porovnání analytického a simulačního přístupu k dané problematice. Zatímco v prvním případě budou studenti určovat parametry výpočtem, pomocí uvedených vzorců, ve druhém případě k určení parametrů systémů využijí simulačního programu. Úkolem studentů je porovnat obě metody a určit jejich výhody a nevýhody. V závěru úlohy je pak uvedeno několik úkolů zaměřených na experimentování se simulačním programem, jejichž smyslem je, aby si studenti uvědomili či ověřili správné pochopení uvedených principů systémů hromadné obsluhy. Samotná laboratorní úloha je uvedena v příloze A.

ZÁVĚR

Tato práce se zabývala problematikou modelů hromadné obsluhy a jejímu použití při návrhu modelu počítačové sítě. Teoretické cíle práce, tj. problematika datových sítí z hlediska teorie hromadné obsluhy, prostředky pro modelování jednotlivých prvků sítě a návrh modelů postihovala první část práce, kde byl mimo jiné popsán obecný model systému hromadné obsluhy a jeho nejdůležitější parametry. Dále zde bylo uvedeno matematické pozadí problematiky – především Markovovy procesy, jejichž využití umožňuje analyzovat jinak značně složité reálné problémy.

Praktickým cílům, kterými bylo vytvoření simulačního programu a návrh laboratorní úlohy, byla věnována druhá polovina práce. Simulace představuje nástroj, který se dnes běžně používá ke stanovení předpokládaného chování a výkonnosti velkých komplexních stochastických systémů. Vytvořit přesný simulační model ovšem není snadné a vyžaduje to rozsáhlé zdroje. Pokud model není přesný, mohou výsledky simulace vést k chybným závěrům. Základním problémem je, že každý model představuje určité zjednodušení skutečnosti, a tedy už v základu je do určité míry chybný. V důsledku toho je výsledek simulace vždy pouze tak dobrý, jak dobře je navržený model simulovaného systému. Pro řešení dané problematiky jsem zvolil využití modelů $M/M/1$ a $M/G/1$. Oba modely obsahují i své rozšířené varianty, aby bylo možné simulovat i modely s konečnou délkou fronty a dvěma prioritními třídami. Tyto modely jsem se rozhodl využít také proto, že je možné k nim odvodit vzorce pro analytické řešení takového systému. Toho se využívá v navržené laboratorní úloze, kde mají studenti za úkol porovnávat analytickou a simulační metodu. V prostředí MATLAB byl vytvořen simulační program, který umožňuje simulovat uzly sítě pomocí výše uvedených modelů na pevně dané topologii sítě. Uživatel má možnost nastavovat jednotlivé vstupní parametry a na výsledných hodnotách a grafech pozorovat, jak tyto vstupní parametry ovlivní výsledek. Správnost funkcí realizujících simulace byla ověřena porovnáním výsledků získaných programem s výsledky získanými pomocí analytického výpočtu. Pro dostatečnou délku simulace a počet běhů byly tyto výsledky velmi podobné a rozdíl se pohyboval pod 5%. Lze tedy usuzovat, že realizace modelů byla provedena správně. Navržená laboratorní úloha má studentům, kteří se s teorií hromadné obsluhy pravděpodobně setkávají poprvé, poskytnout náhled do této problematiky a do modelování prvků počítačové sítě. Úloha je postavena tak, aby student k jejímu provedení nepotřeboval žádné znalosti z této

oblasti, potřebné teoretické poznatky a vzorce nalezne v teoretickém úvodu k práci. Primárním cílem úlohy je, aby student porovnal rozdíl mezi analytickou a simulační metodou. Dále je zde ponechán prostor pro experimentování se simulačním programem, při kterém by si měl student uvědomit či ověřit některé ze základních principů systémů hromadné obsluhy.

SEZNAM POUŽITÉ LITERATURY

- [1] ALFA, A., S. *Queueing Theory for Telecommunications*. New York : Springer, 2010. ISBN 978-1441973139.
- [2] COOPER, R., B. *Introduction to Queueing Theory*. New York : Elsevier North Holland, 1981. ISBN 978-0444003799.
- [3] DUNAYTSEV, R. *Erlang formulas and M/G/1 systems* [online]. 2009. Dostupné z: <<http://www.cs.tut.fi/~dunaytse/lectures.html>>.
- [4] IVERSEN, V., B. *Teletraffic Engineering Handbook*. Dostupné z: <<http://www.tele.dtu.dk/teletraffic>>.
- [5] KESIDIS, G. *An Introduction to Communication Network Analysis*. New Jersey : Wiley-IEEE Press, 2007. ISBN 978-0471371410.
- [6] MOLNÁR, K. *Teorie front* [online]. Dostupné z: <<http://www.utko.feec.vutbr.cz/~molnar/mmos>>.
- [7] TRIVEDI, K., S. *Probability and Statistics with Reliability, Queuing and Computer Science Applications*. New York : John Wiley & Sons, 2002. ISBN 978-0471333418.

