
Математическая теория телетрафика и сети телекоммуникаций

УДК 621.39

Анализ модели $M|G|1|r$ с групповым поступлением и гистерезисным управлением нагрузкой

Э. С. Сопин

*Кафедра систем телекоммуникаций
Российский университет дружбы народов
ул. Миклуто-Маклая, д. 6, Москва, Россия, 117198*

В статье представлены результаты исследования системы массового обслуживания типа $M|G|1|r$ с групповым поступлением заявок и гистерезисным управлением нагрузкой. Получены соотношения для расчета стационарного распределения вероятностей системы, а также формулы для основных вероятностно-временных характеристик, представляющих интерес при анализе гистерезисного управления нагрузкой, применяющегося для контроля перегрузок SIP-серверов в сетях сигнализации.

Ключевые слова: система массового обслуживания, групповое поступление, гистерезисное управление, SIP-сервер.

1. Введение

В современных телекоммуникационных сетях для предотвращения перегрузок крайне необходимым является использование инструментов управления нагрузкой. Одним из самых простых и интуитивно понятных инструментов является механизм порогового управления длиной очереди, который был предложен рабочими группами IETF (Internet Engineering Task Force) для предотвращения перегрузок в сетях сигнализации на базе протокола SIP (Session Initiation Protocol) [1–4].

В статье рассматривается один из вариантов гистерезисного управления нагрузкой с тремя порогами: порог обнаружения перегрузки H , порог снижения перегрузки L и порог сброса нагрузки R (рис. 1).

Механизм работает следующим образом: при достижении длины очереди порога H система переходит в режим перегрузки, при этом интенсивность поступающего потока сообщений снижается. Чтобы режим функционирования не менялся слишком часто, интенсивность поступающего потока не восстанавливается до прежних значений сразу же, как только длина очереди снижается ниже порога H , а только тогда, когда уровень занятости буферной памяти снизится до порога снижения перегрузки L . Аналогично, при достижении длины очереди порога R поступающая нагрузка полностью сбрасывается и восстанавливается только тогда, когда уровень занятости буфера снижается до порога H .

Для построения модели функционирования SIP-сервера необходимо также учитывать групповой характер поступления SIP-сообщений, таких как NOTIFY, применяющихся при предоставлении услуг присутствия [5]. В связи с этим, а также с целью получения более общих результатов, опишем функционирование SIP-сервера в терминах системы массового обслуживания типа $M^{[X]}|G|1|(L, H)|(H, R)$. Анализ подобных моделей без учета группового поступления заявок проведен, например в работах [6, 7]. В данной статье получена система уравнений для расчета стационарного распределения построенной модели при помощи аппарата марковских процессов восстановления [8]. Кроме того, получены формулы для некоторых характеристик, представляющих интерес для анализа показателей качества функционирования SIP-сервера, таких как вероятность нахождения системы в состояниях перегрузки, среднее время пребывания в состояниях перегрузки и сброса нагрузки и средняя длительность цикла управления.

