

Teória informačných sietí

Prvé vydanie

Martin Klimo
Žilinská univerzita v Žiline

Jana Uramová
Žilinská univerzita v Žiline

Juraj Smieško
Žilinská univerzita v Žiline

Obsah

KAPITOLA 1	Modelovanie tokov v sieťach	1
1.1	Deterministický popis IP tokov	4
1.1.1	Použitie lineárneho (σ, ρ) ohraničenia	4
1.1.2	Lineárne ohraničenie pre periodický tok	8
1.1.3	Subaditívne vstupné krivky	10
1.2	Bernoulliho model IP toku	13
1.2.1	Pravdepodobnostné charakteristiky	16
1.2.2	Rozdelenie pravdepodobnosti medzier	20
1.2.3	Identifikácia Bernoulliho procesu	22
1.2.4	Rozdelenie pravdepodobnosti paketových zhlukov	23
	Literatúra	27
	Register	29

1

Modelovanie tokov v sieťach

OBSAH KAPITOLY

1.1	Deterministický popis IP tokov	4
1.1.1	Použitie lineárneho (σ, ρ) ohraničenia	4
1.1.2	Lineárne ohraničenie pre periodický tok	8
1.1.3	Subaditívne vstupné krivky	10
1.2	Bernoulliho model IP toku	13
1.2.1	Pravdepodobnostné charakteristiky	16
1.2.2	Rozdelenie pravdepodobnosti medzier	20
1.2.3	Identifikácia Bernoulliho procesu	22
1.2.4	Rozdelenie pravdepodobnosti paketových zhlukov	23

VZDELÁVACIE CIELE

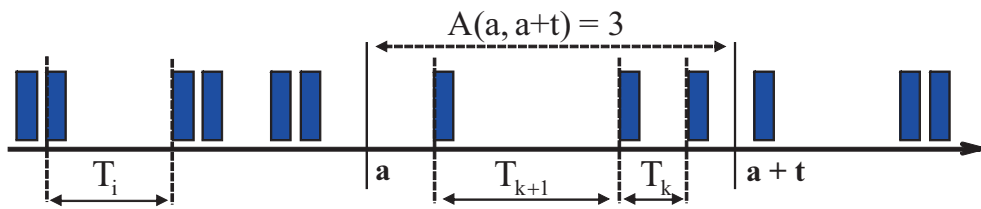
Po dôkladnom preštudovaní tejto kapitoly budete vedieť:

1. Naučiť sa niečo o modelovaní tokov v IP sieťach

Úvod

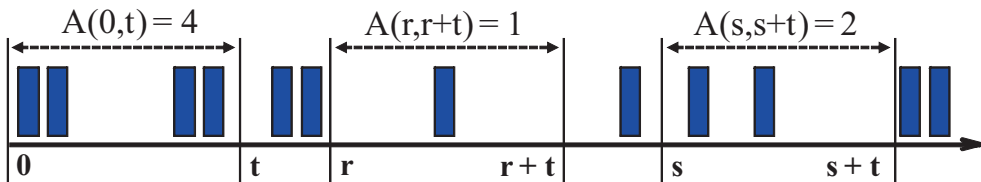
Dôležitou oblasťou výskumu v kontexte sieťových technológií je vývoj modelov IP prevádzky, ktoré môžu byť aplikované na internet, alebo všeobecnejšie na akúkoľvek komunikačnú sieť. Význam týchto modelov môžeme rozdeliť do dvoch skupín. V prvom rade, modely IP prevádzky sú potrebné ako vstup do sieťových simulácií. Tieto simulácie musia byť vykonávané pre účel štúdia a validácie sieťových algoritmov a protokolov, ktoré budú aplikované do reálnej prevádzky a pre účel analýzy ako IP prevádzka reaguje na rôzne podmienky a situácie v sieti, napr. na preťaženie siete a pod. Preto je dôležité, aby navrhnuté modely čo najvernejšie odrážali relevantné charakteristiky prevádzky, ktorú by mali reprezentovať. V druhom rade, dobrý model môže viesť k lepšiemu pochopeniu charakteristík prevádzky, čo môže pomôcť pri navrhovaní smerovačov a iných zariadení, ktoré riadia sieťovú prevádzku.

Paketový tok generovaný nejakým sieťovým zdrojom je jedným zo základných elementov IP siete. Existuje viacero koncepcií ako popísať tok udalostí. Jedným zo spôsobov je popis pomocou tzv. vstupného procesu $A(r, s)$, ktorý predstavuje kumulatívny počet výskytu udalostí (paketov) v intervale $\langle r, s \rangle$. Iný spôsob je použitie premenných T_i , ktoré popisujú dĺžku trvania časovej medzery medzi výskytom jednotlivých udalostí (paketov):



Obr. 1.1: Vstupný proces popísaný pomocou procesu $A(t)$ a premenných T_i .

Vo všeobecnosti budeme predpokladať, že tok predstavujúci IP prevádzku, je náhodný, pričom $A(r, s)$ bude predstavovať nejaký stochastický proces a T_i odpovedajúce náhodné premenné. Ďalším zjednodušením bude uvažovať iba o tzv. stacionárnych náhodných procesoch. Zjednodušene povedané, všetky pravdepodobnostné charakteristiky stacionárneho procesu sú invariantné vzhľadom na čas. Nezáleží, v ktorom časovom intervale sme začali proces pozorovať, záleží iba na dĺžke pozorovania.



Obr. 1.2: Stacionárny vstupný tok

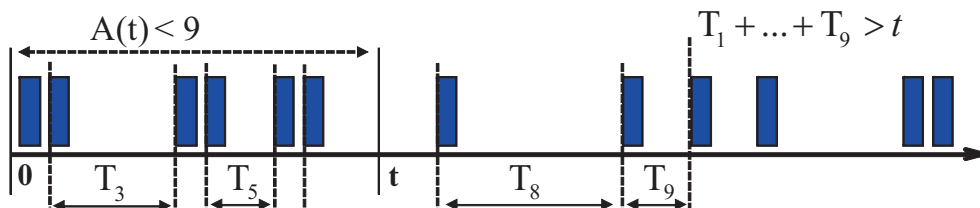
Pre stacionárny proces zavedieme zjednodušené označenie $A(t) = A(0, t)$.

Medzi rozdeleniami pravdepodobnosti stacionárneho procesu $A(t)$ a náhodných premenných T_i , ktoré popisujú medzery medzi udalosťami, platí nasledujúci vzťah. Ak za čas t sa v toku vyskytne menej než k udalostí, potom súčet prvých k

medzier je väčší ako doba t . Preto pravdepodobnosti nastatia uvedených udalostí sa rovnajú.

Pre stacionárny proces $A(t)$ a náhodné premenné T_i popisujúce medzery platí

$$\forall k; \quad \Pr(A(t) < k) = \Pr(T_1 + \dots + T_k > t)$$

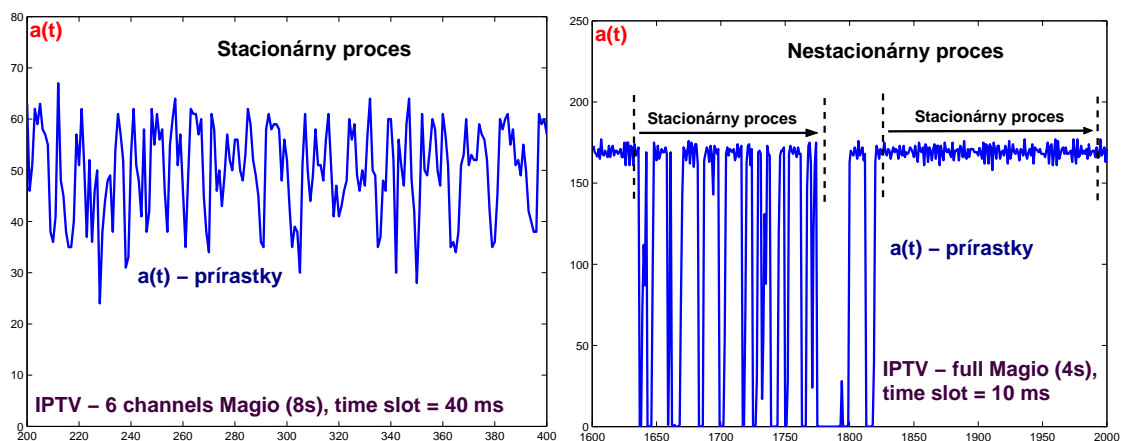


Obr. 1.3: Vzťah medzi procesom $A(t)$ a medzerami T_i

Ak predpokladáme diskretný čas, vstupný proces $A(t)$ sa skladá z tzv. prírastkov $a(i)$, ktoré zaznamenávajú výskyt paketov v čase i . Predpokladáme, že v čase 0 sa nevyskytol žiaden paket:

$$A(t) = \sum_{i=0}^t a(i), \quad a(t) = A(t) - A(t-1), \quad A(0) = a(0) = 0$$

Na nasledujúcich obrázkoch je ukážka merania prírastkov IPTV tokov, ktoré vychádzajú zo zdroja Mágio (ST-com). Na prvom obrázku je meranie 6 kanálov (4s). Podľa priebehu môžeme tok považovať za stacionárny. Na druhom obrázku je nameraná plná prevádzka Mágia (4s). Vidíme, že sa priebeh toku výrazne zmenil. Aj keď ide o nestacionárny tok, dá sa rozdeliť na stacionárne časti.



Obr. 1.4: Príklad stacionárneho a nestacionárneho procesu

Na otestovanie stacionarity procesu by sme potrebovali štatistické nástroje, čo je však mimo rámec tejto publikácie. Čisto laicky však môžeme povedať, že proces je stacionárny ak sa istým spôsobom pravidelne "vlní".

Ak je proces stacionárny, náhodné premenné T_i , ktoré popisujú medzery medzi udalosťami majú rovnaké rozdelenie pravdepodobnosti. Rovnako aj prírastky

toku $a(i)$ majú rovnaké rozdelenie. Ak navyše sú prírastky navzájom nezávislé, potom hovoríme, že proces $A(t)$ je stacionárny s nezávislými prírastkami, resp. (i.i.d) proces (*identical and independent distribution increments*).