Citovaná literatura

Monografie

- [8] BOSE, S., K. *An Introduction to Queueing Systems*. New York : Plenum Publishers, 2002. ISBN 978-0306467349.
- [9] DAIGLE, N., J. *Queueing Theory with Applications to Packet Telecommunication*. Boston : Springer, 2005. ISBN 978-0387228570.
- [10] FLOOD, J., E. *Telecommunications Switching, Traffic and Networks*. New Jersey : Prentice Hall, 1995. ISBN 978-0130333094.

Článek v časopisu

- [11] ERLANG, A., K. *The Theory of Probabilities and Telephone Conversations*. Nyt Tidsskrift for Matematik, 1909, Vol. 20, No. B., s. 33 - 39.
- [12] KENDAL, D., G. *Some problems in the theory of queues*. Journal of the Royal statistical society, 1951, Ser. B 13, s. 151 – 185. ISSN 0964-1998.

Skripta

- [13] KOŘENÁŘ, V. *Stochastické procesy*. Praha : Nakladatelství Oeconomica, 2010. ISBN 978-8024516462.
- [14] TYC, O. *Operační výzkum*. Brno : Mendelova zemědělská a lesnická univerzita, 2003. ISBN 80-7157-726-X.

Elektronické dokumenty

- [15] ADAN, I. *Materiály k předmětu Queueing theory*. Technische Universiteit Eindhoven [online]. [cit 2011-04-22]. Dostupné z: <<http://www.win.tue.nl/~iadan/que/>>
- [16] BARTÍK, V. *Markovské řetězce s diskrétním časem* [online]. [cit 2010-12-10]. Dostupné z: <<http://math.feld.cvut.cz/ftp/bartik/M6b/>>
- [17] *Dimenzování a výpočet ztráty obsluhových systémů se ztrátou*. Katedra telekomunikační techniky ČTVUT-FEL [online]. [cit 2010-12-04]. Dostupné z: <<http://www.comtel.cz/cz/vyuka/tho/index.php>>.
- [18] KONOPKA, L. *Analýza obsluhového systému s prioritami* [online]. 2006-05-01 [cit 2010-12-02]. Dostupné z: <<http://access.feld.cvut.cz/view.php?cisloclanku=2006010201>>.
- [19] NAIN, P. *Basic elements of queueing theory* [online]. [cit 2010-11-25]. Dostupné z: <www.cs.columbia.edu/~misra/COMS6180/>.
- [20] ŠEDIVÁ, B. *Materiály k předmětu Výpočtová statistika*. Katedra matematiky Západočeské univerzity v Plzni [online]. [cit 2011-04-29]. Dostupné z: <home.zcu.cz/~sediva/stav/>.

- [21] *Telecommunications Queueing Theory*. University of Wollongong [online]. [cit 2011-04-29]. Dostupné z: <<http://faraday.elec.uow.edu.au/subjects/autumn/ecte961/>>.
- [22] VIRTAMO, J. *Materiály k přemětu Queueing Theory*. Aalto University [online]. [cit 2011-04-03]. Dostupné z: <<http://www.netlab.tkk.fi/opetus/s383143/kalvot/english.shtml>>.
- [23] VORÁČKOVÁ, Š. *Kendallová klasifikace* [online]. [cit 2010-11-15]. Dostupné z: <<http://www.fd.cvut.cz/departament/k611/PEDAGOG/K611THO.html>>.
- [24] WALCK, CH. *Hand-book on statistical distributions*. Universitet Stockholm [online]. [cit 2011-05-05]. Dostupné z: <<http://www.physto.se/~walck/>>.

SEZNAM OBRÁZKŮ

Obr. 1: Příklad systému hromadné obsluhy	10
Obr. 2: Proces množení a zániku	24
Obr. 3: Schéma modelu $M/M/1$	28
Obr. 4: Charakteristiky modelu $M/M/1$	31
Obr. 5: Schéma modelu $M/G/1/K$	33
Obr. 6: Hlavní okno programu	36
Obr. 7: Nastavení vstupních parametrů uzlu	37
Obr. 8: Vypočtené hodnoty (bez priorit)	39
Obr. 9: Vypočtené hodnoty (s prioritami)	39
Obr. 10: Panel ovládání a celkových výsledků	40
Obr. 11: Zobrazení probíhajícího výpočtu	41
Obr. 12: Graf zpoždění paketů	41
Obr. 13: Graf rozložení zpoždění paketů.....	42
Obr. 14: Graf zaplnění vyrovnávací paměti uzlů	43

