UDC 621.39

Algorithm for Computing Steady-State Probabilities of the Queuing System with Hysteretic Congestion Control and Working Vacations

P. O. Abaev

Telecommunication Systems Department
Peoples' Friendship University of Russia
Miklukho-Maklaya str., 6, Moscow, Russia, 117198

This paper is devoted to investigation of a SIP congestion behavior under overload. The hysteretic overload control mechanism is considered to avoid SIP server collapse and also server working vacations was taken into account. In the paper a queuing model is introduced. The algorithm for computing steady-state probabilities is derived.

Key words and phrases: SIP server, hysteretic overload control, working vacation.

1. Introduction

With growth of deployment of IP based services and increasing use of SIP signaling in NGN, necessitates providing mechanisms handling extreme traffic surges. Overload occurs when the incoming request rate to a SIP server is beyond its processing capacity. If a SIP server becomes overloaded transaction delay increases. SIP requests are retransmitted when adequate responses are not received in a predetermined interval, in order to keep high reliable transmissions of SIP messages over UDP. The retransmission further increases load and it can causes performance degradation of a SIP server, that is, a significant reduce the server throughput, which causes the majority of calls to fail leading to a congestion collapse. At present, the SIP protocol provides a basic limited mechanism for overload control through its 503 (Service Unavailable) response code, which stops the current session request. Unfortunately, this mechanism is not effective and has numerous problems in actual deployment [1].

Various overload control mechanisms based on the basic one, which has different rejection policies and metrics to predict overload condition are introduced in papers [2–5]. For example, in the papers [2,3] queue length based algorithm using two thresholds to detect server states was proposed. For the first time this mechanism was described in [6]. It provides congestion control at the application layer.

The paper deals with a queuing model of the SIP server with working vacations and hysteretic overload control mechanism. Major performance measures of the system can be expressed in terms of the steady-state probabilities. To calculate the probabilities the recursive algorithm is derived.

2. SIP Server Queuing Model

Let us consider a single-server queuing system, depicted in Fig. 1, with working vacations and hysteretic congestion control and denote it as $M|M_2|1|0 < L < R < \infty|WV$ according to the modified Kendall classification. A Poisson customer flow arrives at the system. Customers are queued and take service in accordance with congestion control algorithm. The server operates in two modes: normal and congestion. To detect an overload we introduce two thresholds. When the queue becomes full the system recognizes detecting a congestion and new arrived customers are discarded. So, when the queue length becomes to be less than L, the system recognizes that the congestion is removed and starts putting new customers into the queue. The server takes a

Received 5th May, 2011.
This work was supported by the Russian Foundation for Basic Research, project №10-07-00487-a.

working vacation at the times when the system is empty. During the vacation period new arrived customers are stored in the buffer. The server takes another new vacation if only there is no any new customer in the queue.

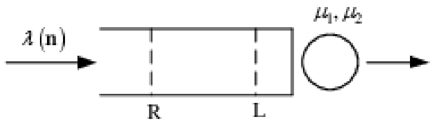

Figure 1. **Queuing Model**

We denote $n_1(t) \in \{1, 2\}$ as the server state in the instant t, where state "1" means the server is busy serving a customer and state "2" the server takes a working vacation. We need also to specify whether control is on or not. Thus we let $n_2(t) \in \{0, 1\}$ equals 0 or 1 to indicate the control is off or on respectively. The occupancy of the queue is denoted by $n_3(t) = \overline{0, R}$. Therefore the Markov process $\mathbf{N}(t) = (n_1, n_2, n_3)$ describes completely the system over the state space

$$\mathcal{N} = \mathcal{N}_0 \cup \mathcal{N}_1, \quad \mathcal{N}_0 = \{\mathbf{n}: \; n_1 = 1, 2; \; n_2 = 0; \; 0 \leqslant n_3 \leqslant R - 1\},$$
$$\mathcal{N}_1 = \{\mathbf{n}: \; (n_1 = 1, 2; \; n_2 = 0; \; 0 \leqslant n_3 \leqslant R - 1) \vee (n_2 = 1; \; n_3 = R)\}.$$

The dependence of the intensity of customer arrivals on the system states is specified by the following relation

$$\lambda(\mathbf{n}) = \lambda \cdot u\left((1 - n_2) \cdot (R - n_3)\right), \quad \mathbf{n} \in \mathcal{N}, \tag{1}$$

where $u(\cdot)$ is the Heaviside function. The qualitative interpretation of the dependence is presented in Fig. 2.