**Stacionárny
proces s
nezávislými
prírastkami**

Ak náhodný proces $A(t)$ má rovnako rozdelené a navzájom nezávislé prírastky $a(i)$, hovoríme, že proces je stacionárny a nezávislý, resp. (i.i.d) proces. V takomto prípade sú aj medzery medzi udalosťami T_i rovnako rozdelené a navzájom nezávislé.

1.1 DETERMINISTICKÝ POPIS IP TOKOV

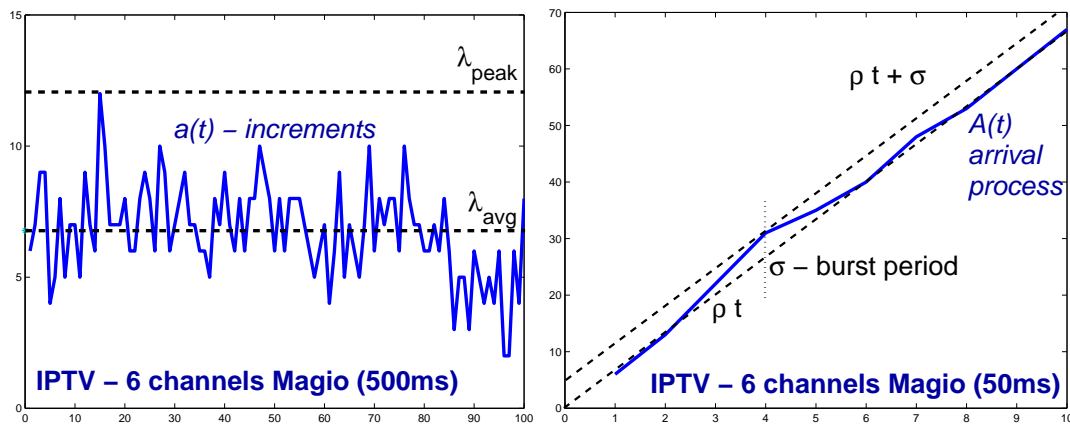
Zostaviť stochastický model pre nameraný IP tok je často komplikované, preto sa budeme najprv zaoberať deterministickými modelmi, ktoré využívajú subaditívne horné ohraňovania pre daný tok.

1.1.1 Použitie lineárneho (σ, ρ) ohraňovania

Na úplne základný popis IP prevádzky sa používajú dva parametre, intenzita toku, resp. stredný počet výskytu udalostí za časovú jednotku (*time slot*), $\lambda_{avg} = \rho$ (*average rate*) a tzv. špičková rýchlosť $\lambda_{peak} = p$ (*peak rate*). Tretí parameter σ (*burst period*), zaviedol Chang v [2]. Predstavuje maximálny rozdiel medzi vstupným procesom $A(t)$ a funkciou intenzity ρt . Burst periódu môžeme chápať ako najväčšiu "vlnu" v dátach, ktorá môže vstúpiť do daného zariadenia.

Pre dané meranie dĺžky n tieto parametre ľahko vypočítame:

$$\lambda_{avg} = \rho = \frac{A(n)}{n}, \quad \lambda_{peak} = \max_{0 \leq t \leq n} [a(t)], \quad \sigma = \max_{0 \leq t \leq n} [A(t) - \rho t]$$



Obr. 1.5: Základné parametre pre prírastky $a(t)$ a kumulatívny proces $A(t)$

Pomocou intenzity toku a burst periódy vieme určiť priamku $\sigma + \rho t$, ktorá daný tok, s ohľadom na dĺžku merania, zhora ohraňuje.

(σ, ρ)
ohraničenie

Hovoríme, že proces $A(t)$ je (σ, ρ) zhora ohraničený, ak platí

$$\forall s, t; \quad 0 \leq s \leq t; \quad A(t) - A(s) \leq \rho(t - s) + \sigma$$

PRÍKLAD 1.1

Majme nameraný nejaký paketový tok o dĺžke 15. Prírastky toku v jednotlivých časových slotoch sú zaznamenané v tabuľke:

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$a(t)$	2	3	3	4	6	6	5	4	1	2	3	2	2	1	1

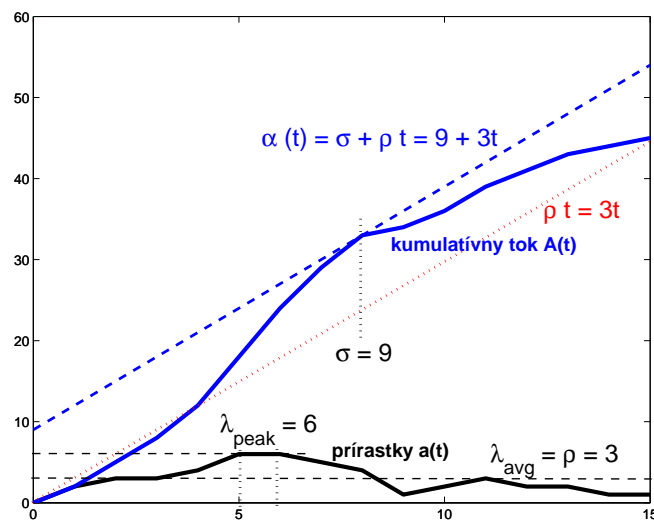
Jednoducho zistíme špičkovú a priemernú intenzitu:

$$\lambda_{peak} = \rho = 6, \quad \lambda_{avg} = \frac{1}{15} \sum_{t=1}^{15} a(t) = \frac{45}{15} = 3$$

Pre výpočet burst parametra musíme vytvoriť kumulatívny tok $A(t)$ a porovnať jeho hodnoty s funkciou ρt , ktorá predstavuje priemernú intenzitu toku za časový interval $\langle 0, t \rangle$.

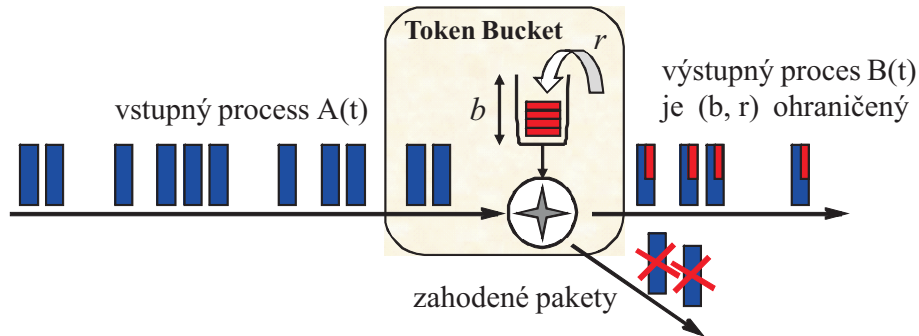
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$A(t)$	2	5	8	12	18	24	29	33	34	36	39	41	43	44	45
ρt	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45
$A(t) - \rho t$	-1	-1	-1	0	3	6	8	9	7	6	6	5	4	2	0

Pre daný tok je burst perióda $\sigma = \max_{0 \leq t \leq n} [A(t) - \rho t] = 9$ pre $t = 8$, a na ohraničenie toku môžeme použiť priamku $\alpha(t) = 9 + 3t$.



Obr. 1.6: Ohraničenie vstupného toku funkciou $\alpha(t) = 9 + 3t$

Na riadenie vstupu IP tokov do chrbticovej siete sa používajú tzv. Token Buckety (TB) alebo Peak limitéry (PR), ktoré sú priamo implementované na vstupných portoch prístupového uzla. Z hľadiska modelovania Token Bucket predstavuje zásobník na povolenia (tokeny) o veľkosti b , ktoré sú generované intenzitou r . Vstupujúci paket je ďalej preposlaný do siete, ak sa v token buckete nachádza aspoň jedno povolenie. V prípade, že zásobník je prázdny, prichádzajúci paket je zahodený. Token Bucket za jednotku času môže vyslať maximálne $b + r$ paketov, preto výstupný tok je ohraničený krivkou $\gamma_{b,r}(t) = b + rt$.



Obr. 1.7: Schéma zariadenia Token Bucket

Zariadenie Peak Limiter je verziou Token Bucketu s parametrom $b = 0$, to znamená, že neexistuje zásobník na povolenia a zásobník prepustí do siete len taký objem pravádzky, koľko dokáže vygenerovať povolení za časovú jednotku. Peak Limiter je vlastne obmedzovač špičkovej kapacity pre danú prevádzku. Na rozdiel od Token Bucketu neumožňuje "našetriť" si kapacitu počas prázdných prechádzok v toku. V oboch zariadeniach dochádza k stratám paketov v prípade, keď nie je k dispozícii žiaden token, preto ohraničenie pre výstupný tok musí byť menšie, nanajvýš rovné ohraničeniu vstupného toku.

PRÍKLAD 1.2

Nech daný paketový tok o dĺžke $n = 15$ prechádza cez Peak Limitér, ktorý umožňuje vyslať do siete len prevádzku so špičkovou intenzitou $r = 4$ pakety za časovú jednotku. Určíme prírastky výstupného toku $b(t)$ a lineárne ohraničenie pre výstupný proces $B(t)$:

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$a(t)$	2	3	3	4	6	6	5	4	1	2	3	2	2	1	1
$b(t)$	2	3	3	4	4	4	4	4	1	2	3	2	2	1	1
$B(t)$	2	5	8	12	16	20	24	28	29	31	34	36	38	39	40

Parametre výstupného toku sú

$$\lambda_{peak} = 4, \quad \lambda_{avg} = \varrho = \frac{40}{15} = 2.667, \quad \sigma = \max_{0 \leq t \leq n} [B(t) - \varrho t] = 6.667$$

Burst perióda vo výstupnom toku $B(t)$ nastáva v rovnakom časovom okamihu $t = 8$ ako pri vstupnom toku $A(t)$, pričom dosahuje nižšiu hodnotu vďaka obmedzeniu špičkovej intenzity peak limitérom. Výstupný tok je ohraničený priamkou $\beta(t) = 6.667 + 2.667t$, pričom platí $\beta(t) < \alpha(t), \forall t$.