SEZNAM POUŽITÝCH ZKRATEK, VELIČIN A SYMBOLŮ

FIFO	First In First Out, režim fronty
FTP	File Transport Protocol, protokol pro přenos souborů v síti
LIFO	Last In First Out, režim fronty
PRI	Priority, režim fronty s prioritami
QoS	Quality of Service, termín užívaný pro rezervaci a řízení datových toků
RTP	Real-time Transport Protocol, protokol pro paketizaci multimediálních toků v reálném čase
SHO	Systém hromadné obsluhy
SIRO	Selection In Random Order, režim fronty s náhodným výběrem
TCP	Transmission Control Protocol, transportní protokol (spolehlivý)
UDP	User Datagram Protocol, transportní protokol (nespolehlivý)
VoIP	Voice over Internet Protokol, internetová telefonie
$\mathbf{A}(n)$	Matice intenzit přechodu
c_v	Koeficient variace
$E(L)$	Průměrný počet požadavků v systému
$E(L_q)$	Průměrný počet požadavků ve frontě
$E(R)$	Zbývající doba obsluhy požadavku
$E(S)$	Průměrná doba obsluhy
$E(W)$	Průměrná doba strávená v systému
$E(W_q)$	Průměrná doba čekání ve frontě
$\mathbf{0}$	Nulový vektor, nulová matice
PASTA	Poisson Arrivals See Time Averages, vlastnost systémů s poissonovským rozložením dob příchodů
$\mathbf{p}(n)$	Vektor pravděpodobností v určitém okamžiku n
$\mathbf{P}(n)$	Matice pravděpodobnosti přechodu
$p_0(t)$	Pravděpodobnost, že v systému v čase t bude 0 požadavků
P_B	Pravděpodobnost blokování
p_k	Pravděpodobnost, že v libovolném časovém okamžiku je ve frontě právě k požadavků

$p_n(t)$	Pravděpodobnost, že v systému v čase t bude právě n požadavků
$p_{n,n+1}(t, t + \Delta t)$	Pravděpodobnost přechodu ze stavu n do stavu $n+1$ během časového intervalu Δt
$P_n(T)$	Pravděpodobnost vstupní n jednotek do systému v intervalu T
$q(t)$	Hustota rozdělení pravděpodobnosti dob mezi příchody požadavků
T	Množina časových okamžiků
$X(t)$	Náhodná veličina
Δt	Časový interval
λ	Intenzita příchodů, intenzita vyskytujících se jevů
ρ	Intenzita provozu (provozní zatížení)
μ	Intenzita obsluhy, intenzita úmrtí procesu množení a zániku

SEZNAM PŘÍLOH

A. Laboratorní úloha	I
B. Obsah přiloženého média	XII
C. Disk DVD	

A. LABORATORNÍ ÚLOHA

SIMULACE ZPOŽDĚNÍ V PAKETOVÉ SÍTI POMOCÍ TEORIE HROMADNÉ OBSLUHY

Cíl

Získat základní povědomí o systémech hromadné obsluhy a jejich využití při návrhu či optimalizaci paketových sítí. Praktické stanovení parametrů obsluhového systému analyticky a pomocí simulace.

Teoretický úvod

Se systémy hromadné obsluhy (SHO) se setkáváme v mnoha oblastech reálného života. Příkladem systému hromadné obsluhy může být prodejna, kde proud požadavků tvoří zákazníci, kteří přicházejí nakupovat a obslužnými zařízeními jsou prodavači, kteří je obsluhují. Obecně lze tedy SHO popsat jako modely reálných systémů, jejichž funkce spočívá ve vykonávání obsluhy (tj. poskytování určité služby) pro velké počty přicházejících požadavků. Obsluhu realizují prvky označované jako kanály obsluhy (obslužné linky), před kanálem obsluhy se zpravidla vytváří fronta požadavků čekajících na poskytnutí služby.

Hlavním úkolem těchto modelů je určit výkonnostní charakteristiky systému, např. střední dobu strávenou požadavkem v systému, střední délku fronty apod. Zkoumají se tedy především vlastnosti a chování přicházejících požadavků a časová náročnost na poskytované služby.

Použití SHO je typické pro modelování komunikačních systémů, dopravních systémů, výrobních procesů, počítačových systémů, apod.

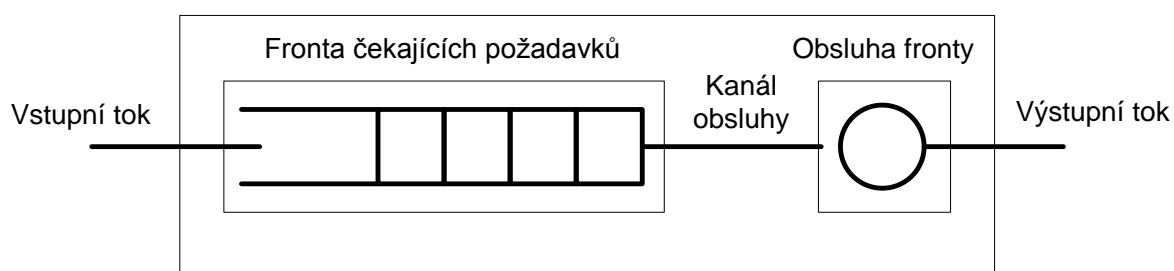
U modelů hromadné obsluhy je často cílem měření vzájemných vazeb mezi základními ukazateli charakterizujícími kvalitu a efektivitu systému hromadné obsluhy a nalezení optimálního provozního režimu vzhledem ke zvolenému kritériu – optimalizace systému zpravidla vede podstatnému snížení možnosti vzniku front, nebo snížení celkových ztrát z důvodu čekání. Na druhou stranu je také nutné se zabývat i ekonomickou stránkou problému – snížení pravděpodobnosti vzniku front je ve většině případů dosaženo rozšířením kapacity obslužného zařízení, tedy za cenu navršení prostředků na systém. Cílem je tedy nalézt optimální relaci mezi určitou mírou kvality poskytovaných služeb (představovanou

dobou obsluhy, ztrátami a dalšími kvalitu snižujícími parametry) a náklady na dosažení dané kvality služby (kdy se náklady pak promítnou do ceny služby pro zákazníka).