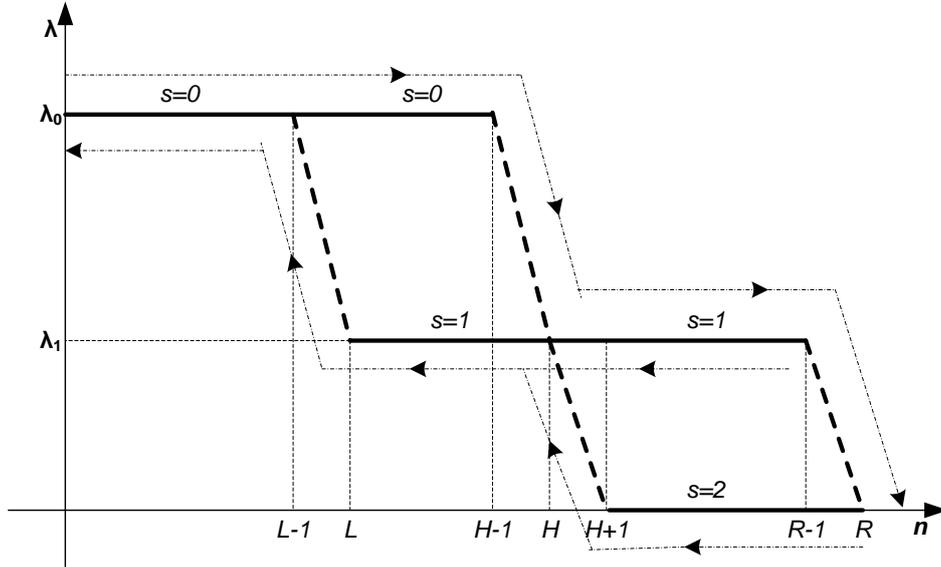


Рис. 1. Гистерезисное управление нагрузкой

2. Построение модели

Обозначим $X(t)$ — двумерный случайный процесс с множеством состояний S

$$S = \left\{ (j, s) \left| \begin{array}{ll} j = \overline{0, R}, & s = 0 \\ j = \overline{L, R}, & s = 1 \\ j = \overline{H+1, R}, & s = 2 \end{array} \right. \right\}$$

и его подмножествами $S_i = \{(j, s) \in S | s = i\}, i = \{0, 1, 2\}$, где j — число заявок в системе, а s — режим ее функционирования в момент времени $t \geq 0$. Заявки поступают на прибор группами, поток групп заявок является пуассоновским с интенсивностью $\lambda_s, s = \{0, 1, 2\}$, причем $\lambda_0 > \lambda_1, \lambda_2 = 0$. В каждой группе поступает случайное число заявок с вероятностью l_i того, что поступит ровно $i \geq 0$ заявок. Длительность обслуживания является случайной величиной с функцией распределения $B(x)$ и средним $b^{(1)} < \infty$.

Пусть $t_1 < t_2 < \dots$, где t_n — момент окончания обслуживания n -ой заявки. Для упрощения анализа будем считать, что режим функционирования s может меняться только в моменты $t_n, n > 0$. Тогда состояния дискретного случайного процесса $X(t_n + 0)$ образуют вложенную цепь Маркова. Множество состояний $X(t_n + 0)$ и его подмножества описываются следующим образом:

$$\tilde{S} = \left\{ (j, s) \left| \begin{array}{ll} j = \overline{0, H-2}, & s = 0 \\ j = \overline{L, R-2}, & s = 1 \\ j = \overline{H+1, R-1}, & s = 2 \end{array} \right. \right\};$$

$$\tilde{S}_i = \{(j, s) \in \tilde{S} | s = i\}, i = \{0, 1, 2\}.$$

Обозначим стационарные распределения процессов $X(t)$ and $X(t_n + 0)$ соответственно $\{p_{j,s}\}$ и $\{q_{j,s}\}$, где

$$p_{j,s} = \lim_{t \rightarrow \infty} P\{X(t) = (j, s)\};$$

$$q_{j,s} = \lim_{n \rightarrow \infty} P\{X(t_n + 0) = (j, s)\}.$$

Для нахождения переходных вероятностей цепи Маркова обозначим β_k^s вероятность того, что за время обслуживания одной заявки в режиме s в систему поступит k заявок, т.е.

$$\beta_k^s = \int_0^{\infty} e^{-\lambda_s x} \frac{(\lambda_s x)^k}{k!} d(B(x)).$$

