

СРАВНЕНИЕ ОДНОПороГОВОЙ И ДВУПороГОВОЙ СИСТЕМ ДЛЯ УПРАВЛЕНИЯ ПЕРЕГРУЗКАМИ SIP-СЕРВЕРА

Таланова М.О., Щербакова А.Е.

Российский университет дружбы народов, matalanova@gmail.com, aewerbakova@gmail.com

Построены модели в виде систем с одним и двумя порогами для управления перегрузками. Проведено сравнение некоторых характеристик их функционирования.

Ключевые слова: SIP-сервер, гистерезисное управление нагрузкой, LBOC.

Введение

Управление нагрузкой является основным инструментом предотвращения перегрузок в сетях SIP-серверов (Session Initiation Protocol). Одним из механизмов является гистерезисное управление [1]. Мы исследовали механизм LBOC (Loss-Base Overload Control), при котором осуществляется межузловое управление сигнальной нагрузкой между SIP-серверами.

Системы с гистерезисным управлением перегрузками

Процесс обработки сообщений SIP-сервером можно описать с помощью систем массового обслуживания $M|M|1|L|B$ (рис. 1), и $M_2|M_2|L, H|B$ (рис. 2) с конечной очередью размера B и гистерезисным управлением нагрузкой с порогами L , $1 \leq L < B$, и H , $1 \leq H < B$.



Рис. 1. Система $M|M|1|L|B$ [2]

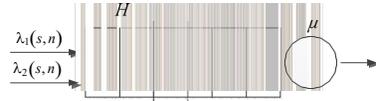


Рис. 2. Система $M_2|M_2|L, H|B$ [3]

На систему $M|M|1|L|B$ поступает пуассоновский поток заявок с интенсивностью λ , а на систему $M_2|M_2|L, H|B$ - два потока с интенсивностями λ_1 и λ_2 , время обслуживания заявки на приборе имеет экспоненциальное распределение с параметром μ . При поступлении на обслуживание заявка сохраняет место в очереди, дисциплина выбора заявок из очереди на обслуживание FCFS. Система с одним порогом может функционировать в одном из двух режимов: режиме нормальной нагрузки ($s=0$) или режиме перегрузки ($s=1$), а система с двумя порогами в одном из трех режимов: режиме нормальной нагрузки ($s=0$), режиме перегрузки ($s=1$) и режиме сброса нагрузки ($s=2$).

Состояние систем имеет вид (s, n) , где статус нагрузки: $s \in \{0, 1\}$ или $s \in \{0, 1, 2\}$

соответственно, а количество заявок в очереди: $n, n = \overline{0, B}$. Функционирование систем описывают марковские процессы $X(t), t > 0$, с пространствами состояний $X = X_0 \cup X_1$ для системы с одним порогом и $X = X_0 \cup X_1 \cup X_2$ для системы с двумя порогами.

Графики интенсивности входящего потока, зависящей от состояния систем, представлены на рис. 3. При нахождении системы с одним порогом во множестве X_0 интенсивность входящего потока заявок равна $\lambda > 0$, во множестве $X_1 - \lambda = 0$. При нахождении системы с двумя порогами во множестве X_0 интенсивность входящего потока заявок равна $\lambda > 0$, в множестве $X_1 -$ уменьшается таким образом, что $\lambda' = \rho\lambda$,

где $p = 1 - q$ – доля заявок, которые принимаются на обслуживание в множестве X_1 .

Если система находится в множестве X_2 , то $\lambda = 0$.

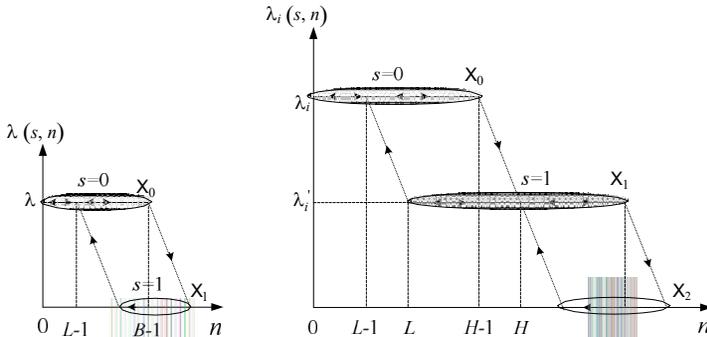


Рис. 3. Функции интенсивности $\lambda(s, n)$ потока заявок для двух систем

Стационарные вероятности $p_{s,n}$ состояний процесса $X(t)$ для двух систем вычисляются через систему уравнений равновесия [3].