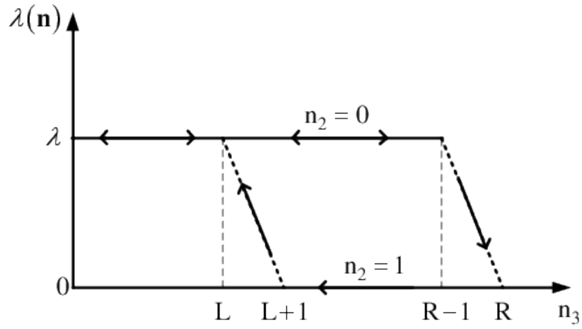

Figure 2. **Hysteretic Congestion Control Mechanism**

We assume the customer service time and the vacation durations to be exponentially distributed with parameters μ_1 and μ_2 respectively. We consider that stationary probabilities $p_{n_1 n_2 n_3} = \lim\limits_{t \to \infty} P\{\mathbf{N}(t) = \mathbf{n}\}$, $\mathbf{n} \in \mathcal{N}$ exist, and satisfy the system of equilibrium equations

$$\lambda p_{200} = \mu_1 p_{100}, \tag{2}$$

$$(\lambda + \mu_2) p_{20i} = \lambda p_{20i-1}, \quad i = \overline{1, R-1}, \tag{3}$$

$$\mu_2 p_{21R} = \lambda p_{20R-1}, \tag{4}$$

$$(\lambda + \mu_1)\, p_{10i} = \lambda u\,(i)\, p_{10i-1} + \mu_1 p_{10i+1} + \mu_2 p_{20i+1},\ i = \overline{0, L-1},\ i = \overline{L+1, R-2}, \tag{5}$$

$$(\lambda + \mu_1)\, p_{10L} = \lambda p_{10L-1} + \mu_1 p_{10L+1} + \mu_2 p_{20L+1} + \mu_1 p_{11L+1}, \tag{6}$$

$$(\lambda + \mu_1)\, p_{10R-1} = \lambda p_{10R-2}, \tag{7}$$

$$\mu_1 p_{11R} = \lambda p_{10R-1}, \tag{8}$$

$$p_{11i} = p_{11i+1}, \quad i = \overline{L+1, R-2}, \tag{9}$$

$$\mu_1 p_{11R-1} = \mu_1 p_{11R} + \mu_2 p_{21R}. \tag{10}$$

In the next section we derive a recursive algorithm for efficient calculation of the probabilities $p_{n_1} n_2 n_3$.

3. Algorithm for Computation of Steady-State Probabilities

In the real situation the number of states is so large that problem of performance evaluation becomes intractable. Therefore we propose a method for calculation of the state probabilities. We express $p_{n_1 n_2 n_3}$ in terms of p_{200}, i.e., $p_{n_1 n_2 n_3} = x_{n_1 n_2 n_3} p_{200}$. The coefficients $x_{n_1 n_2 n_3}$ fulfill conditions formulated in the lemma bellow. In order to find probability p_{200}, we make use of the normalization condition $\sum\limits_{\mathbf{n} \in \mathcal{N}} p_{n_1 n_2 n_3} = 1$.

