

## Модел за определяне параметрите на алгоритмите за управление на претоварването в IP мрежи

Георги Христов, Теодор Илиев

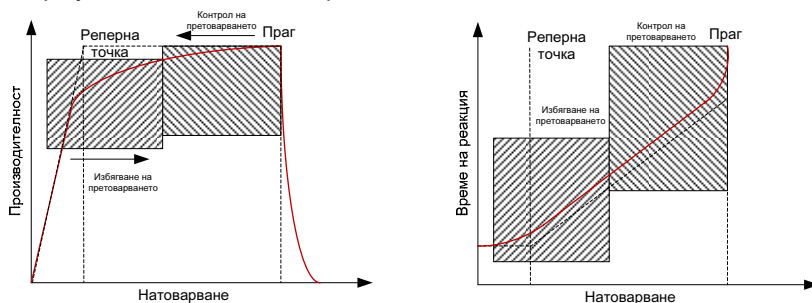
**A model for determination of the parameters of the congestion control algorithms in IP - based networks:** Congestion avoidance mechanisms allow a network to operate in the optimal region of low delay and high throughput, thereby, preventing the network from becoming congested. This is different from the traditional congestion control mechanisms that allow the network to recover from the congested state of high delay and low throughput. Both congestion control and congestion avoidance mechanisms are basically resource management problems. They can be formulated as system control problems in which the system senses its state and feeds this back to its user who adjust their controls.

**Key words:** Congestion avoidance, Congestion control, Flow control, Fairness, Network performance;

### 1. ВЪВЕДЕНИЕ

Претоварването е основният фактор в разпределението на ресурсите, който намалява ефективността при новите и старите технологии, които се използват съвместно в комуникационните мрежи. Прилагането на нови преносни среди в добре известни структури увеличава наличната честотна лента. Подобен пример е прилагането на оптични линии за връзка в локални комуникационни мрежи. Важно е да се отбележи, че те работят съвместно и със старите преносни среди, с тясна честотна лента, като например усуканата двойка кабели. Тази хетерогенност води до нарушаване на синхронизацията при обработката на заявките в обслужващите устройства, увеличават се опашките и оттам и претоварването в мрежите.

Традиционните схеми за управление на претоварването спомагат за бързо възстановяване на комуникационните системи след възникване на колизии [5]. На фиг. 1 е показана зависимостта на производителността и времето за отговор на мрежата при увеличаване на натоварването.



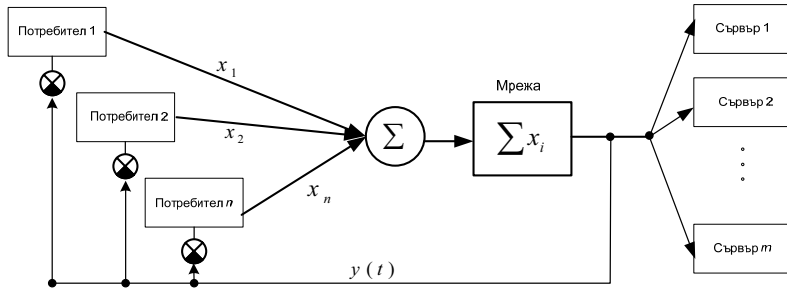
Фигура 1 Зависимост на параметрите на мрежата от натоварването

Добре проектираната схема за избягване на претоварването осигурява стабилна работа на системата около реперната точка, като следи за промяната на времето за реакция на мрежата и предприема действия за намаляване на натоварването при сигнали за излизане от това състояние.

Ресурсните устройства в мрежите следят за своето натоварване и определят дали то е под или над тяхното оптимално ниво. В зависимост от нивото на натоварване се изпраща двоичен сигнал (1 = превишаване на капацитета, 0 = под оптималното ниво) към потребителите, които с помощта на алгоритъм за увеличаване или намаляване приспособяват своя трафик към наличните ресурси. Тази двоична обратна връзка се осъществява, чрез добавяне на бит в заглавната част на пакетите [4].

## 2. МОДЕЛ НА СИСТЕМА С КОНТРОЛ НА ПРЕТОВАРВАНЕТО

Разглежда се комуникационна структура дадена на фиг.2, която се споделя от  $n$  потребители и състоянието на системата се определя от броя на пакетите в нея. Операциите, които се извършват се дефинират като дискретни във времето с много малки времеви интервали  $\Delta t$  между тях. Ако по време на интервала  $t$ ,  $i$  потребител натоварва мрежата с  $x_i(t)$ , тогава за общото натоварване на мрежата в стеснения участък се получава  $\sum x_i(t)$ , а състоянието на системата се описва, чрез  $n$  размерния вектор  $x(t) = \{x_1(t), x_2(t), \dots, x_n(t)\}$ .