Pokud přijmeme určité zjednodušující podmínky, lze SHO koncipovat jako matematický model, jehož chování je možné vyjádřit pomocí vzorců (analytické řešení). Pro rozsáhlejší SHO a pro detailnější modelování jejich vlastností je většinou nutné řešení pomocí simulace.

SHO, jeho prvky a parametry

Základní struktura a prvky modelu jednoduchého SHO jsou patrné z následujícího obrázku.



Obr. 1: Jednoduchý systém hromadné obsluhy

Aby bylo možné vypočítat základní charakteristiky systému hromadné obsluhy, je potřeba adekvátním způsobem popsat i jeho hlavní prvky. Stochastický charakter mají především dva prvky systému – vstupní proud požadavků a doba trvání obsluhy. Lze je proto popsat pouze přibližně pomocí známých rozdělení pravděpodobnosti. Při popisu ostatních prvků systému se zpravidla problémy tohoto druhu nevyskytují. Základní charakteristiky, které sledujeme při modelování SHO, zpravidla zahrnují průměrný počet požadavků v systému a ve frontě, průměrnou dobu strávenou v systému, průměrnou dobu čekání ve frontě či pravděpodobnost blokování.

Vstupní proud požadavků

Vlastnosti a intenzita vstupního proudu mají rozhodující vliv na činnost systému hromadné obsluhy a na jeho základní charakteristiky. Požadavky přicházejí do SHO nejčastěji v náhodných okamžicích, $t_0 < t_1 < t_2 \dots$, takže vstupní proud požadavků lze obecně charakterizovat jako stochastický proces. Délky intervalů mezi jejich příchody τ_n představují hodnoty náhodných veličin, pomocí které lze vstupní proud požadavků popsat (většinou pomocí distribuční funkce pravděpodobnostního rozdělení $F_a(t) = p(\tau \leq t)$). Údaje získané z SHO velmi často vyhovují předpokladu Poissonova rozdělení počtu vstupujících požadavků

za určitý interval. Tomuto rozložení počtu požadavků odpovídá exponenciální rozdělení intervalů mezi jejich příchody.

Pokud je zdroj, ze kterého přicházejí požadavky do systému obsluhy omezený (konečný), jde o uzavřený systém, v opačném případě se jedná o otevřený systém. Zdroj požadavků je považován za neomezený i v případě, že nelze s jistotou určit, kolik požadavků si může vyžádat obsluhu.

Vstupní proud může být charakterizován třemi veličinami:

- distribuční funkcí pravděpodobnostního rozdělení intervalů mezi příchody - $F_a(t)$,
- střední hodnotou intervalu mezi příchody - $E(\tau) = E(A) = \frac{1}{\lambda}$, λ představuje intenzitu příchodů,
- koeficientem variace - $c_{vA} = \frac{\sigma(t)}{E(A)}$, který číselně charakterizuje nahodilost příchodů.

Pro zcela pravidelné příchody má hodnotu 0, pro příchody odpovídající Poissonovu rozložení má hodnotu 1. $\sigma(t)$ je směrodatná odchylka intervalu mezi příchody.

Doba obsluhy a režim obsluhy

Stejně jako v případě vstupního proudu, je při modelování systému obsluhy potřeba popsat i dobu obsluhy jednotlivých požadavků. Dobu obsluhy jednotlivých požadavků lze také pokládat za náhodnou veličinu, charakterizovanou opět pomocí distribuční funkce určitého rozdělení, tj. $F_s(t) = p(t_s \leq t)$, kde t_s je doba obsluhy. Obvykle se doba obsluhy řídí exponenciálním rozdělením

Počet a uspořádání obslužných linek představuje další faktor, na kterém závisí efektivnost provozu systému hromadné obsluhy. Probíhá-li obsluha pouze na jedné obslužné lince, jedná se o systém s jednoduchou obsluhou (jednokanálový systém obsluhy). Lze-li obsluhovat současně několik požadavků, jde o systém s vícenásobnou obsluhou neboli o vícekanálový systém obsluhy. V systémech s více obslužnými linkami se zpravidla používá paralelní uspořádání, existuje ovšem i řazení sériové.