PRÍKLAD 1.3

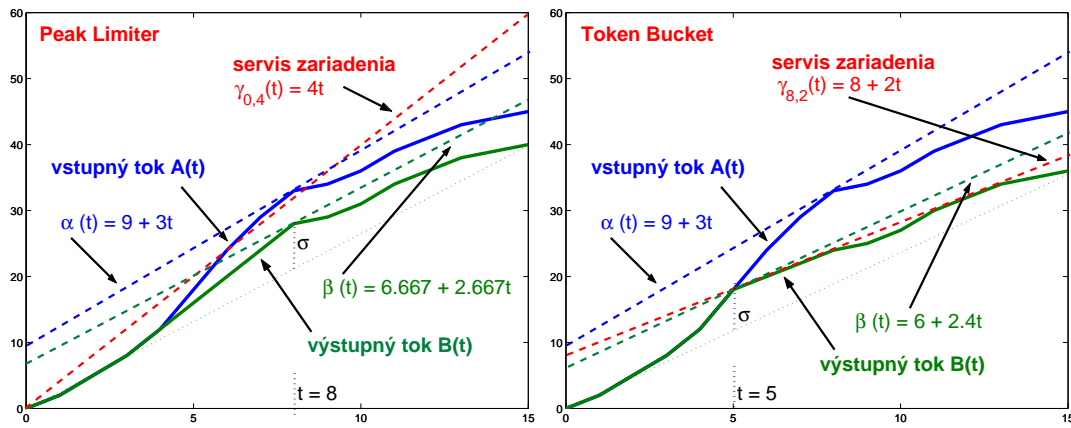
Nech rovnaký paketový tok ako v predchádzajúcom príklade, prechádza cez Token Bucket s parametrami $r = 2$ a $b = 8$. Premenná $tb(t)$ bude popisovať počet tokenov v zásobníku v danom časovom slot. Určíme parametre a lineárne ohraňenie pre výstupný proces $B(t)$:

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$a(t)$	2	3	3	4	6	6	5	4	1	2	3	2	2	1	1
$tb(t)$	8	7	6	4	0	0	0	0	1	1	0	0	0	1	2
$b(t)$	2	3	3	4	6	2	2	2	1	2	3	2	2	1	1
$B(t)$	2	5	8	12	18	20	22	24	25	27	30	32	34	35	36

Parametre výstupného toku sú

$$\lambda_{peak} = 6, \quad \lambda_{avg} = \varrho = \frac{36}{15} = 2.4, \quad \sigma = \max_{0 \leq t \leq n} [B(t) - \varrho t] = 6$$

Výstupný tok $B(t)$ je ohraňený $\beta(t) = 6 + 2.4t$, pričom platí $\beta(t) < \alpha(t)$.



Obr. 1.8: Ohraňenie výstupného toku funkciou $\beta(t)$ pri prechode cez Peak Limiter a Token Bucket

Na predchádzajúcich obrázkoch sú zobrazené vstupné a výstupné toky pre zariadenia Peak Limiter a Token Bucket. Výstupné toky sú samozrejme navzájom odlišné. V prvom prípade sme obmedzili iba špičkovú intenzitu, preto ak hodnota prírastkov je menšia alebo rovná $r = 4$, výstupný tok kopíruje v čase vstupný tok. Burst perióda nastane v rovnakom časovom okamihu, ale s menšou hodnotou. V druhom prípade sme okrem povolenej intenzity $r = 2$ mohli využívať aj povolenia z token bucketu, $b = 8$. To znamená, že špičková intenzita toku môže byť až $\lambda_{avg} = 10$ za podmienky, že predtým bola dostatočne dlhý čas slabá prevádzka, menšia alebo rovná ako $r = 2$. Úlohu bucketu na povolenia môžeme chápať ako nejakú rezervu, ktorá sa v prípade príchodu relatívne veľkej záťaže (burst periódy) oproti intenzite r do zariadenia.

V obidvoch zariadeniach dochádza k stratám paketov. V prípade presného merania môžeme počet stratených paketov vyrátať presne rozdielom $A(t) - B(t)$. Pri praktických situáciách môžeme určiť iba horný odhad strát.

Na obrázkoch ďalej vidíme, že pre výstupné toky máme niekoľko lineárnych ohraňení. Funkciu $\beta(t)$ sme získali priamo pomocou výpočtu parametrov toku.

Samozrejme výstupné toky ohraňuje funkcia $\alpha(t)$, ktorá pôvodne popisovala vstupné toky $A(t)$, pričom platí $\beta(t) < \alpha(t), \forall t$. Pribudlo nám ešte ohraňenie pomocou funkcie $\gamma_{b,r}(t)$, ktoré popisuje servis daného zariadenia, ktoré v čase t_0 môže vyslať do siete maximálne $b + rt_0$ paketov. Preto výstupný tok je ohraňovaný aj priamkou $\gamma_{b,r}(t)$, ktorá ale už nie je v jednoznačnom vzťahu s funkciami $\alpha(t)$ a $\beta(t)$, preto nám zostáva na výber, ktorú z uvedených priamok je rozumnejšie ponechať ako ohraňenie výstupného toku.

1.1.2 Lineárne ohraňenie pre periodický tok

V praxi sa môžeme stretnúť s tokmi, ktoré vykazujú istý periodický charakter prenášaných paketov. V ideálnom prípade môžeme uvažovať o deterministickom periodickom toku, ktorého prírastky v čase sa po nejakej perióde p opakujú.

Periodický tok

Vstupný proces $A(t)$ nazveme periodický s periódou $p > 0$, ak pre prírastky toku $a(i)$ platí

$$t > 0; \quad a(t + p) = a(t)$$

Určíme horné ohraňenie pre rozdiel $A(t) - A(s)$. Číslo $A(p)$ predstavuje počet paketov, ktoré sa vyskytli počas periódy p . Funkcia $\lceil x \rceil$ zaokrúhli reálne číslo x na najmenšie celé číslo, ktoré je väčšie alebo rovné číslu x .

$$A(t) - A(s) \leq \left\lceil \frac{t-s}{p} \right\rceil A(p) \leq \left(\frac{t-s}{p} + 1 \right) A(p) = A(p) + \frac{A(p)}{p}(t-s)$$

Využili sme horné ohraňenie pre počet periód p , ktoré sa nachádzajú v intervale o dĺžke $t - s$. Odvodenie si vysvetlíme na nasledujúcom príklade.

PRÍKLAD 1.4 Majme daný periodický tok $A(t)$ s periódou p .

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	...
$a(t)$	16	14	4	3	2	1	1	0	16	14	4	3	2	1	1	...
$A(t)$	16	30	34	37	39	40	41	41	57	71	75	78	80	81	82	...

Dĺžka periódy je $p = 8$ a počas periódy sa vyskytne $A(8) = 41$ paketov. Budeme uvažovať o rozdele hodnôt napríklad $A(35) - A(14)$:

$$A(35) - A(14) = [4 \cdot A(8) + A(3)] - [A(8) + A(6)] = 117$$

Rozdiel ohraňíme pomocou zaokrúhlenia počtu periód v intervale o dĺžke $t - s$ smerom nahor:

$$A(35) - A(14) = 117 \leq \left\lceil \frac{35-14}{8} \right\rceil A(8) = \lceil 2.625 \rceil A(8) = 3 \cdot A(8) = 123$$

V poslednom kroku zväčšíme presný počet periód o jednotku:

$$A(35) - A(14) = 117 \leq 123 \leq (2.625 + 1) \cdot A(8) = 148.625$$

Číslo 148.625 je hodnota horného ohraničenia toku $A(t)$ v bode $t - s = 21$:

$$\alpha_p(t-s) = A(p) + \frac{A(p)}{p}(t-s) = 41 + 5.125 \cdot (t-s) \Rightarrow \alpha_p(21) = 148.625$$

ohraničenie periodického toku

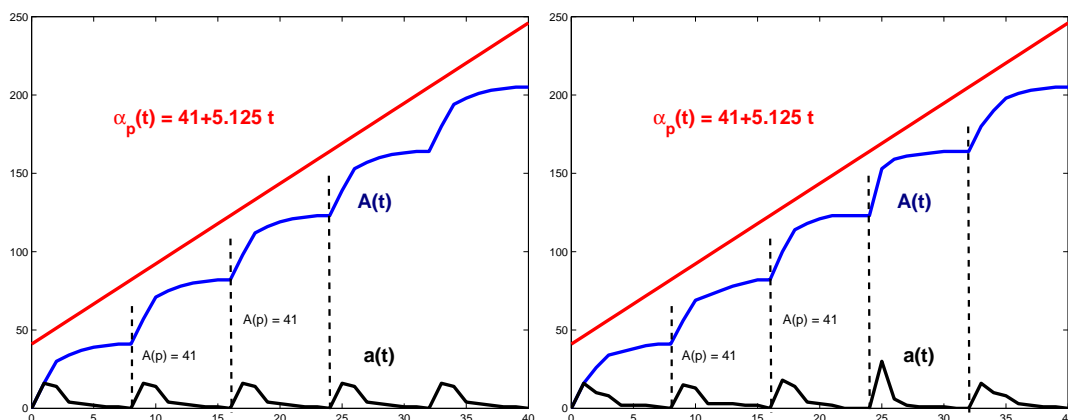
Nech je vstupný proces $A(t)$ periodický s periódou $p > 0$. Proces $A(t)$ je zhora ohraničený priamkou $\alpha_p(t)$:

$$A(t) - A(s) \leq \alpha_p(t-s) = A(p) + \frac{A(p)}{p}(t-s) = \sigma + \varrho(t-s)$$

kde $\sigma = A(p)$ je burst parameter a $\varrho = A(p)/p$ priemerná intenzita toku.

Na určenie burst parametra sme nepoužili namerané hodnoty, ale sme využili periodicitu toku. Pri priamom výpočte musí mať meranie dĺžku celočíselného násobku periódy p , v opačnom prípade môžeme získať zavádzajúce ohraničenie v závislosti na tom, v akej časti periódy bolo meranie ukončené.