Фигура 2 Модел на система за управление на  $n$  потребители, споделящи общи ресурси

Ако системата се намира в състояние около реперната точка, тогава  $x_i(t)$  на  $i$  потребител ще зависи от разпределението на ресурсите в мрежата. По време на интервалите  $\Delta t$  системата обновява своето ниво на натоварване и изпраща двоични сигнали по канала за обратна връзка  $y(t)$ , които се интерпретират от потребителите по следния начин:

$$y(t) = \begin{cases} 0 \rightarrow \text{Увеличаване на натоварването} \\ 1 \rightarrow \text{Намаляване на натоварването} \end{cases} \quad (1)$$

На базата на приетите сигнали потребителите променят интензивността на своите заявки със стойност  $u_i(t)$  и следователно:

$$x_i(t+1) = x_i(t) + u_i(t) \quad (2)$$

Промяната  $u_i(t)$  представя контрола на  $i$  потребител, която е функция на текущото състояние на потребителя и сигнала, получен по канала за обратна връзка:

$$u_i(t) = f(x_i(t), y(t)) \quad (3)$$

В системата не е възможно всеки потребител да следи промяната в интензивността на заявките на останалите и следователно  $u_i(t)$  не може да бъде изразено като функция на  $x_j(t)$ ,  $j \neq i$ . В най-общия случай управляващата функция  $f(\dots)$  притежава линеен или нелинеен характер. Съгласно (2) математическото описание на действието на функцията придобива вида:

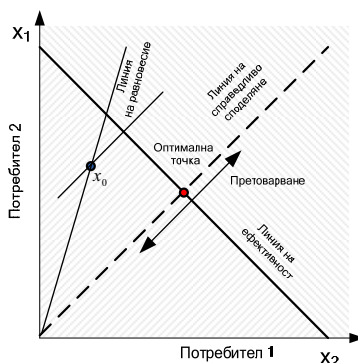
$$x_i(t+1) = \begin{cases} a_i + b_i x_i(t) & \text{ако } y(t) = 0 \rightarrow \text{Увеличение} \\ a_D + b_D x_i(t) & \text{ако } y(t) = 1 \rightarrow \text{Намаление} \end{cases} \quad (4)$$

където  $a_i$ ,  $b_i$ ,  $a_D$ ,  $b_D$  са параметри определящи схемата за управление.

### 3. ОПРЕДЕЛЯНЕ НА ПАРАМЕТРИТЕ ЗА КОНТРОЛ

Анализът на възможните схеми за линейно управление се извършва чрез използване на метод с  $n$  размерен вектор. В описания метод два потребителя използват общи системни ресурси [3].

На фиг. 3 е показано разпределените ресурси на два потребителя  $\{x_1(t), x_2(t)\}$ , представени като точки  $x_1, x_2$  в двумерното пространство. Всяка равнинна област, за която е изпълнено условието  $x_1+x_2=X$ , представлява област на ефективно действие на системата. Всяка точка, разположена върху линията на ефективност изпълнява това условие. При  $x_1=x_2$ , се наблюдава справедливо разпределение на ресурсите. Точката с координати  $(X/2, X/2)$  представлява оптималната работна точка. Целта на управлението е да приведе функционирането на системата в тази точка, без значение от началното състояние. Всички точки, затворени в пространството между линията на ефективност и координатните оси описват неефективното използване на ресурсите. Точката  $x_0$  е част от тази зона [1, 2, 7].



Фигура 3. Векторно представяне на система с два потребителя

За правилното функциониране на дадена система, е необходимо тя да реагира коректно на всеки подаден сигнал в обратната връзка. Нивата на този сигнал могат да се зададат по следния начин :

$$\begin{aligned} y(t)=0 &\rightarrow \sum x_i(t+1) > \sum x_i(t) \\ y(t)=1 &\rightarrow \sum x_i(t+1) < \sum x_i(t) \end{aligned} \quad (5)$$

Чрез дефинираните параметри за управление, (5) може да се запише по следния начин:

$$\begin{aligned} na_i + (b_i + 1) \sum x_i(t) &> 0 \quad \forall n \text{ и } \forall \sum x_i(t) \\ na_D + (b_D - 1) \sum x_i(t) &< 0 \quad \forall n \text{ и } \forall \sum x_i(t) \end{aligned} \quad (6)$$

За привеждане на системата в зоната на ефективна работа е необходимо да са изпълнени следните условия:

$$b_i > 1 - \frac{na_i}{\sum x_i(t)} \quad (7)$$

$$b_D < 1 - \frac{na_i}{\sum x_i(t)} \quad \forall n \text{ и } \forall \sum x_i(t) \quad (8)$$