Veličiny a vztahy charakterizující jeden kanál obsluhy:

- distribuční funkce pravděpodobnostního rozdělení doby obsluhy - $F_s(t)$,
- střední doba obsluhy - $E(t_s) = E(S) = \frac{1}{\mu}$, μ představuje intenzitu obsluhy,
- koeficient variace doby obsluhy - $c_{vS} = \frac{\sigma(t_s)}{E(S)}$, který číselně charakterizuje nahodilost (resp. pravidelnost) obsluhy. Pro shodné doby obsluh má hodnotu 0, pro obsluhy

s exponenciálním rozdělením má hodnotu 1. $\sigma(t_s)$ je směrodatná odchylka doby obsluhy.

Čekání ve frontě a režim výstupu z fronty

Podle „trpělivosti“ požadavků na obsluhu se SHO dělí na systémy bez čekání a systémy s čekáním. První typ je charakterizován tak, že pokud jsou všechna zařízení obsluhy obsazena, požadavek do systému nevstoupí a rezignuje na obsluhu. Ve druhém typu systému každý požadavek, který vstoupí, buď čeká tak dlouho, dokud neopustí systém po úspěšném ukončení obsluhy, nebo čeká alespoň určitou dobu a teprve není-li během této doby obsloužen, opouští systém.

Mezi systémy obsluhy s čekáním patří jak systémy s neomezenou délkou fronty, tak systémy s omezenou kapacitou fronty (popř. systému). Je-li počet míst ve frontě a tím v systému omezen, pak při jejich plném obsazení není dalším požadavkům přicházejícím do systému obsluha poskytnuta – to je charakteristické pro systémy se ztrátami. Při neomezené délce fronty jsou naopak všechny požadavky, které přicházejí do systému a jsou ochotny čekat, obslouženy.

Způsob, jakým se požadavky vstupující do systému řadí do fronty, popř. pořadí, v němž jsou obsluhovány, charakterizují režim výstupu z fronty. Většinou jde o pevně určené pořadí, kam patří dva základní režimy:

- a) FIFO (fronta) – požadavky jsou obsluhovány podle pořadí vstupu do systému
- b) LIFO (zásobník) – požadavky jsou obsluhovány v obráceném pořadí, než v jakém vstoupily do systému

Není-li stanoveno pevné pořadí vstupu požadavků do obsluhy, může režim fronty buď zohledňovat důležitost požadavků či naléhavost jejich obsluhy vyjádřenou různým stupněm priority (PRI), anebo jsou požadavky k obsluze přiřazovány zcela náhodně, bez jakýchkoliv předem daných pravidel (SIRO). V prvním případě se jedná o systémy s prioritami, v druhém o systémy s neuspořádanou frontou.

K popisu fronty slouží dva parametry:

- střední délka fronty - $E(L_q)$, tj. střední počet požadavků ve frontě
- střední doba čekání požadavků ve frontě - $E(W_q)$.

Kendallova klasifikace SHO

SHO lze klasifikovat z různých hledisek. Nejčastěji používanými kritérii jsou charakter a typ rozdělení vstupního proudu požadavků, charakter a typ rozdělení doby

obsluhy, režim obsluhy, čekání ve frontě, počet míst ve frontě a režim výstupu z fronty. Standardní klasifikace základních typů systémů hromadné obsluhy byla navržena D. G. Kendalllem. Používá třímístného kódu, založeného na kombinaci dvou písmen a jedné číslice ve formě $A/B/s$, kde první písmeno A označuje rozdělení intervalů mezi příchody požadavků, druhé písmeno B slouží k označení rozdělení dob trvání obsluhy a s udává počet paralelních zařízení obsluhy v systému. Nejčastěji používanými symboly jsou:

M – exponenciální rozdělení (Markovova typu),

E_k – Erlangovo rozdělení k -tého řádu,

D – deterministické rozdělení,

U – rovnoměrné rozdělení,

G – obecné rozdělení.

Například $M/M/1$ označuje SHO s exponenciálním rozdělením intervalů mezi příchody požadavků i dob obsluhy s jednou obslužnou linkou. Obdobně systém typu $D/H_2/1$ je jednokanálový s pravidelným vstupním proudem požadavků a s hyperexponenciálním rozdělením druhého řádu dob trvání obsluhy.

Veličiny a vztahy pro elementární SHO

Na obr. 2 jsou znázorněné tyto veličiny:

$E(L)$ – střední počet požadavků v systému,

$E(L_q)$ – střední počet požadavků ve frontě,

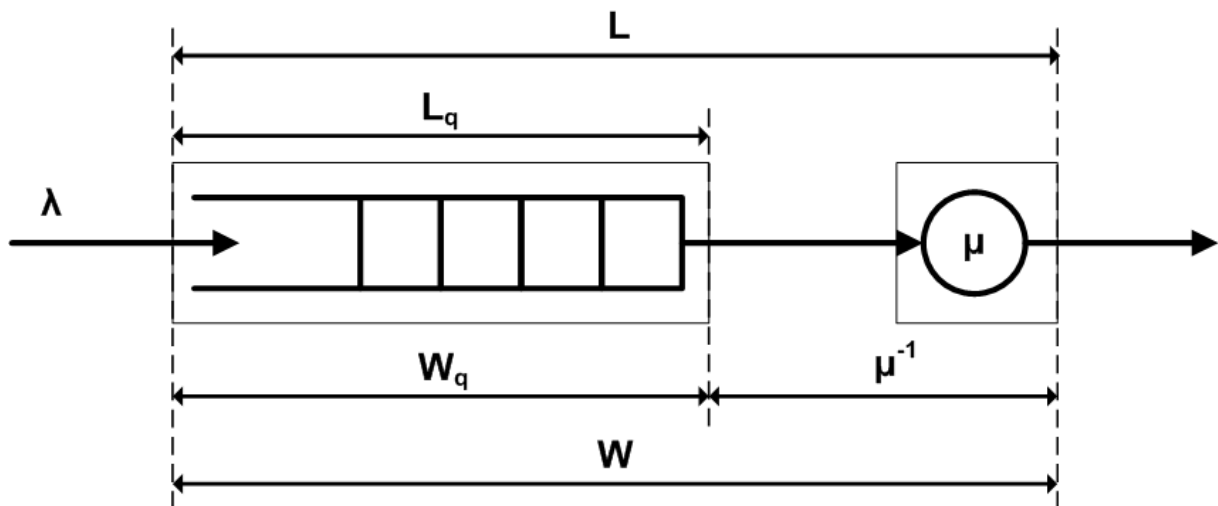
$E(W)$ – střední doba strávená v systému,

$E(W_q)$ – střední doba strávená v systému

Mezi těmito veličinami platí následující vztahy:

$$E(L) = E(L_q) + \rho \quad (1)$$

$$E(W) = E(W_q) + E(S) \quad (2)$$



Obr. 2: Charakteristiky modelu SHO

Intenzitu provozu lze při znalosti intenzity vstupního toku a intenzity obsluhy definovat jako $\rho = \frac{\lambda}{\mu}$. Aby byl systém stabilní, musí být splněna podmínka stability $\rho < 1$. Z uvedeného vztahu plyne, že pro stabilitu musí platit $\lambda < \mu$, tedy obsluha musí být rychlejší než příchody.

Pro elementární SHO dále platí Littleův teorém, který říká, že očekávaný počet zákazníků v systému se rovná součinu intenzity vstupního toku a očekávaného času, který zákazníci v systému stráví.

$$E(L) = \lambda \cdot E(W) \quad (3)$$

$$E(L_q) = \lambda \cdot E(W_q) \quad (4)$$

Důležité je, že stačí určit jednu (kteroukoliv) z nich a ostatní tři lze jednoduše vypočítat viz uvedené vztahy.

Model M/M/1

Vstupní proud požadavků je poissonovský, lze ho charakterizovat jediným parametrem - λ , (jde o parametr exponenciálního rozdělení $F(t) = 1 - e^{-\lambda t}$). Obdobně doba obsluhy má exponenciální rozdělení s parametrem μ . Zatížení SHO je potom $\rho = \frac{\lambda}{\mu}$, a pro dosažení stacionárního režimu činnosti musí být menší než 1.

Střední počet požadavků v SHO lze určit ze vztahu:

$$E(L) = \frac{\lambda}{\mu - \lambda} \quad (5)$$

Ostatní parametry lze odvodit pomocí vztahů uvedených v předchozích odstavcích:

- Průměrný počet požadavků ve frontě:

$$E(L_q) = \frac{\lambda^2}{\mu(\mu - \lambda)} \quad (6)$$

- Průměrná doba strávená v systému:

$$E(W) = \frac{1}{\mu - \lambda} = \frac{1}{\mu(1 - \rho)} \quad (7)$$

- Průměrná doba čekání ve frontě:

$$E(W_q) = \frac{\lambda}{\mu(\mu - \lambda)} \quad (8)$$

Pro model s prioritami jsou pak střední doby, které požadavek stráví v systému dány pomocí těchto vztahů (první a druhá prioritní třída):

$$E(W_1) = \frac{\frac{1 + \rho_2}{\mu}}{1 - \rho_1} \quad (9)$$

$$E(W_2) = \frac{\frac{(1 - \rho_1(1 - \rho_1 - \rho_2))}{\mu}}{(1 - \rho_1)(1 - \rho_1 - \rho_2)} \quad (10)$$

Model M/G/1

Vstupní proud požadavků je poissonovský, lze ho charakterizovat jediným parametrem - λ , (jde o parametr exponenciálního rozdělení $F(t) = 1 - e^{-\lambda t}$). Doba obsluhy může být specifikována distribuční funkcí $F(t)$ nebo hustotou $f(t)$. Protože průměrná doba čekání požadavku závisí na prvních dvou momentech doby obsluhy, stačí nám v praxi pro určení střední doby čekání znát střední dobu obsluhy (nebo intenzitu obsluhy) a koeficient variace. Zatížení SHO je potom $\rho = \frac{\lambda}{\mu}$.