Lemma 1. *The coefficients $x_{n_1 n_2 n_3}$ obey the following relations*

$$x_{200} = 1, \quad x_{20i} = \left(\frac{\lambda}{\lambda + \mu_2}\right)^i, \quad i = \overline{1, R-1}, \quad x_{21R} = \frac{\lambda}{\mu_2} \cdot \left(\frac{\lambda}{\lambda + \mu_2}\right)^{R-1}, \tag{11}$$

$$\begin{gathered} x_{100} = \frac{\lambda}{\mu_1}, \\ x_{10i+1} = \mu_1^{-1} \left[(\lambda + \mu_1)\, x_{10i} - \lambda u\,(i)\, x_{10i-1} - \mu_2 x_{20i+1} \right], \quad i = \overline{0, L-1}, \end{gathered} \tag{12}$$

$$x_{11L+i} = \frac{\lambda A_{R-L-2} + \mu_2 x_{21R}}{\mu_1 + \lambda B_{R-L-2}}, \quad i = \overline{1, R-L-1}, \tag{13}$$

$$x_{10L+i} = A_{i-1} - B_{i-1} x_{11L+1}, \quad i = \overline{1,\ R-L-1}, \tag{14}$$

$$x_{11R} = \frac{\lambda}{\mu_1} x_{10R-1}, \tag{15}$$

where A_i and B_i are given by

$$\begin{gathered} A_0 = \mu_1^{-1} \left(\lambda x_{20L} + \lambda x_{10L} - \mu_2 x_{20L+1} \right), \\ A_i = \mu_1^{-1} \left(\lambda x_{20L+i} + \lambda A_{i-1} - \mu_2 x_{20L+i+1} \right), \quad i = \overline{1, R-L-2}, \end{gathered} \tag{16}$$

$$B_0 = 1, \quad B_i = 1 + \frac{\lambda}{\mu_1} B_{i-1}, \quad i = \overline{1, R-L-2}. \tag{17}$$

Proof. We can clearly see that expressions (11) and (12) are derived from equations (2)–(5). Let us consider the way of formulas derivation for the remaining coefficients x_{11i}, $i = \overline{L+1, R}$ and x_{10i}, $i = \overline{L+1, R-1}$. At first we obtain the auxiliary equation by summing up the first L equations from (5)

$$\mu_1 p_{10L} - \lambda p_{10L-1} = \mu_1 p_{100} - \mu_2 \sum_{i=0}^{L-1} p_{20i+1}. \tag{18}$$

Dividing both sides of (18) by μ_1 and substituting (11) we find

$$p_{10L} - \frac{\lambda}{\mu_1} p_{10L-1} = \frac{\lambda}{\mu_1} \left(\frac{\lambda}{\lambda + \mu_2} \right)^L \cdot p_{200} = \frac{\lambda}{\mu_1} p_{20L}. \quad (19)$$

We rewrite equation (6) as

$$p_{10L+1} = \frac{\lambda}{\mu_1} p_{10L} + p_{10L} - \frac{\lambda}{\mu_1} p_{10L-1} - \frac{\mu_2}{\mu_1} p_{20L+1} - p_{11L+1}, \quad (20)$$

and in view of (19) equation (20) takes the form

$$p_{10L+1} = \mu_1^{-1} \left(\lambda x_{10L} + \lambda x_{20L} - \mu_2 x_{20L+1} \right) p_{200} - p_{11L+1} \quad (21)$$

Then we rewrite expression (21) using formulas (16) and (17)

$$p_{10L+1} = A_0 p_{200} - B_0 p_{11L+1}. \quad (22)$$

Applying (22) into equation (5) when $i = L + 1$ yields

$$p_{10L+2} = \mu_1^{-1} \left(\lambda A_0 p_{200} + \mu_1 A_0 p_{200} - \lambda p_{10L} - \mu_2 p_{20L+2} \right) - p_{11L+1} \left(B_0 + \frac{\lambda}{\mu_1} \right). \quad (23)$$

Then substituting the expression (16) for A_0 in place of the coefficient of μ_1 in (23) we find

$$p_{10L+2} = \mu_1^{-1} \left(\lambda A_0 p_{200} + \lambda p_{20L} - \mu_2 p_{20L+1} - \mu_2 p_{20L+2} \right) - p_{11L+1} \left(1 + \frac{\lambda}{\mu_1} \right). \quad (24)$$