Uvažovaný periodický tok je veľmi idealistický. V praxi dochádza k narušeniu periodicity vďaka rôznym náhodným faktorom. Napríklad výskyt paketov počas jednej periódy je náhodný, ale celkový počet paketov v periódach ostáva rovnaký. Aj v takomto prípade je použitie funkcie $\alpha_p(t)$ vhodným ohraničením:

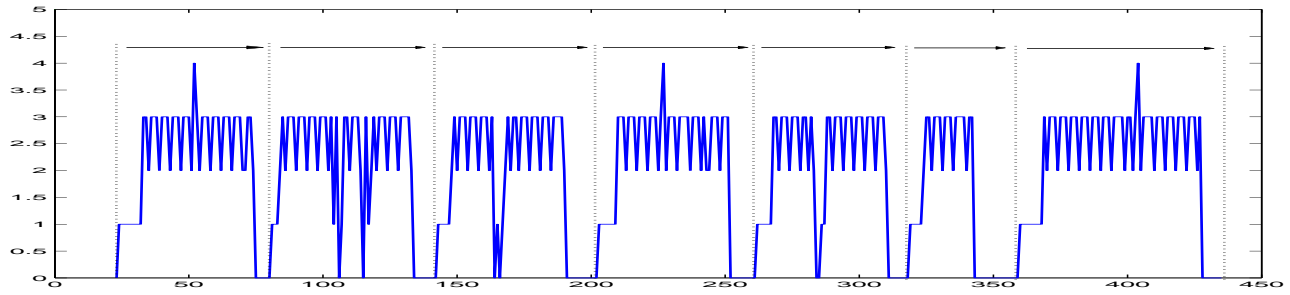


Obr. 1.9: Lineárne ohraničenie dvoch typov periodického toku s rovnakou hodnotou $A(p)$

Na druhom grafe vidíme, že rovnaké ohraničenie $\alpha_p(t)$ ostalo odolné voči náhodnému výskytu paketov počas jednotlivých periód.

V reálnej prevádzke sa môže tok, v ktorom výskyt paketov je náhodný bez obmedzenia na nejakú danú dĺžku periód, javiť v istom zmysle ako periodický. Na nasledujúcom obrázku je ukážka IPTV toku, ktorý predstavuje program Markíza zo zdroja ST-COM Mágio.

Časový slot má dĺžku 10 ms, celková dĺžka ukážky je 4500ms=4.5s, prírastky toku v slote majú hodnoty $a(i) = 0, 1, \dots, 4$. Vidíme, že tok je v istom zmysle periodický. Tok sme rozdelili do 7 periód, pričom za ukončenie periódy sme považovali sekvenciu viacej než 4 núl. Nová perióda potom začína prvým nenulým



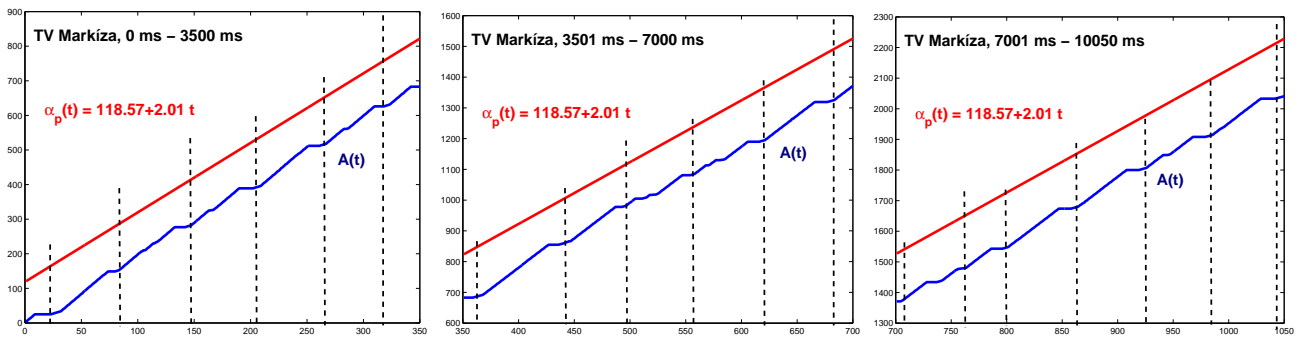
Obr. 1.10: Záznam IPTV Markíza zo zdroja ST=COM Mágioa o dĺžky 4.5s

prírastkom. V nasledujúcej tabuľke sú zaznamenané veľkosti jednotlivých periód p_i , počet paketov v perióde $A(p_i)$ a intenzita výskytu paketov ϱ_i v perióde:

i - perióda	1	2	3	4	5	6	7
p_i [ts=10ms]	57	62	60	59	57	41	77
$A(p_i)$ [paket]	124	128	112	123	114	57	172
ϱ_i [paket/ts]	2.17	2.06	1.87	2.08	2.00	1.39	2.23

Pre lineárne ohraňenie $\alpha_p(t)$ periodického toku použijeme priemery z nameovaných hodnôt, $\hat{p} = 59$ [ts] a $\hat{A}(p) = 118.574$ [paket]:

$$\alpha_p(t) = \hat{A}(p) + \frac{\hat{A}(p)}{\hat{p}} t = 118.574 + 2.0097t$$



Obr. 1.11: Lineárne ohraňenie pre periodický tok TV Markíza o dĺžke 10,05s

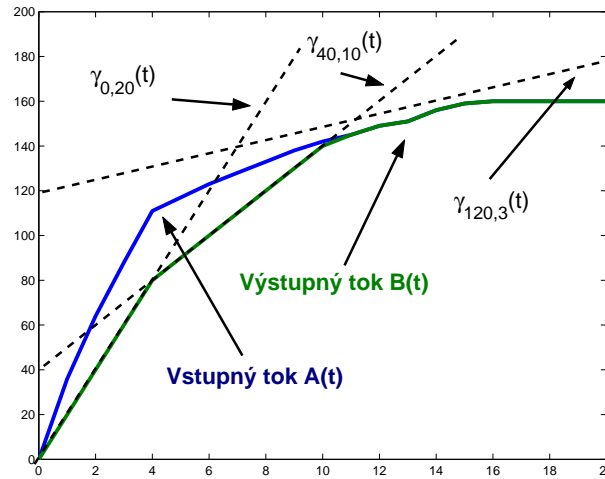
Aj keď sme na určenie funkcie $\alpha_p(t)$ použili iba prvých 7 periód toku (5.4 s), vďaka nemennosti toku (stacionarite), ohraňenie je vhodné pre celú dĺžku merania. Vďaka hodnote burst parametra funkcia $\alpha_p(t)$ garantuje horné ohraňenie s rezervou o veľkosti počtu výskytov paketov počas jednej periódy.

1.1.3 Subaditívne vstupné krivky

V predchádzajúcich kapitolách sme ukázali, že pre tok vychádzajúci z daného zariadenia môžeme určiť niekoľko lineárnych ohraňení $\gamma_{i,j}(t)$. Pre zachovanie linearitu musíme z nich vybrať jedno najvhodnejšie. Avšak pre získanie lepšieho modelu pre tok je optimálne použiť minimum z ohraňení $\gamma_{i,j}(t)$.

Na obrázku je znázornený prechod vstupného toku $A(t)$ cez tri Token Buckety, ktorých práca je popísaná pomocou funkcií $\gamma_{i,j}(t)$. Výstupný tok $B(t)$ je ohraňovaný všetkými troma funkciami $\gamma_{i,j}(t)$. Pre získanie presnejšieho modelu pre výstupný tok použijeme funkciu $\alpha(t)$, ktorá je funkcionálnym minimom všetkých ohraňujúcich funkcií,

$$\alpha(t) = \min [\gamma_{0,20}(t), \gamma_{40,10}(t), \gamma_{120,3}(t)],$$



Obr. 1.12: Prechod toku $A(t)$ cez tri Token Buckety

subaditívna funkcia

Funkcia $f(x)$ je subaditívna, ak platí: $\forall s, t \geq 0; \quad f(t+s) \leq f(t) + f(s)$

PRÍKLAD 1.5

Funkcia $f(x) = x$ je subaditívna:

$$f(t+s) = t+s \leq f(t) + f(s)$$

Funkcia $f(x) = x^2$ nie je subaditívna:

$$f(t+s) = (t+s)^2 = t^2 + 2ts + s^2 \geq t^2 + s^2 = f(t) + f(s)$$

Funkcia $f(x) = \sqrt{x}$ je subaditívna, overíme nerovnosť

$$\sqrt{t+s} \leq \sqrt{t} + \sqrt{s} \quad \Rightarrow \quad t+s \leq t + 2\sqrt{ts} + s \quad \Rightarrow \quad 0 \leq 2\sqrt{ts}$$

subaditívna vstupná krivka

Subaditívnu funkciu $\alpha(t)$ nazveme subaditívna vstupná krivka toku $A(t)$, ak

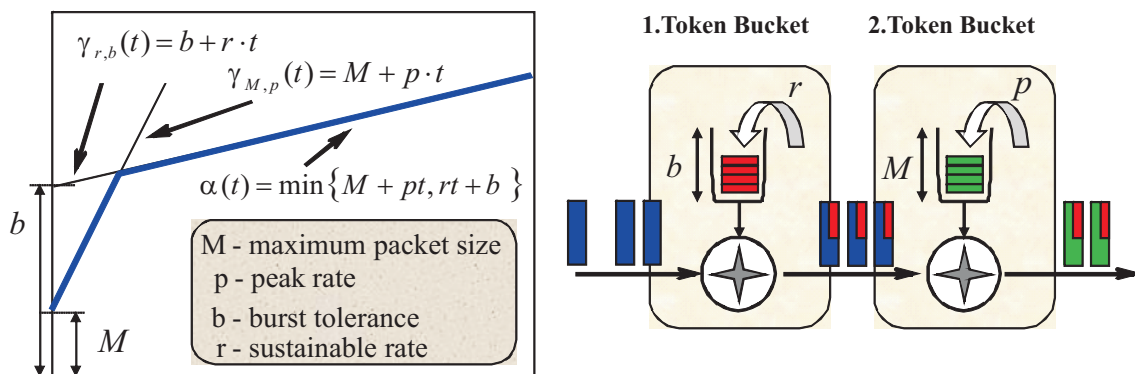
$$\forall s, t \geq 0; \quad 0 \leq s \leq t; \quad A(t) - A(s) \leq \alpha(t-s)$$

Treba si uvedomiť, že všetky funkcie $\alpha(t)$, ktoré sú funkcionálnym minimom lineárnych ohraňčení $\gamma_{b_i, r_i}(t)$, $\alpha(t) = \min_{\forall i} [\gamma_{b_i, r_i}(t)] = \min_{\forall i} [b_i + r_i t]$ sú konkávne funkcie, a dá sa dokázať, že konkávna funkcia je aj subaditívna.

Podľa odporúčaní ITU-T [3] sa pre služby IntServ používa tzv. T-SPEC (špecifikácia) pomocou parametrov: M - maximálna veľkosť paketu, p - špičková intenzita, b - burst parameter a r - priemerná intenzita.