Връзката между линейното управление и справедливото разпределение на ресурсите може да се изрази като:

$$F(x(t+1)) = \frac{(\sum x_i^2(t+1))}{n(\sum x_i^2(t+1))} = \frac{(\sum a + bx_i(t))^2}{n \sum (a + bx_i(t))^2} = \frac{(\sum c + x_i(t))^2}{n \sum (c + x_i(t))^2}, \quad (9)$$

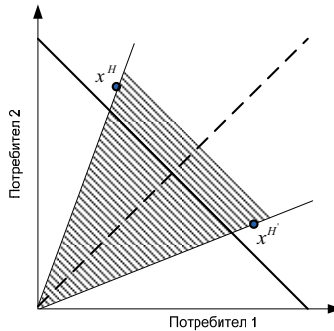
където:  $c = a/b$

Критерият за справедливост се удовлетворява при  $c > 0$  [6], което е изпълнено единствено при следните условия:

$$\begin{aligned} a_i &\geq 0, & b_i &\geq 0 \\ a_D &\geq 0, & 0 \leq b_D < 1 \end{aligned} \quad (10)$$

където коефициентите  $a_i$ ,  $b_i$  и  $a_D$ ,  $b_D$  не могат да приемат едновременно стойност 0.

Съгласно този критерий всички състояния между линията за справедливо споделяне на ресурсите и правата, преминаваща през началото на координатната система и точката  $x^H$ , са с по-висока степен на справедливост от  $x^H$ . В сила е същото правило и за огледалната точка на  $x^H$ . За да се удовлетвори критерия за справедливо разпределение, е необходимо състоянието на системата да е затворено между двете правите свързващи точките  $x^H$  и  $x^{H'}$ .



Фигура. 4 Векторно представяне на изпълнението на критерия за справедливо разпределение на ресурсите

Наличието на изискване за отсъствие на информация, различна от тази на обратната връзка  $y(t)$  относно системното състояние, допълнително ограничава набора от възможните схеми за линейно управление. Това изискване налага следното условие:

$$\begin{aligned} y(t) = 0 &\rightarrow x_i(t+1) > x_i(t) \quad \forall i \\ y(t) = 1 &\rightarrow x_i(t+1) < x_i(t) \quad \forall i \end{aligned} \quad (11)$$

Представено чрез коефициентите  $a$ ,  $b$  може да се запише по следния начин:

$$\begin{aligned} a_i + (b_i - 1)x_i(t) &> 0 \quad \forall x_i(t) \geq 0 \\ a_D + (b_D - 1)x_i(t) &< 0 \quad \forall x_i(t) \geq 0 \end{aligned}$$

Тези два математически израза ограничават условията (10):

$$\begin{aligned} a_i &> 0 & b_i &\geq 1 \\ a_D &= 0 & 0 \leq b_D < 1 \end{aligned} \quad (12)$$

Условието в (10) могат да гарантират ефективна работа на системата чрез изпълнение на изискванията за разпределеност. Това не се изпълнява единствено в случая, когато всички потребители реагират на системата едновременно. С цел избягване на такава ситуация се прави допускането за следните ограничения:

$$a_i + (b_i - 1)X_{\max} > 0$$

$$\begin{aligned} N_{\max} a_D + (b_D - 1)X_{\min} < 0 \\ N_{\min} < X < X_{\max}, \end{aligned} \quad (13)$$

където  $N_{\max}$  е горната граница на броя на потребителите, ползващи общите ресурси. Когато условието (13) е удовлетворено, не може да се изпълни равенството  $\sum x_i(t+1) = \sum x_i(t)$ .

Допуска се, че условия (13) са изпълнени за случая, когато  $y(t)=0$ :

$$\begin{aligned} na_i + b_i x_i(t) < x_i(t) \quad \forall i \\ na_i + (b_i - 1) \sum x_i(t) < 0 \end{aligned} \quad (14)$$

Възможно е стойностите на коефициентите  $a_i$ ,  $b_i$  и параметъра  $x_i$  да бъдат положителни и неравенството (14) да е изпълнено, само когато  $b_i - 1 < 0$ . Неравенството (14) може да се изрази и чрез заместване на  $\sum x_i$  с  $X_{\max}$ :

$$nx_i + (b_i - 1)X_{\max} < 0 \quad (15)$$

(15) нарушава (13), което противоречи на направеното допускане.