Střední délku fronty lze určit ze vztahu:

$$E(L_q) = \frac{1 + c_v^2}{2} \cdot \frac{\rho^2}{1 - \rho} \quad (11)$$

Ostatní parametry lze odvodit ze vztahů uvedených v předchozích odstavcích:

- Průměrný počet požadavků v systému

$$E(L) = \rho + \frac{1 + c_v^2}{2} \cdot \frac{\rho^2}{1 - \rho} \quad (12)$$

- Průměrná doba čekání ve frontě

$$E(W_q) = \frac{1 + c_v^2}{2} \cdot \frac{\rho}{1 - \rho} \cdot E(S) \quad (13)$$

- Průměrná doba strávená v systému

$$E(W) = \left(1 + \frac{1 + c_v^2}{2} \cdot \frac{\rho}{1 - \rho} \right) \cdot E(S) \quad (14)$$

Koeficient c_v kvalifikuje nahodilost (pravidelnost) doby obsluhy:

- pro exponenciální rozdělení doby obsluhy je $c_v=1$ ($M/G/1$ přejde v $M/M/1$),
- pro shodné (nenáhodné) doby obsluhy je $c_v=0$ ($M/G/1$ přejde v $M/D/1$) a střední délka fronty vyjde dvakrát menší než pro $M/M/1$ se stejným zatížením,
- pro „nepravidelné“ doby obsluhy (popsané např. erlangovým či gaussovským rozdělením s danou střední hodnotou a směrodatnou odchylkou) je $0 < c_v < 1$ a střední délka fronty $E(L_q)$ vychází někde mezi výše uvedenými případy,
- pro případ, kdy $c_v > 1$ (hyperexponenciální nebo diskretní rozdělení), může vyjít střední délka fronty „horší“ než pro $M/M/1$.

Pro model s prioritami jsou pak střední doby požadavků jednotlivých priorit, které požadavek stráví v systému dány pomocí těchto vztahů:

$$E(W_1) = \frac{\rho_1 \cdot E(R_1) + \rho_2 \cdot E(R_2)}{1 - \rho_1} + E(S_1) \quad (15)$$

$$E(W_2) = \frac{\rho_1 \cdot E(R_1) + \rho_2 \cdot E(R_2)}{(1 - (\rho_1 + \rho_2))(1 - \rho_1)} + E(S_2) \quad (16)$$

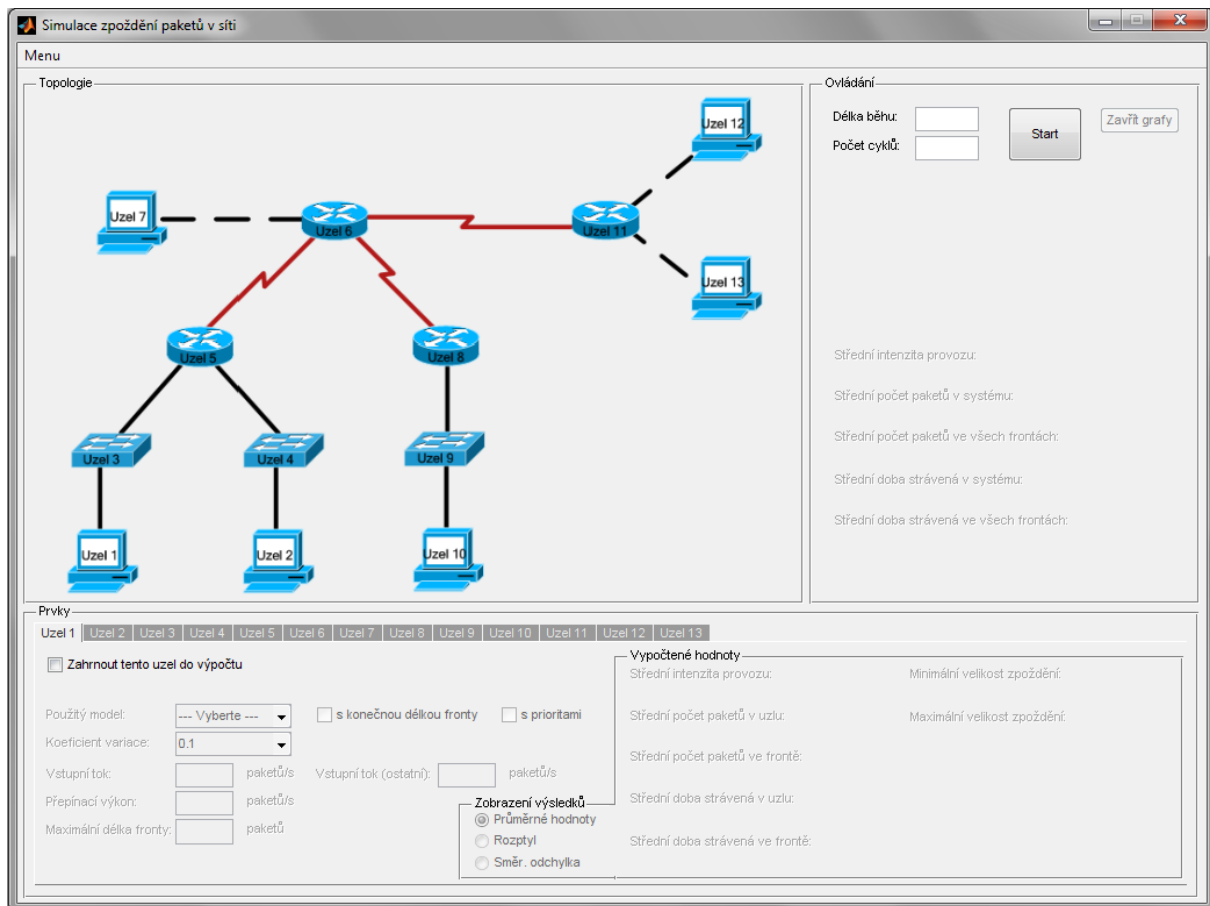
kde $E(R)$ značí zbývající dobu obsluhy požadavku, kterou lze vypočítat podle vztahů:

$$E(R) = \lambda \cdot \frac{1}{2} E(S^2),$$

$$E(S^2) = (1 + c_v^2) \cdot E(S)^2.$$

Simulační program

Simulační program umožňuje modelování jak jednoho konkrétního systému (uzlu), tak průchod paketu sítě (simulace více sériově řazených uzlů). Hlavní okno programu je na obr. 3.



Obr. 3: Hlavní okno simulačního programu

Při práci s programem uživatel nejdříve vybere uzel či uzly, které chce simulovat a zadá jejich vstupní parametry. Před spuštěním simulace je potřeba zadat také délku simulace a počet cyklů do příslušných polí.

Po dokončení simulace je zobrazena trojice grafů: graf zpoždění paketů, graf rozložení zpoždění paketů a graf zaplnění vyrovnávací paměti uzlu.

U každého uzlu jsou též zobrazeny výsledné parametry (lze zobrazit střední hodnoty, rozptyl a směrodatnou odchylku).

Podrobnější návod pro práci s programem je součástí samotného programu.

Zadání

- 1) Podle parametrů, které vám zadá vyučující, určete charakteristiky jednoduchého systému hromadné obsluhy:
 - a) analyticky (tj. pomocí vztahů uvedených v této úloze) určete
 - zatížení systému,
 - střední počet požadavků akumulovaných v SHO a ve frontě,
 - střední dobu trvání průchodu požadavku SHO a střední dobu čekání požadavku ve frontě.
 - b) pomocí simulačního programu určete
 - zatížení systému
 - střední délku fronty
 - střední dobu čekání požadavku ve frontě
 - stanovte počet běhů, popř. délku simulace, při které se získané údaje nebudou lišit o více než 10% od analytického řešení.
- 2) Podle parametrů zadaných vyučujícím určete charakteristiky systému hromadné obsluhy s prioritami
 - a) analyticky určete
 - střední dobu trvání průchodu požadavků obou priorit SHO a střední dobu čekání požadavků ve frontě.
 - b) pomocí simulačního programu
 - střední dobu trvání průchodu požadavků obou priorit SHO a střední dobu čekání požadavků ve frontě,
 - u simulace porovnejte získané hodnoty s hodnotami střední doby průchodu požadavků a čekání bez priorit
- 3) Experimentujte se simulačním programem:
 - porovnejte výsledky modelů $M/M/1$ a $M/G/1$, popište rozdíly,
 - nasimulujte systém s konečnou délkou fronty – podle získaných výsledků a s pomocí vzorce pro pravděpodobnost blokování vypočtěte, jak dlouhá by musela být fronta pro dosažení ztrátovosti $< 1\%$, pravděpodobnost blokování je pro $M/M/1$ systém dána vztahem

$$P_B = p_K = \frac{(1 - \rho)\rho^K}{1 - \rho^{K+1}}.$$

a pro $M/G/1$ vztahem

$$P_B = \frac{\rho^K}{K! \sum_{k=0}^K \frac{\rho^k}{k!}}$$

co se stane, pokud bude provozní zatížení > 1 ? Pokuste se daný scénář nasimulovat. Výsledky popište.

B. OBSAH PŘILOŽENÉHO MÉDIA

Na přiloženém médiu se kromě elektronické verze tohoto dokumentu také nachází vytvořený program pro simulaci zpoždění paketů v síti. Program je umístěn ve složce `simulator`.

Pro správnou funkci programu je třeba mít na PC nainstalovaný MATLAB, nebo nainstalovat Matlab Compiler Runtime v7.7, pomocí souboru `MCRInstall.exe`, který se nachází v adresáři `simulator`. Samotný program se pak spouští souborem `main.exe` umístěným ve složce `simulator\program`. Program byl testován s Matlabem v7.5 a MCR v7.7, na jiných konfiguracích nemusí být funkční. Zdrojové kódy jsou umístěny ve složce `simulator\program\main_mcr\main`.