Кроме того, введем l_i^k — вероятность того, что в k группах поступит ровно i заявок. Нетрудно убедиться, что распределение $\{l_i^k\}$ является k -кратной сверткой распределения $\{l_i\}$:

$$l_i^k = \sum_{n=0}^i l_n l_{i-n}^{k-1}.$$

Тогда система уравнений равновесия для распределения $\{q_{j,s}\}$ принимает вид

$$q_{j,0} = q_{0,0} \sum_{i=1}^{j+1} l_i \sum_{k=0}^{j-i+1} l_{j-i+1}^k \beta_k^0 + \sum_{i=1}^{\min(j+1, H-2)} q_{i,0} \sum_{k=0}^{j-i+1} l_{j-i+1}^k \beta_k^0 +$$

$$+ \delta_{j, L-1} \beta_0^0 q_{L,1}, \quad j = \overline{0, H-2},$$

$$q_{j,1} = q_{0,0} \sum_{i=1}^{j+1} l_i \sum_{k=0}^{j-i+1} l_{j-i+1}^k \beta_k^0 + \sum_{i=1}^{H-2} q_{i,0} \sum_{k=0}^{j-i+1} l_{j-i+1}^k \beta_k^0 +$$

$$+ \sum_{i=L}^{\min(j+1, R-2)} q_{i,1} \beta_{j-i+1}^1 + \delta_{j, H} q_{H+1,2}, \quad j = \overline{H-1, R-2},$$

$$q_{j,1} = \sum_{i=L}^{j+1} q_{i,1} \sum_{k=0}^{j-i+1} l_{j-i+1}^k \beta_k^1, \quad j = \overline{L, H-2},$$

$$q_{R-1,2} = q_{0,0} \sum_{i=1}^{\infty} l_i \sum_{j=\max(0, R-i)}^{\infty} \sum_{k=0}^j l_j^k \beta_k^0 + \sum_{i=1}^{H-2} q_{i,0} \sum_{j=R-i}^{\infty} \sum_{k=0}^j l_j^k \beta_k^0 +$$

$$+ \sum_{i=L}^{R-2} q_{i,1} \sum_{j=R-i}^{\infty} \sum_{k=0}^j l_j^k \beta_k^1,$$

$$q_{j,2} = q_{R-1,2}, \quad j = \overline{H+1, R-2},$$

где $\delta_{i,j}$ — символ Кронеккера.

Теперь, используя распределение $\{q_{j,s}\}$ по вложенной цепи Маркова, можно выписать уравнения для вероятностного распределения $\{p_{j,s}\}$ в произвольные моменты времени процесса $X(t)$.

$$\begin{aligned}
p_{0,0} &= C^{-1} \frac{1}{\lambda_0} q_{0,0}, \\
p_{j,0} &= \frac{C^{-1}}{\lambda_0} \left\{ q_{0,0} \sum_{i=0}^j l_i \sum_{k=0}^{j-i} l_{j-i}^k \left(1 - \sum_{n=0}^k \beta_n^0 \right) + \right. \\
&\quad \left. + \sum_{i=1}^{\min(j, H-2)} q_{i,0} \sum_{k=0}^{j-i} l_{j-i}^k \left(1 - \sum_{n=0}^k \beta_n^0 \right) \right\}, \quad j = \overline{1, R-1}, \\
p_{R,0} &= \frac{C^{-1}}{\lambda_0} \left\{ q_{0,0} \sum_{i=0}^{\infty} l_i \sum_{j=\max(0, R-i)}^{\infty} \sum_{k=0}^j l_j^k \left(1 - \sum_{n=0}^k \beta_n^0 \right) + \right. \\
&\quad \left. + \sum_{i=1}^{H-2} q_{i,0} \sum_{j=R-i}^{\infty} \sum_{k=0}^j l_j^k \left(1 - \sum_{n=0}^k \beta_n^0 \right) \right\}, \\
p_{j,1} &= \frac{C^{-1}}{\lambda_1} \sum_{i=L}^{\min(j, R-2)} q_{i,1} \sum_{k=0}^{j-i} l_{j-i}^k \left(1 - \sum_{n=0}^k \beta_n^1 \right), \quad j = \overline{L, R-1}, \\
p_{R,1} &= \frac{C^{-1}}{\lambda_1} \sum_{i=L}^{R-2} q_{i,1} \sum_{j=R-i}^{\infty} \sum_{k=0}^j l_j^k \left(1 - \sum_{n=0}^k \beta_n^1 \right), \\
p_{j,2} &= C^{-1} b^{(1)} q_{R-1,2}, \quad j = \overline{H+1, R},
\end{aligned} \tag{2}$$

где $C = \left(\frac{1}{\lambda_0} + b^{(1)} \right) q_{0,0} + b^{(1)} (1 - q_{0,0}) = b^{(1)} + \frac{1}{\lambda_0} q_{0,0}$.