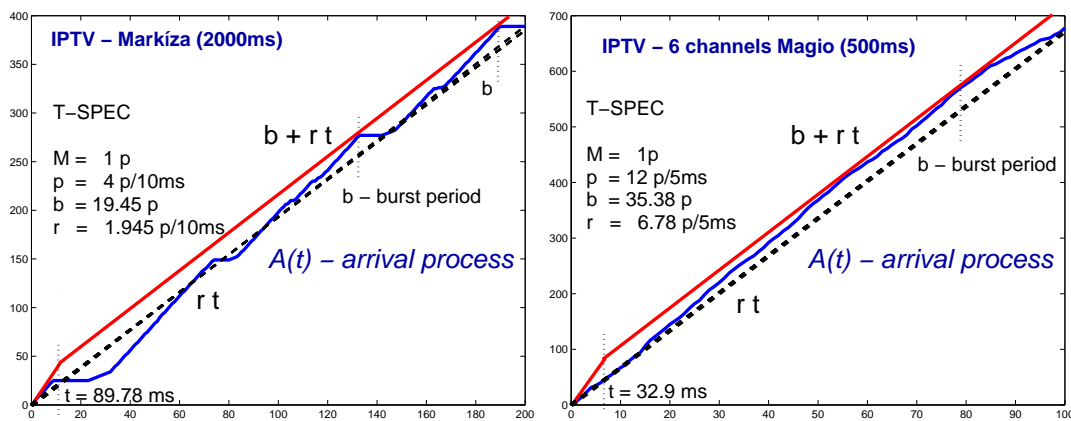
Vstupný tok je vlastne riadený dvoma Token Bucketmi, ktoré ohraňčujú tok krivkami $\gamma_{M,p}(t) = M + p \cdot t$ a $\gamma_{b,r}(t) = b + r \cdot t$. Výsledná krivka je subaditívna:

$$\alpha(t) = \min [M + p \cdot t, b + r \cdot t] \quad \text{resp.} \quad \alpha = \gamma_{M,p} \oplus \gamma_{b,r} \quad (\oplus - \text{minimum})$$



Obr. 1.13: T-špecifikácia pre služby IntServ, obmedzenie toku dvoma Token Bucketmi

Na obrázkoch je ukážka určenia T-špecifikácie pre dve krátke vzorky TV Markíza a 6 kanálov TV. Vďaka použitému kódovaniu v zdroji boli veľkosti paketov rovnaké, preto sme nastavili parameter $M = 1$.



Obr. 1.14: T-špecifikácia pre dva rôzne typy IPTV tokov

V rámci projektu [6] sme pracovali s rôznymi vzorkami záznamov IPTV zo zdroja ST-Magio o veľkosti 2 minúty. V prípade použitia celého záznamu napr. pre 6 kanálov sa zmenia len druhé dve hodnoty $b = 194p$, $r = 6p/5ms$. Priemerná intenzita r ostala v podstate zachovaná, zväčšil sa burst parameter b

oproti polsekundovému záznamu, počas 2 minútového záznamu sa pravdepodobne vyskytla v toku relatívne veľká dátová vlna.

V tejto kapitole sme sa zaoberali deterministickým popisom IP tokov. Najjednoduchší spôsob je použiť lineárne ohraňenie, ktorého parametre, stredná hodnota a burst perióda, vyplývajú buď z charakteru toku, alebo ho ovplyvňujú obmedzovače, ktoré riadia vstup toku do siete. Pre presnejší popis toku sme zaviedli subaditívne vstupné krivky, ktoré budú neskôr tvoriť základ pre použitie Network Kalkulu. Tento nástroj umožňuje z deterministického hľadiska analyzovať a dimenzovať rôzne situácie v sieti.

Vo všeobecnosti však použitie deterministických horných ohraňení môže byť veľmi štedré pri pridelení nutných prostriedkov na prevádzku. Napríklad ojedinelý výskyt burst periódy môže spôsobiť neadekvátne zvýšenie horného ohraňenia, pričom však potlačenie (zahodenie paketov) tejto periódy by možno spôsobilo len minimálne straty, ktoré by boli ďaleko pod hranicou požadovanej kvality obsluhy QoS.

Použitie deterministických ohraňení je jednoduché, efektívne a často plne postačujúce. V snahe však získať podrobnejší popis paketových tokov, napríklad kvôli presnejšiemu dimenzovaniu zariadení, siahneme po stochastických modeloch. Na ich určenie potrebujeme analýzu pravdepodobnostných charakteristík meraných tokov. Základné typy stochastických modelov uvedieme v nasledujúcej kapitole.

1.2 BERNOULLIHO MODEL IP TOKU

Najjednoduchší stochastický model paketového toku, resp. IP prevádzky, predstavuje Bernoulliho proces. Je definovaný dvoma vlastnosťami - výskyt udalostí (paketov, rámcov, buniek, ATM cells, bitov ...) v toku je navzájom nezávislý, a pravdepodobnosť výskytu udalosti v časom slotu sa v čase nemení, čiže je pre všetky sloty rovnaká. Predpokladáme, že v časovom slotu sa môže vyskytnúť iba jedna udalosť. Tento predpoklad je odôvodnený samotnou technológiou IP siete, v kde na najspodnejšej fyzickej vrstve sa nemôžu naraz vyskytnúť 2 bity, bez ohľadu na to, aký rýchly prenos je. Časový slot, v ktorom môže nastať výskyt maximálne jedného paketu (rámca, bunky), budeme nazývať *elementárny slot*.

Bernoulliho proces

Bernoulliho proces nazveme stochastický proces, v ktorom sa je výskyt paketov nezávislý s rovnakou pravdepodobnosťou p pre každý elementárny slot. Nech $q = 1 - p$. Prírastky toku $a(i)$ v elementárnom časovom slotu sú náhodné premenné s alternatívnym rozdelením:

$$\Pr(a(i) = k) = p^k(1 - p)^{1-k} = p^k q^{1-k}, \quad k = 0, 1$$

Náhodnú premennú $a(i)$ nazývame tiež Bernoulliho premenná resp. Bernoulliho indikátor (nadobúda hodnoty 0 a 1, čiže indikuje výskyt paketu v slotu), alebo hovoríme, že prírastky toku $a(i)$ majú alternatívne rozdelenie, $a(i) \sim \text{Alt}(p)$.

Počet udalostí (napr. paketov resp. bitov) v časovom slotu dĺžky n (jeden slot predstavuje súčet n elementárnych slotov) popíšeme pomocou náhodnej premennej $A(n)$. Z elementárnej pravdepodobnosti vieme, že jej rozdelenie je binomické:

Binomické rozdelenie

Pravdepodobnosť, že náhodná premenná $A(n)$, ktorá popisuje počet výskytu paketov n elementárnych slotov, nadobudne hodnotu k , sa riadi binomickým rozdelením:

$$A(n) \sim Bi(n, p); \quad \Pr(A(n) = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n$$

PRÍKLAD 1.6 Nech nasledujúca postupnosť núl a jednotiek predstavuje výskyt bitov, resp. paketov v nejakom IP toku na sieti

100101001000011111001100000101111010001110101101110000010100111010

Z týchto dát vieme odhadnúť pravdepodobnosť výskytu paketu p v jednom časovom slotu. Počet všetkých vzoriek je 66 (čo je samozrejme málo, ale postačuje pre ukážku), počet výskytu paketu je 31, preto $p = 32/66 = 0.4849$. Ak by sme ten istý tok merali v dvoch časových slotoch, z predchádzajúceho merania dostaneme nasledujúci záznam:

1 1 1 0 1 0 1 2 2 0 2 0 0 1 1 2 1 1 0 2 1 1 2 1 2 0 0 1 1 0 2 1 1

Odhadneme rozdelenie pravdepodobnosti výskytu paketov v dvoch slotoch:

$$p_0 = 9/33 = 0.2727 \quad p_1 = 16/33 = 0.4849 \quad p_2 = 8/32 = 0.2424$$

Ak použijeme ako model pre meraný tok Bernoulliho proces s pravdepodobnosťou $p = 0.4849$ (a s predpokladom že výskyt paketov je nezávislý) dostávame binomické rozdelenie pravdepodobnosti:

$$p_0 = p^2 = 0.2351 \quad p_1 = 2pq = 0.4995 \quad p_2 = q^2 = 0.2653$$

Aj bez použitia štatistických metód (napr. testu dobrej zhody o rovnosti rozdelení) vidíme, že pravdepodobnosti odhadnuté z pozorovania a pravdepodobnosti získané z Bernoulliho procesu sa na seba "podobajú".

Ak by sme realizovali meranie tej istej realizácie toku v troch časových slotoch, záznam výskytu paketov by bol

1 2 1 0 2 3 1 1 0 2 3 1 1 2 2 2 2 0 1 1 3 1

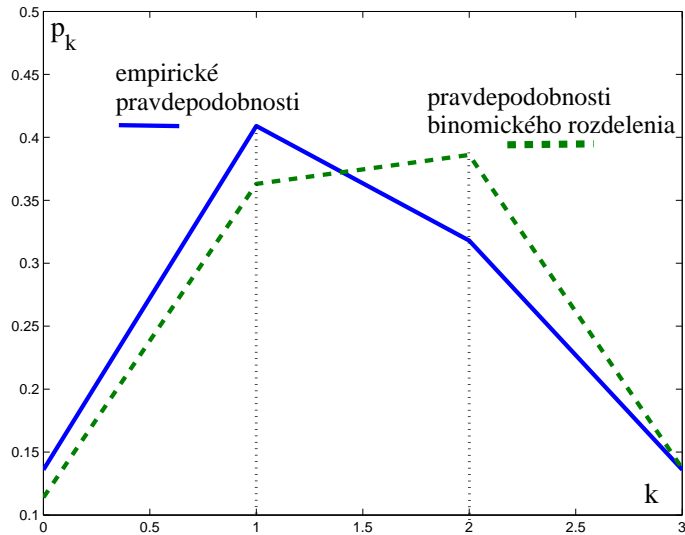
Odhadneme rozdelenie pravdepodobnosti výskytu paketov v troch slotoch:

$$p_0 = \frac{3}{22} = 0.136, \quad p_1 = \frac{9}{22} = 0.409, \quad p_2 = \frac{8}{22} = 0.318, \quad p_3 = \frac{3}{22} = 0.136$$

Použijeme Bernoulliho proces s $p = 0.4849$ ako model pre meraný tok.