Приема се, че допускането е за  $y(t)=1$ . Тогава:

$$\begin{aligned} na_D + b_D x_i(t) < x_i(t) \quad \forall i \\ na_D + (b_D - 1) \sum x_i(t) < 0 \end{aligned} \quad (16)$$

Ако  $b_D < 1$ , лявата страна на неравенство (16) приема отрицателни стойности. Действието, което се предприема от системата е в посока от началото на координатната система към позиция  $X_{\min}$ . При тези обстоятелства неравенство (16) добива вида:

$$N_{\max} a_D + (b_D - 1)X_{\min} > 0 \quad (17)$$

Това нарушава (13) и противоречи на направеното допускане.

Въз основа на този анализ се предлагат нови условия за линейно управление, които са по-малко ограничителни от изразените в (12) и по-ограничителни от изразените в (10):

$$\begin{aligned} a_i > 0 \quad b_i \geq 1 - \frac{a_i}{X_{\max}} \\ 0 \leq a_D < (1 - b_D) \frac{X_{\min}}{X_{\max}} \quad 0 \leq b_D < 1 \end{aligned} \quad (18)$$

Изразът (9) е монотонно увеличаваща функция на  $c=a/b$ . Следователно малки стойности за  $a$  и големи стойности за  $b$ , могат да доведат до по-бързо привеждане на системата в зоната на справедливо споделяне на ресурсите. За изпълнение на това изискване е необходимо  $a_D=0$ . Следователно за справедливо разпределение на ресурсите системата остава неизменна за всяка една стъпка на намаляване. В този случай, параметърът  $b_D$  не оказва влияние. При увеличаване интензивността на

потребителите на системата, малки стойности за  $b_i$ , водят до по-бързо достигане състоянието на справедливо разпределение на мрежовите ресурси.

### ЗАКЛЮЧЕНИЕ

Схемите за контрол на претоварването обезпечават работата на комуникационните мрежи в зоната на малки закъснения и висока производителност. Това се постига посредством наблюдение и изпращане на информация до потребителите за текущото ниво на натоварване.

В статията са формулирани набор от условия, които трябва да бъдат задоволени от всяка схема за увеличаване/намаляване на използваните мрежови ресурси от страна на потребителите, с цел гарантиране на справедливото им разпределение. За да се обезпечи ефективното използване на системата при така дефинираните стойности на ключовите параметри в алгоритмите за управление на претоварването е необходимо процеса на намаляване на използваните ресурси да има мултипликативен, а процеса на увеличаване адитивен характер.

*Публикуваните резултати са получени при работата по договор № ВУ-ТН-105/2005 г. на МОН на тема „Планиране на мултимедийни телекомуникационни мрежи с управление на трафика и качеството на обслужване“*

### ЛИТЕРАТУРА

- [1] Allman M., Eddy W., Ostermann S., Estimating Loss Rates with TCP. ACM SIGMETRICS Performance Evaluation Review, vol.31, no.3, pp. 12 – 24, 2003.
- [2] Allman M., Paxson V., Stevens W, TCP congestion control. IETF RFC 2581, 1999.
- [3] Floyd, S., Fall K., Promoting the use of end-to-end congestion control in the Internet, IEEE/ACM Transactions on Networking, vol.7, no.5, pp. 458 – 472, 1999.
- [4] Floyd S., Jacobson V., Random Early Detection gateways for congestion avoidance, IEEE/ACM Transactions on Networking, vol.1, no.4, pp. 397 – 413, 1993.
- [5] Jacobson V., Congestion avoidance and control. ACM SIGCOMM Computer Communication Review, vol.25, no.1, 1995, pp. 157 – 187
- [6] Misra A., Baras J., Ott T., Generalized TCP congestion avoidance and its effect on bandwidthsharing and variability, Proceedings of the IEEE Global Telecommunications Conference GLOBECOM'00, pp. 329 – 337, 2000.
- [7] Vicisano L., Rizzo L., Crowcroft J., TCP-like congestion control for layered multicast data transfer, Proceedings of the IEEE Conference Computer and Communications Societies INFOCOMM'98, pp. 996 – 1003, 1998.

### За контакти:

ас. инж. Георги Валентинов Христов, катедра “Комуникационна техника и технологии”, Русенски университет “Ангел Кънчев”, тел.: 082-888 663, e-mail: [ghristov@mbox.contact.bg](mailto:ghristov@mbox.contact.bg)

гл. ас. д-р инж. Теодор Божидаров Илиев, катедра “Комуникационна техника и технологии”, Русенски университет “Ангел Кънчев”, тел.: 082-888 663, e-mail: [tiliev@ecs.ru.acad.bg](mailto:tiliev@ecs.ru.acad.bg)

**Докладът е рецензиран.**