3. Анализ вероятностно-временных характеристик

Теперь мы можем найти вероятность нахождения системы в режиме перегрузки суммированием соответствующих вероятностей распределения $\{p_{j,s}\}$.

$$P(S_1) = \sum_{(j,s) \in S_1} p_{j,s} = \sum_{j=L}^R p_{j,1} = C^{-1} b^{(1)} \sum_{j=L}^{R-2} q_{j,1}. \tag{3}$$

Аналогично вычисляется вероятность нахождения системы в режиме сброса нагрузки.

$$P(S_2) = \sum_{(j,s) \in S_2} p_{j,s} = \sum_{j=H+1}^R p_{j,2} = C^{-1} b^{(1)} \sum_{j=H+1}^{R-1} q_{j,2}. \tag{4}$$

Найдем теперь среднее время τ_{12} пребывания системы в множестве состояний перегрузки и сброса нагрузки. Пусть \mathbf{P}_0 — матрица переходных вероятностей цепи Маркова $X(t_n + 0)$ на подмножестве \tilde{S}_0 размерности $(H-1) \times (H-1)$.

$$(\mathbf{P}_0)_{ij} = \begin{cases} \sum_{i=1}^{j+1} l_i \sum_{k=0}^{j-i+1} l_{j-i+1}^k \beta_k^0, & i = 0, \quad j = \overline{0, H-2}; \\ \sum_{k=0}^{j-i+1} l_{j-i+1}^k \beta_k^0, & i = \overline{1, H-2}, \quad j = \overline{i-1, H-2}; \\ 0, & j < i-1. \end{cases}$$

Рассмотрим поведение процесса $X(t_n + 0)$ в подмножестве состояний \tilde{S}_0 . Выходя из подпространства \tilde{S}_1 в \tilde{S}_0 , система всегда попадает в состояние $(L - 1, 0)$, поэтому начальное распределение в подпространстве \tilde{S}_0 можно выразить вектором $\mathbf{e}_L^T = \left(\underbrace{0, \dots, 0}_{L-1}, 1, \underbrace{0, \dots, 0}_{H-L-1} \right)$. Пусть также $\mathbf{a}^T = (a_0, \dots, a_{H-2})$, где $a_i = 1 - \sum_{j=0}^{H-2} (\mathbf{P}_0)_{ij}$ — вероятность того, что система, находясь в состоянии $(i, 0)$, выйдет из подпространства \tilde{S}_0 на следующем шаге цепи Маркова. Тогда $\mathbf{e}_L^T \mathbf{P}_0^n \mathbf{a}$ — вероятность того, что система выйдет из подмножества \tilde{S}_0 ровно через $n + 1$ шагов процесса $X(t_n + 0)$, а среднее число шагов $\tilde{\tau}_0$ имеет вид

$$\tilde{\tau}_0 = \sum_{n=0}^{\infty} (n + 1) \mathbf{e}_L^T \mathbf{P}_0^n \mathbf{a} = \mathbf{e}_L^T (\mathbf{I} - P_0)^{-1} \mathbf{1}. \quad (5)$$

Чтобы получить среднее время τ_0 нахождения процесса $X(t)$ в подмножестве S_0 , необходимо домножить выражение (5) на среднюю длительность одного шага цепи Маркова $X(t_n + 0)$ в подмножестве состояний \tilde{S}_0 , в результате

$$\tau_0 = \left(b^{(1)} + \frac{1}{\lambda_0} \frac{q_{0,0}}{P(\tilde{S}_0)} \right) \mathbf{e}_L^T (\mathbf{I} - P_0)^{-1} \mathbf{1}. \quad (6)$$