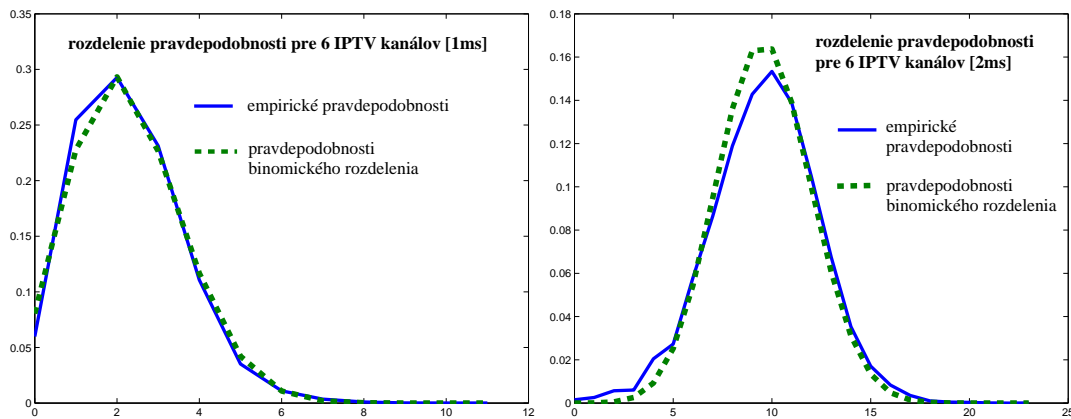
$$p_0 = p^3 = 0.114, \quad p_1 = 3p^2q = 0.363, \quad p_2 = 3pq^2 = 0.386, \quad p_3 = q^3 = 0.137$$

Aj pri malom počte dát (22) vidíme výraznú podobnosť medzi odhadnutými pravdepodobnosťami a hodnotami, ktoré sme získali z Bernoulliho procesu.



Obr. 1.15: Porovnanie odhadnutých a binomických pravdepodobností

Na nasledujúcich obrázkoch sa vrátíme k meraniam získaných zo zdroja ST-Magio, konkrétne k záznamu 6 kanálov IPTV kanálov a porovnáme tento záznam s Bernoulliho modelom IP toku.



Obr. 1.16: modelovanie toku 6 kanálov IPTV Bernoulliho procesom

Na prvom obrázku je znázornené rozdelenie pravdepodobností výskytu paketov, pričom časový slot je 1ms. Maximálny počet paketov v 1ms sa rovnal 11 poketov, a pravdepodobnosť výskytu 1 paketa je 0.2048. Namerané hodnoty sme aproximovali binomickým rozdelením $Bi(11, 0.2048)$. Na druhom obrázku sa časový slot rovnal 8ms a na aproximáciu sme použili rozdelenie $Bi(23, 0.4179)$. To znamená, že maximálne sa vyskytlo 23 paketov za 8ms, a pravdepodobnosť výskytu

paketa je 0.4179. Aj bez prevedenia štatistických testov môžeme aproximáciu v obidvoch prípadoch považovať za dostatočnú.

V predchádzajúcej kapitole sme sa zaoberali IPTV tokom Markíza, teda iba jedným kanálom. Tento tok vykazoval nezameniteľné periodické vlastnosti, aj keď dĺžka periódy a intenzita toku v periódach boli náhodné. Periodicita je typická pre IPTV toky. Na predchádzajúcom obrázku sa zaoberáme 6 kanálmi Mágia, pričom pri ujme na všeobecnosti môžeme predpokladať, že každý z kanálov môže v istej miere vykazovať periodické vlastnosti. Avšak pri súčte šiestich kanálov periodičnosť jednotlivých kanálov zaniká, a prejavuje sa Bernoulliho vlastnosť, výskyt jednotlivých paketov je navzájom nezávislý (jednotlivé pakety patria k rôznym kanálom).

1.2.1 Pravdepodobnostné charakteristiky

Základné číselné charakteristiky náhodnej premennej X sú **stredná hodnota** EX a **disperzia** DX . Pre náhodnú premennú $a(i)$, ktorá popisuje prírastky v elementárnom slotе, ich určíme veľmi jednoducho:

$$Ea(i) = 0q + 1p = p, \quad Da(i) = (0 - p)^2q + (1 - p)^2p = pq(q + p) = pq$$

Momenty prírastkov $a(i)$

Nech náhodná premenná $a(i)$ má alternatívne rozdelenie pravdepodobnosti $\Pr(a(i) = k) = p^k(1 - p)^{1-k}$, $k = 0, 1$. Potom stredná hodnota a disperzia premennej $a(i)$ sú: $Ea(i) = p$, $Da(i) = p(1 - p) = pq$

Strednú hodnotu náhodnej premennej $A(n)$, ktorá popisuje výskyt paketov v slotе o dĺžke n , môžeme získať rôznymi spôsobmi. Najbežnejší spôsob je priamo použitie definície strednej hodnoty pre diskretnú náhodnú premennú:

$$\begin{aligned} EA(n) &= \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^n k \frac{n!}{(n-k)!k!} p^k q^{n-k} = \\ &= npq \sum_{k=1}^n \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} q^{n-k+1} = npq \sum_{k=0}^{n-1} \frac{(n-1)!}{(n-1-k)!k!} p^k q^{n-k-1} = \\ &= \sum_{k=0}^n k \binom{n-1}{k} p^k q^{n-1-k} = npq(p + q)^{n-1} = npq \end{aligned}$$

Podobne, s použitím rozkladu $k^2 = k(k-1) + k$ by sme z rozdelenia vypočítali druhý moment $EA(n)^2$ a z neho disperziu $DA(n) = EA(n)^2 - (EA(n))^2$. Postup je zdĺhavý, a čitateľ ho môže nájsť v ľubovoľnej učebnici o elementárnej pravdepodobnosti

Pri znalosti elementárnej pravdepodobnosti môžeme k získanému výsledku dospieť jednoduchou elementárnou úvahou, ktorá vychádza priamo z definície Bernoulliho procesu. Pravdepodobnosť výskytu paketu v jednom slotе je p . Túto pravdepodobnosť môžeme tiež pokladať za strednú hodnotu výskytu paketu v slotе. Preto stredná hodnota výskytu paketov v n -slotoch je np .

Pre určenie disperzie $DA(n)$ použijeme tvrdenie: *disperzia súčtu nezávislých náhodných premenných je súčet disperzií jednotlivých premenných*. Keďže výskyt paketov v Bernoulliho procese je nezávislý, disperzia $A(n)$ je súčtom disperzií výskytu paketov v jednotlivých slotoch:

**Momenty
náhodnej
premennej
 $A(n)$**

Nech náhodná premenná $A(n)$ má binomické rozdelenie pravdepodobnosti $\Pr(A(n) = k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, \dots, n$. Potom stredná hodnota a disperzia premennej $A(n)$ sú: $EA(n) = np$, $DA(n) = npq$

Náhodnú premennú jednoznačne popisuje jej rozdelenie pravdepodobnosti, resp. distribučná funkcia. Ďalšími spôsobmi, ako popísať náhodnú premennú je použitie rôznych vytvárajúcich funkcií. Najčastejšie sa uvádza Momentová vytvárajúca funkcia $\varphi(\theta)$ (*MGF - moment generation function*), ktorá je strednou hodnotou parametrickej exponenciálnej transformácie náhodnej premennej. My sa budeme zaoberať kumulatívnu vytvárajúcou funkciou, ktorá je prirodzeným logaritmom MGF, pretože priamo súvisí s dimenzovaním kapacity linky pomocou efektívnej šírky pásma, s ktorým sa stretneme v neskorších kapitolách.

**Kumulatívna
vytvárajúca
funkcia CGF**

Kumulatívna vytvárajúca funkcia (*CGF - cumulant generation function*), pre nezápornú náhodnú premennú X , má tvar:

$$\lambda(\theta) = \ln \varphi(\theta) = \ln E[e^{\theta X}] = \ln E \left[\sum_{\forall k} e^{\theta k} p_k \right]$$

Kumulatívna vytvárajúca funkcia pomocou n -tých derivácií kumuluje, resp. vytvára lineárne kombinácie n -tých počiatočných momentov premennej X . Nám bude stačiť informácia, že prvá derivácia v nule je stredná hodnota EX a druhá derivácia v nule je disperzia $DX = EX^2 - (EX)^2$.

**Výpočet
momentov
pomocou
CGF**

$$\lambda'(0) = \ln E[e^{\theta X}]|_{\theta=0} = \frac{E[Xe^{\theta X}]}{E[e^{\theta X}]} \Big|_{\theta=0} = EX$$

$$\lambda''(0) = \frac{E[X^2 e^{\theta X}] E[e^{\theta X}] - E[Xe^{\theta X}]^2}{E[e^{\theta X}]^2} \Big|_{\theta=0} = EX^2 - (EX)^2 = DX$$

PRÍKLAD 1.7

Určíme CGF $\lambda(\theta)$ pre náhodnú premennú $a(i)$:

$$\lambda(\theta) = \ln \left[e^{\theta \cdot 0} P(a(i) = 0) + e^{\theta \cdot 1} P(a(i) = 1) \right] = \ln [q + pe^\theta] \quad (1.1)$$

Vypočítame prvú a druhú deriváciu v bode $\theta = 0$:

$$\lambda'(0) = \frac{pe^\theta}{q + pe^\theta} \Big|_{\theta=0} = p, \quad \lambda''(0) = \frac{pe^\theta(q + pe^\theta) - (pe^\theta)^2}{(q + pe^\theta)^2} \Big|_{\theta=0} = pq$$

Dostali sme samozrejme strednú hodnotu $Ea(i)$ a disperziu $Da(i)$.

PRÍKLAD 1.8

Určíme CGF $\lambda(\theta)$ pre náhodnú premennú $A(n)$:

$$\begin{aligned} \lambda(\theta) &= \ln \left[\sum_{k=0}^n e^{\theta k} \binom{n}{k} p^k q^{n-k} \right] = \ln \left[\sum_{k=0}^n \binom{n}{k} (pe^\theta)^k q^{n-k} \right] = \ln [q + pe^\theta]^n \\ &\Rightarrow \lambda(\theta) = n \ln [q + pe^\theta] \end{aligned} \quad (1.2)$$

Vypočítame prvú a druhú deriváciu v bode $\theta = 0$:

$$\lambda'(0) = n \frac{pe^\theta}{q + pe^\theta} \Big|_{\theta=0} = np, \quad \lambda''(0) = n \frac{pe^\theta(q + pe^\theta) - (pe^\theta)^2}{(q + pe^\theta)^2} \Big|_{\theta=0} = npq$$

Dostali sme samozrejme strednú hodnotu $EA(n)$ a disperziu $DA(n)$.