In view of the expression for p_{20Q} and p_{20Q+1} given by (11) equation (24) takes form

$$p_{10L+2} = \mu_1^{-1} \left(\lambda A_0 x_{200} + \lambda x_{20L+1} - \mu_2 x_{20L+2} \right) p_{200} - p_{11L+1} \left(1 + \frac{\lambda}{\mu_1} \right). \quad (25)$$

Continuing this procedure, we finally arrive at recursive expressions

$$p_{10L+i} = A_{i-1} p_{200} - B_{i-1} p_{11L+1}, \quad i = \overline{1, R - L - 1}. \quad (26)$$

Thereby, the probability p_{10L+i} depends on p_{200} and p_{11L+1} only, so the aim is to express p_{11L+1} in terms of p_{200}. Combining the equations (8)–(10) yields the following formula

$$p_{11L+1} = \ldots = p_{11R-1} = \frac{\lambda}{\mu_1} p_{10R-1} + \frac{\mu_2}{\mu_1} p_{21R}. \quad (27)$$

Substituting the expression for p_{10R-1} from (26) into (27) we obtain

$$p_{11L+1} = \ldots = p_{11R-1} = \frac{\lambda}{\mu_1} \left(A_{R-L-2} p_{200} - B_{R-L-2} p_{11L+1} \right) + \frac{\mu_2}{\mu_1} p_{21R}. \quad (28)$$

From equation (28) the probability p_{11L+1} obeys the formula

$$p_{11L+1} = \frac{\lambda A_{R-L-2} + \mu_2 x_{21R}}{\mu_1 + \lambda B_{R-L-2}} \cdot p_{200}.$$

Consequently, we have derived formulas (13)–(15), and the lemma is proved.

4. Conclusion

In this paper the queuing model with the hysteretic overload control mechanism is considered. The model also takes into account server working vacations. To make the computation of the probability more simple the recursive algorithm was proposed. The subject of future study is to specify SIP server model taking into account the peculiarity of the arrival streams of SIP messages, service rates depending of the message types and the hysteretic overload control.

References

1. *Rosenberg J.* Requirements for Management of Overload in the Session Initiation Protocol // RFC5390. — 2008.
2. *Ohta M.* Overload Protection in a SIP Signaling Network // International Conference on Internet Surveillance and Protection. — 2006. — Pp. 205–210.
3. Queuing Strategies for Local Overload Control in SIP Server / R. Garroppo, S. Giordano, S. Spagna, S. Niccolini / IEEE Global Telecommunications Conference. — 2009.
4. *Abaev P.* On SIP Session Setup Delay Modeling in Next Generation Networks / IEEE International Congress on Ultra Modern Telecommunications and Control System (ICUMT). — 2010. — Pp. 1125–1131.
5. *Shen C., Schulzrinne H., Nahum E.* Session Initiation Protocol (SIP) Server Overload Control: Design and Evaluation // Lecture Notes in Computer Science. — Springer, 2008. — Vol. 5310. — Pp. 149–173.
6. ITU-T Recommendation Q.704: Signalling System No.7 – Message Transfer Part, Signalling Network Functions and Messages. — 1996.

УДК 621.39

Алгоритм расчета стационарных вероятностей СМО с гистерезисным управлением и прогулками прибора

П. О. Абаев

Кафедра систем телекоммуникаций
Российский университет дружбы народов
ул. Миклухо-Маклая, 6, г. Москва, Россия, 117198

В статье исследуется модель функционирования SIP сервера в условиях перегрузки. Для контроля перегрузки сервера применяется механизм гистерезисного управления нагрузкой. Разработана модель функционирования SIP сервера в условиях перегрузки с прогулками прибора и гистерезисным управлением нагрузкой. Предложен алгоритм расчета стационарных вероятностей функционирования системы.

Ключевые слова: SIP сервер, гистерзисное управление, прогулка прибора.