Поскольку в стационарном режиме система входит и выходит из подмножества состояний S_0 одинаковое число раз, тогда τ_{12} — среднее время пребывания в подмножестве $S_1 \cup S_2$ определяем по формуле

$$\tau_{12} = \tau_0 \frac{P(S_1 \cup S_2)}{P(S_0)}. \quad (7)$$

Теперь из формул (6) и (7) легко найти среднюю длительность τ цикла управления:

$$\tau = \tau_0 + \tau_{12}. \quad (8)$$

4. Численный анализ

На основании полученных в [9] параметров гистерезисного управления нагрузкой был проведен численный анализ характеристик τ и τ_{12} . В нашем примере использовались следующие величины порогов: $L = 73$, $H = 84$, $R = 100$, а время обслуживания заявок детерминировано и равно 5 мс.

На рис. 2 представлен график зависимости среднего времени нахождения в режиме перегрузки τ_{12} и средней длительности цикла управления τ в зависимости от интенсивности нагрузки в случае без группового поступления.

На рис. 3 представлен аналогичный график для случая группового поступления заявок, количество заявок в группе имеет геометрическое распределение с параметром 0,5.

Из графиков видно, что при небольших $\rho < 1,3$ длительность цикла управления в большей степени зависит от времени нахождения системы в режиме нормального функционирования, но с ростом нагрузки доля времени в режиме перегрузки значительно возрастает. Стоит отметить также, что при небольших перегрузках ($\rho < 1,5$) исследуемые характеристики для систем с групповым поступлением заявок или без практически совпадают, однако при увеличении перегрузки разница становится существенной.

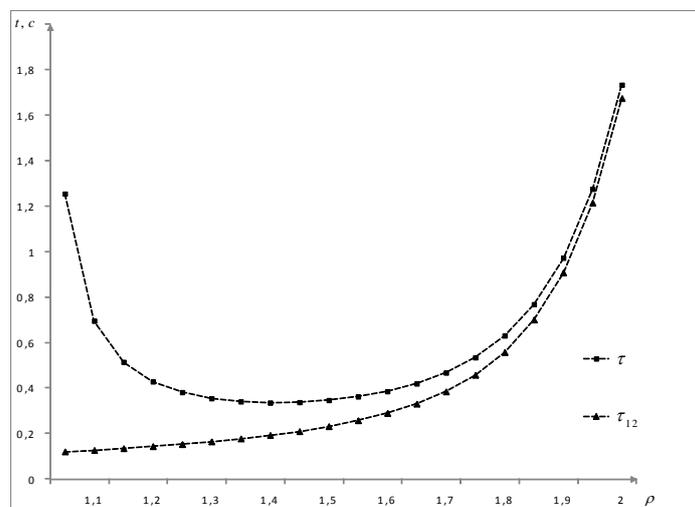


Рис. 2. Характеристики гистерезисного управления для систем с ординарным потоком

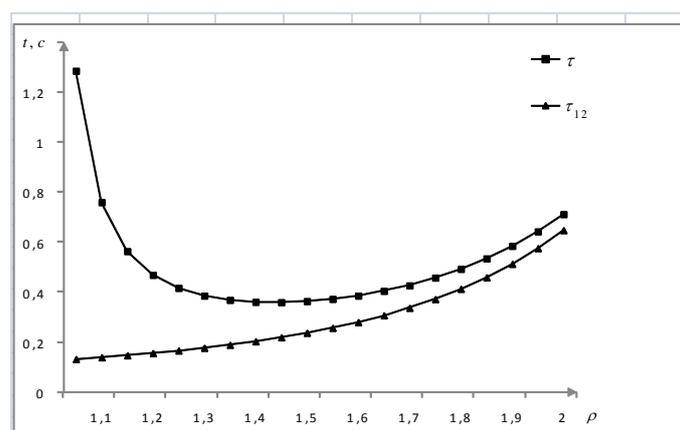


Рис. 3. Характеристики гистерезисного управления для систем с групповым поступлением заявок