Výsledok potvrdil vetu: *CGF funkcia súčtu nezávislých náhodných premenných je súčtom CGF funkcií jednotlivých premenných.* Nech $X = \sum X_k$,

$$\lambda_X(\theta) = \ln E \left[e^{\theta \sum_k X_k} \right] = \ln E \left[\prod_k e^{\theta X_k} \right] = \ln \prod_k E \left[e^{\theta X_k} \right] = \sum_k \ln E \left[e^{\theta X_k} \right]$$

**CGF pre
Binomické
rozdelenie**

Kumulatívna vytvárajúca funkcia náhodnej premennej, ktorá má Binomické rozdelenie $A(n) \sim Bi(n, p)$ má tvar

$$\lambda(\theta) = n \ln [q + pe^\theta]$$

Ďalšia funkcionálna charakteristika náhodnej premennej, ktorú budeme používať, je **distribučná funkcia** $F(x)$:

$$F(x) = P(X \leq x) = \sum_{k; k \leq x} p_k$$

Pre náhodnú premennú K^n má tvar:

$$F(x) = P(A(n) \leq x) = \sum_{k=0}^x \binom{n}{k} p^{n-k} q^k$$

Pri úlohách o dimenzovaní má význam tzv. doplnková distribučná funkcia

$$\bar{F}(x) = 1 - F(x) = P(X > x)$$

ktorá vyjadruje pravdepodobnosť prekročenie istej "kritickej" hodnoty x pre daný proces, ktorý je popísaný náhodnou premennou X .

PRÍKLAD 1.9 Cez informačný kanál prenášame 8 bitové slová. Pravdepodobnosť chybného prenosu je 0.1, pričom výskyt chyby v jednotlivých bitoch sa navzájom neovplyvňuje. Kodek na prímači dokáže opraviť 2 chybné bity. Aká je pravdepodobnosť úspešného prenosu?

V tomto prípade, vďaka nezávislosti, Bernoulliho proces tvorí výskyt jednotlivých chýb v bitoch. Ich počet budeme modelovať náhodnou premennou E , ktorá sa riadi binomickým rozdelením $Bi(8, 0.1)$:

$$p_k = \Pr(E = k) = \binom{8}{k} 0.1^k 0.9^{8-k}$$

Nech udalosť A predstavuje nastatie úspešného prenosu, to znamená, že všetky bity sú bezchybné. Kodek dokáže opraviť maximálne 2 chybné bity

$$\Pr(A) = \Pr(E \leq 2) = p_0 + p_1 + p_2 = 0.9^8 + 8 \cdot 0.9^7 \cdot 0.1 + 28 \cdot 0.9^6 \cdot 0.1^2 = 0.9619$$

Pravdepodobnosť úspešného prenosu, resp. výskytu maximálne dvoch chybných bitov v 8-bitovom slove, je 0.9619.

PRÍKLAD 1.10 Majme nameraný nejaký reálny IP tok, ktorý má formu záznamu počtu paketov, resp. prírastkov toku $a(i) = 0, 1, \dots, 4$, v časovom slotu o dĺžke 20 ms:

4100200000 0012444444 0142300104 0332110002 4103401124 4330034400

Jednoduchým spôsobom získame dva základné parametre na popis toku:

$$\lambda_{peak} = 4 \text{ [paket]}, \quad \lambda_{avg} = \frac{102}{60} = 1.7000 \text{ [paket/20ms]}$$

Chceme overiť, či Bernoulliho model je vhodný na popis nameraného toku. Bernoulliho model závisí od dvoch parametrov, n a p . Vypočítame ich z parametrov nameraného toku:

$$\lambda_{peak} = n = 4 \text{ [paket]}, \quad \lambda_{avg} = np \Rightarrow p = \frac{\lambda_{avg}}{n} = 0.4250$$

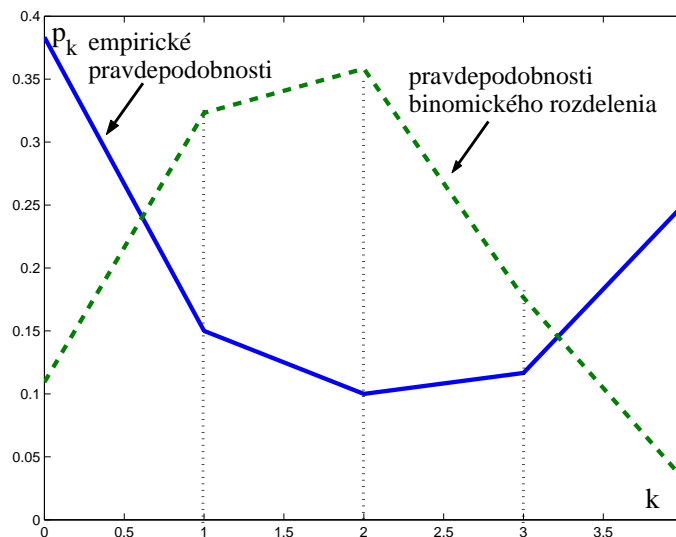
Hodnota 0.4250 je pravdepodobnosť výskytu paketu elementárnom časovom slotu, ktorý predstavuje jednu štvrtinu časového slotu meraného toku, našom

príklade teda 5 ms. Maximálny počet paketov je 4 za 20 ms, teda paket sa vyskytol v každom elementárnom 5 ms slot. Tak to predpokladá model, v realite nemusí byť 20 ms rovnomerne rozdelený na elementárne sloty. Meranie je ale obmedzené na 20 ms, neumožňuje analyzovať proces v rámci 20 ms, preto predpoklady modelu o rovnosti elementárnych tokov môžeme akceptovať.

V tabuľke uvedieme empirické pravdepodobnosti \bar{p}_k vypočítané z dát a pravdepodobnosti binomického rozdelenia p_k s odhadnutými hodnotami n a p :

k	0	1	2	3	4
\bar{p}_k	0.3833	0.1500	0.1000	0.1167	0.2500
p_k	0.1093	0.3232	0.3583	0.1766	0.0326

Medzi pravdepodobnosťami panuje minimálna zhoda, preto s istotou môžeme tvrdiť, že nameraný tok neodpovedá Bernoulliho modelu.



Obr. 1.17: Porovnanie odhadnutých a binomických pravdepodobností

V praxi, napríklad pri použití programu Wireshark, môžeme dostať záznam reálneho toku vo forme hodnôt, ktoré predstavujú doby výskytov paketov. Pri vhodnej voľbe časového slotu môžeme dostať záznam prírastkov toku s vopred zvoleným píkom. Ak by sme za veľkosť časového slotu zobrali veľkosť minimálnej medzery, dostali záznam iba 0 a 1. V takomto prípade môžeme analyzovať medzery medzi paketmi v meranom toku.

1.2.2 Rozdelenie pravdepodobnosti medzier

Medzeru v Bernoulliho procese budeme definovať ako postupnosť núl (v elementárnom slotu sa nevyskytol paket), ktorú ukončí výskyt jednotky (paketu). Napríklad v postupnosti 110101000111 sa vyskytujú dve medzery o dĺžke 1, a jedna medzera o dĺžke 3. Pravdepodobnosť výskytu medzery o dĺžke 3 je vďaka nezávislosti rovná $\Pr(T = 3) = q^3p$. Pre úplnosť pravdepodobnostného modelu musíme uvažovať aj o medzere dĺžky nula, ktorej pravdepodobnosť je $\Pr(T = 0) = p$.

**Rozdelenie
pravdepodob-
nosti
medzier**

Nech náhodná premenná T popisuje veľkosť medzery v Bernoulliho toku. Rozdelenie pravdepodobnosti n.pr. T je geometrické:

$$T \sim Geo(q); \quad P(T = t) = (1 - p)^t p = q^t p, \quad t = 0, 1, 2, \dots$$

Vypočítame strednú hodnotu dĺžky medzier:

$$ET = \sum_{t=0}^{\infty} t q^t p = p q \sum_{t=1}^{\infty} t q^{t-1} = p q \sum_{t=1}^{\infty} \frac{d}{dq} q^t = p q \frac{d}{dq} \frac{q}{1 - q} = \frac{p q}{(1 - q)^2} = \frac{q}{p}$$

Pri výpočte sme zamenili poradie sumácie a derivovania ($q < 1$) a použili sme vzťah $1 - q = p$. Pri výpočte disperzie by sme postupovali podobným spôsobom, ale technicky náročnejším, v priebehu sumácie je potrebné urobiť deriváciu druhého rádu. My si ukážeme podstatne jednoduchší spôsob výpočtu použitím kumulatívnej vytvárajúcej funkcie.

PRÍKLAD 1.11

Pomocou kumulatívnej vytvárajúcej funkcie vypočítajte strednú hodnotu a disperziu n.pr. T .

Najprv určíme CGF funkciu pre geometrické rozdelenie:

$$\begin{aligned} \lambda(\theta) &= \ln \left[\sum_{t=0}^{\infty} e^{\theta t} q^t p \right] = \ln \left[\sum_{t=0}^{\infty} (e^{\theta} q)^t p \right] = \ln \left[\frac{p}{1 - e^{\theta} q} \right] = \\ &= \ln p - \ln [1 - e^{\theta} q], \quad \theta < -\ln q \end{aligned}$$

Vypočítame prvú a druhú deriváciu v bode $\theta = 0$:

$$\lambda'(0) = \frac{e^{\theta} q}{1 - e^{\theta} q} \Big|_{\theta=0} = \frac{q}{p}, \quad \lambda''(0) = \frac{e^{\theta} q (1 - e^{\theta} q) + (e^{\theta} q)^2}{(1 - e^{\theta} q)^2} \Big|_{\theta=0} = \frac{q}{p^2}$$

Dostali sme $ET = q/p$ a $DT = q/p^2$.