5. Заключение

В статье для анализа показателей качества функционирования SIP-сервера построена модель типа $M|G|1$ с гистерезисным управлением и групповым поступлением заявок. На основании модели были выписаны системы уравнений, позволяющие вычислять стационарное распределение вероятностей системы, получены формулы основных вероятностно-временных характеристик: вероятности нахождения в режиме перегрузки и режиме сброса нагрузки, среднее время возврата из режима перегрузки и средняя длительность цикла управления. Проведенный численный анализ позволяет утверждать, что групповое поступление заявок является существенным фактором, влияющим на характеристики гистерезисного управления нагрузкой и, следовательно, должно учитываться при построении более точных моделей.

Литература

1. *Rosenberg J.* RFC 5390. Requirements for Management of Overload in the Session Initiation Protocol. — 2008.
2. *Hilt V., Noel E., Shen C., Abdelal A.* RFC 6357. Design Considerations for Session Initiation Protocol (SIP) Overload Control. — 2011.
3. *Montagna S., Pignolo M.* Load control techniques in SIP signaling networks using multiple thresholds. // 13th International Telecommunications Network Strategy and Planning Symposium NETWORKS. — Budapest, Hungary: 2008. — 1-17 p.
4. *Абаев П. А., Гайдамака Ю. В., Самуёлов К. Е.* Гистерезисное управление нагрузкой в сетях сигнализации. // Вестник РУДН. Серия «Математика. Информатика. Физика». — 2011. — С. 55–73. [Абаев П. А., Gaidamaka Y. V., Hysteretic Overload Control in a SIP Signaling Network // Bulletin of Peoples' Friendship University of Russia. Series «Series Mathematics. Information Sciences. Physics». — 2011. — № 4. — P. 55–73]
5. IMS presence server: traffic analysis & performance modeling / C. Chi, R. Hao, D. Wang, Z. Cao // Proc. of the International Conference on Network Protocols. — Orlando, Florida, USA: 2008. — Pp. 63–72.
6. *Gaidamaka Y., Samouylov K., Sopin E.* Analysis of $M|G|1$ queue with hysteretic load control // XXX International Seminar on Stability Problems for Stochastic Models, the Autumn Session of the VI International Seminar on Applied Problems of Probability Theory and Mathematical Statistics related to Modeling of Information Systems. — Svetlogorsk: 2012. — Pp. 87–89.
7. *Roughan M., Pearce C. E. M.* A martingale analysis of hysteretic overload control // Advances in Performance Analysis. — 2000. — Pp. 1–30.
8. *Самуёлов К., Сопин Э.* К анализу системы $M[X]|G|1|r$ с прогулками прибора // Вестник РУДН. Серия Математика. Информатика. Физика. — 2011. — С. 91–97. [Samouylov K. E., Sopin E. S. On Analysis of $M[X]|G|1|r$ Queuing System // Bulletin of Peoples' Friendship University of Russia. Series «Series Mathematics. Information Sciences. Physics». — 2011. — № 1. — P. 91–97]
9. Simulation of Overload Control in SIP Server Networks / P. O. Abaev, Y. V. Gaidamaka, A. V. Pechinkin et al. // Proc. of the 26th European Conference on Modelling and Simulation ECMS. — Koblenz: 2012. — Pp. 533–539.

UDC 621.39

Analysis of $M|G|1|r$ Queue with Batch Arrival and Hysteretic Overload Control

E. S. Sopin

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

The paper presents $M|G|1|r$ queuing system analysis with batch arrival and hysteretic load control. We provide a system of equations for steady-state probability distribution and derive formulas for system characteristics that are of interest considering hysteretic load control mechanism in SIP-based signalling networks

Key words and phrases: queuing system, batch arrival, hysteretic load control, SIP server.