**Momenty
náhodnej
premennej T**

Nech náhodná premenná T má geometrické rozdelenie pravdepodobnosti $\Pr(T = t) = q^t p$, $t = 0, 1, 2, \dots$. Stredná hodnota a disperzia sú:

$$ET = \frac{q}{p}, \quad DT = \frac{q}{p^2}$$

Určíme distribučnú funkciu geometrického rozdelenia:

$$F(t) = P(T \leq t) = 1 - P(T > t) = 1 - \sum_{k=t+1}^{\infty} q^k p = 1 - q^{t+1}$$

Doplnková distribučná funkcia má tvar:

$$\bar{F}(t) = \Pr(T > t) = 1 - F(t) = q^{t+1}$$

Geometrické rozdelenie má vlastnosť "zabúdanie na minulosť". Pravdepodobnosť, že medzera medzi paketmi bude ešte trvať viac než čas t , ak vieme že už trvala τ , nezáleží od τ :

$$P(T > t + \tau | T > \tau) = \frac{P(T > t + \tau)}{P(T > \tau)} = \frac{q^{t+\tau+1}}{q^{\tau+1}} = q^t = P(T \geq t)$$

Výsledná pravdepodobnosť je závislá iba od času t , rozdelenie "zabudlo", že medzera už nejaký čas trvala. Táto vlastnosť sa používa pri tvorbe Markovových reťazcov s diskretným časom.

Koeficient variabilnosti

Koeficientom variabilnosti náhodnej premennej T označíme $\nu_T = \frac{\sigma_T}{ET}$

Budeme predpokladať, že náhodná premenná T je nezáporná. Koeficient ν_T vyjadruje mieru variabilnosti procesu popísaného premennou T . Ak $\nu_T \rightarrow 0$, proces je "skoro" deterministický, rozptyl σ_T sa blíži k nule a medzery môžeme považovať za rovnako veľké. Ak $\nu_T = 1$, je splnená nutná podmienka pre Poissonov proces, ak $\nu_T \rightarrow \infty$, dostávame sa do oblasti ťažkých chvostov (Heavy-tailed distribution) a samopodobnosti.

Pre Bernoulliho proces platí

$$\nu_T = \frac{\sigma_T}{ET} = \frac{\sqrt{q/p^2}}{q/p} = \frac{1}{\sqrt{1-p}}$$

Keďže parameter p je pravdepodobnosť, $p \in (0, 1)$, koeficient variabilnosti pre geometrické rozdelenie je $\nu_T \in (1, \infty)$.

1.2.3 Identifikácia Bernoulliho procesu

Predchádzajúca kapitola poskytuje návod, ako určiť parameter p z 0/1 záznamu:

$$\frac{1}{\sqrt{1-p}} = \frac{\sigma_T}{ET} \Rightarrow p = 1 - \frac{(ET)^2}{DT}$$

Nutnou podmienkou pre modelovanie pomocou Bernoulliho procesu je, aby pre odhadnutú hodnotu p platilo $p \in (0, 1)$. Parameter p môžeme vypočítať aj ako strednú hodnotu prírastkov $a(i)$, $p = Ea(i)$. Ak odhadnuté hodnoty parametra podstatne odlišné, skúmaný tok netvorí Bernoulliho proces.

1.2.4 Rozdelenie pravdepodobnosti paketových zhlukov

Paketový zhluk v Bernoulliho procese budeme definovať ako postupnosť jednotiek (v elementárnom slotu sa vyskytol paket), ktorú ukončí výskyt nuly. Veľkosť zhluku budeme popisovať náhodnou premennou Z . Ide vlastne o rovnakú koncepciu ako pri úvahách od medzierách v predchádzajúcej kapitole, s tým, že nahradíme 0 za 1 a naopak. Distribúcia dĺžky paketových zhlukov má preto geometrické rozdelenie:

$$Z \sim Geo(p); \quad P(Z = z) = p^z q, \quad z = 0, 1, 2, \dots,$$

Geometrickým rozdelením sme sa zaoberali v predchádzajúcej kapitole. Popis paketových zhlukov sa od popisu medzier medzi paketmi líši iba parametrom rozdelenia $p = 1 - q$. Momenty a funkcionálne charakteristiky sme vypočítali v predchádzajúcej kapitole.

Charakteristiky náhodnej premennej Z

Nech náhodná premenná Z popisuje veľkosť paketových zhlukov v Bernoulliho toku. Náhodná premenná má potom geometrické rozdelenie pravdepodobnosti $\Pr(Z = z) = p^z q$, $z = 0, 1, 2, \dots$. Stredná hodnota, disperzia sú, kumulatívna vytvárajúca funkcia a distribučná funkcia majú tvar:

$$EZ = \frac{p}{q}, \quad DZ = \frac{p}{q^2}, \quad \lambda(\theta) = \ln q - \ln [1 - e^\theta p], \quad F(z) = 1 - p^{z+1}$$

Poznámka: V Bernoulliho procese distribúcia počtu paketov v časovom slotu o veľkosti n má rozdelenie binomické, medzery medzi paketmi a veľkosti paketových zhlukov sa riadia podľa geometrického rozdelenia, pričom medzi strednou hodnotou medzier a strednou hodnotou veľkosti paketových zhlukov je vzájomná väzba (viď nižšie). Tento fakt sa môže použiť ako nutná podmienka pri testovaní, či sa reálna prevádzka môže modelovať Bernoulliho procesom.

$$A(n) \sim Bi(n, p), \quad T \sim Geo(q), \quad Z \sim Geo(p), \quad EZ = \frac{1}{ET}$$

PRÍKLAD 1.12 V Bernoulliho toku je pravdepodobnosť výskytu paketu v 1 ms je rovná 0.7. Aká je pravdepodobnosť, že veľkosť paketového zhluku nie je väčšia ako 2?

$$\Pr(Z \leq 2) = 1 - \Pr(Z > 2) = 1 - 0.7^3 = 0.657$$

Aká je pravdepodobnosť, že nasledujú za sebou práve 3 pakety?

$$\Pr(Z = 3) = 0.7^3 \cdot 0.3 = 0.1029.$$

Aká je pravdepodobnosť, že medzera medzi paketmi je minimálne 4 ms?

$$\Pr(T \geq 4) = 0.3^4 = 0.0081$$

Aká je pravdepodobnosť, že sa v priebehu 3 ms vyskytne aspoň 1 paket?

$$\Pr(A(3) \geq 1) = 1 - \Pr(A(3) = 0) = 1 - 0.3^3 = 0.973$$

Aká je pravdepodobnosť, že sa v priebehu 6 ms vyskytnú práve 3 pakety?

$$\Pr(A(6) = 3) = \binom{6}{3} 0.7^3 \cdot 0.3^3 = 0.18522$$

Aký je stredný počet paketov v 100 ms?

Výsledok:

$$EA(100) = 100 \cdot 0.7 = 70p/100ms$$

Aká je pravdepodobnosť, že veľkosť paketového zhluky je väčšia než 3?

$$\Pr(Z > 3) = 0.7^4 = 0.2401$$

Aká je pravdepodobnosť, že veľkosť paketového zhluky je maximálne 5?

$$\Pr(Z \leq 5) = 1 - \Pr(Z > 5) = 1 - 0.7^6 = 0.88235.$$

Aká je pravdepodobnosť, že v 3 ms budú maximálne 2 pakety?

$$\Pr(A(3) \leq 2) = p_0 + p_1 + p_2 = 0.7^3 + 3 \cdot 0.7 \cdot 0.3^2 + 3 \cdot 0.7^2 \cdot 0.3 = 0.973.$$

Aká je pravdepodobnosť, že 5 ms slot nie je prázdny?

$$\Pr(A(5) \geq 0) = 1 - p_0 = 1 - 0.3^5 = 0.83193.$$

Aká je pravdepodobnosť, že medzera medzi paketmi je minimálne 6 ms a maximálne 8 ms?

$$\Pr(6 \leq T \leq 8) = 0.3^6 \cdot 0.7 + 0.3^7 \cdot 0.7 + 0.3^8 \cdot 0.7 = 0.00071.$$

Bernoulliho proces predstavuje základný model pre popis IP tokov s diskretným časom. Má dve základné vlastnosti, pravdepodobnosť výskytu paketu v časovom slotu sa v čase nemení, a výskyt paketov je navzájom nezávislý. Preto je proces stacionárny, jeho pravdepodobnostné charakteristiky nezávisia na čase, a má nezávislé prírastky. Rozdelenie pravdepodobnosti počtu paketov v n slotoch sa riadi Binomickým rozdelením, ktoré je dvojparametrické, preto pri analýze nameraného toku môžeme použiť dva základné parametre, strednú intenzitu a peak. V prípade, že máme k dispozícii podrobný záznam meraného toku so znalosťou o výskyte paketov v čase, môžeme pri analyzovaní použiť fakt, že veľkosť paketových zhlukov a medzery medzi paketami sa riadia Geometrickými rozdeleniami s navzájom doplnkovými parametrami p a q .

Opodstatnenosť použitia Bernoulliho toku je hlavne na modelovanie súčtu viacerých IP tokov. Toto tvrdenie sme demonštrovali na ukážke aproximácie rozde-

lenia pravdepodobnosti výskytu paketov v IPTV toku, ktorý predstavoval súčet 6 kanálov zo zdroja Mágio.

V súčasnosti je však Bernoulliho model často nepostačujúci, napríklad keď IP tok vykazuje istú zhlukovitosť výskytu paketov. Vhodným modelom pre takúto situáciu sa budeme zaoberať v nasledujúcej kapitole.

Literatúra

- [1] J.L.Boudec, P.Thiran *Network Calculus*, Springer Verlag, 2004
- [2] Cheng Shang Chang, *Performance Guarantees in Communication Networks*, Springer Verlag, 2000
- [3] ITU-T Recommendation Y.1315, *Global information infrastructure, Internet protocol aspect and Next-Generation Networks* 09/2006
- [4] Jozef Kalas, *Markovove reťazce - skriptum*, Polygrafické stredisko UK v Bratislave, 1993
- [5] F. P. Kelly, *Notes on Effective Bandwidth*, "Stochastic Networks: Theory and Application", pages 141-168, Oxford University Press, 1996
- [6] T-COM: *Project: Modeling of IP traffic*, Research report, 2007-2008

Register

EB - Efektívna šírka pásma, 45,
95, 96, 103

Markovov Modulovaný Poisso-

nov Proces, 103

Poissonov tok, 96, 100

THO - Teória hromadnej ob-
sluhy, 96, 101