Вероятностно-временные характеристики систем

Получены формулы для расчета следующих вероятностно-временных характеристик систем. Для системы с одним порогом вероятность того, что SIP-сервером сообщения на обслуживание не принимаются:

$$B(X_1) = \sum_{n=L}^B p_{1,n}. \quad (1)$$

Для системы с двумя порогами вероятность того, что SIP-сервер перегружен и не принимает неприоритетные сообщения на обработку:

$$B(X_1) = q \cdot \sum_{k=L}^{B-1} p_{1,k}. \quad (2)$$

Вероятность того, что SIP-сервер перегружен, не принимает сообщения на обработку:

$$B(X_2) = \sum_{k=H+1}^B p_{2,k}. \quad (3)$$

Среднее время пребывания системы в множестве состояний перегрузки:

- для системы $M|M|L/B$:

$$M\tau_1 = \mu^{-1}(B - L + 1). \quad (4)$$

- для системы $M_2|M_2|L, H/B$:

$$M\tau = \int_0^t \left(\mu \hat{P} \begin{matrix} (t) \\ (1, H-1)(0, L-1) \end{matrix} - (\lambda' + \mu) \hat{P} \begin{matrix} (t) \\ (1, H)(0, L-1) \end{matrix} + \lambda \hat{P} \begin{matrix} (t) \\ (1, H+1)(0, L-1) \end{matrix} \right) dt, \quad (5)$$

где вероятности переходов находятся с помощью матрицы $\hat{P}(t) = e^{\hat{\Lambda}t}$, $\hat{\Lambda}$ – матрица интенсивностей переходов усеченного процесса с пространством состояний $X' = X_1 \cup X_2 \cup \{(0, L-1)\}$.

Найдено значение 95% квантиль величин $M\tau_1$ и $M\tau_{12}$. Вектор вероятностей усеченного процесса: $\hat{\mathbf{p}}^T(t) = \hat{\mathbf{p}}^T(0)\hat{\mathbf{P}}(t)$. Начальный вектор $\hat{\mathbf{p}}(0)$ для $M \uparrow \downarrow \downarrow$ равен

$$\hat{\mathbf{p}}^{(s,n)}(0) = \begin{cases} 1, & (s,n) = (1,B) \\ 0, & (s,n) \neq (1,B); \end{cases} \text{ для } M_2 \downarrow M_2 \downarrow L, H, B \quad - \quad \hat{\mathbf{p}}^{(s,n)}(0) = \begin{cases} 1, & (s,n) = (1,H) \\ 0, & (s,n) \neq (1,H). \end{cases} \quad [1]$$

Тогда функция распределения $F_\tau(t)$ случайной величины τ для обеих систем равна $F_\tau(t) = \hat{P}_{(0,L-1)}(t)$, $t \geq 0$. Значение 95% квантиль величины τ для обеих систем находится из соотношения $F_\tau(\tau^{0.95}) = 0.95$.

Выводы

Возьмем исходные данные: $B=100$, $\rho=1.2$, $\mu=200 \text{ с}^{-1}$, для системы $M \uparrow \downarrow \downarrow$ | | $L=86$, для системы $M_2 \downarrow M_2 \downarrow L, H, B$ $L, H = \langle 44, 52 \rangle$, $q=0.3$. По формулам (1)-(5) получены результаты:

- для $M \uparrow \downarrow \downarrow$ | | L, B : $B(X_1) = 0.166667$, $M\tau_1 = 0.075 \text{ с}$, $\tau^{0.95} = 0.10944 \text{ с}$.

- для $M_2 \downarrow M_2 \downarrow L, H, B$: $B(X_1 \cup X_2) = 0.16668$, $M\tau_{12} = 0.280547 \text{ с}$, $\tau^{0.95} = 0.9 \text{ с}$.

Для $M_2 \downarrow M_2 \downarrow L, H, B$ время возврата системы в множество состояний нормальной нагрузки больше, чем для $M_2 \downarrow M_2 \downarrow \downarrow \downarrow \downarrow$. Для $M_2 \downarrow M_2 \downarrow L, H, B$ 95% квантиль времени возврата системы в множество нормальной нагрузки меньше, чем для $M_2 \downarrow M_2 \downarrow \downarrow \downarrow \downarrow$. Вероятность пребывания систем в множестве состояний перегрузки почти одинакова.

Литература

1. Абаев П.О., Гайдамака Ю.В., Самуйлов К.Е. Гистерезисное управление нагрузкой в сетях сигнализации // «Вестник РУДН. Серия «Математика. Информатика. Физика»» – М.: Изд-во РУДН. – 2011. – №4. - С. 55-73.
2. Таланова М.О. Анализ простейшей модели порогового управления перегрузками SIP-сервера // Сборник избранных трудов VIII Международной научно- практической конференции «Современные информационные технологии и ИТ- образование». Под ред. В.А. Сухомлина, 8-10 ноября 2013 г. – М.: ИНТУИТ.РУ, 499- 504 с.
3. Таланова М.О. Построение модели механизма просеивания нагрузки с двумя порогами для управления перегрузками в сети прокси-серверов // Материалы Всероссийской конференции с международным участием «ИТТММBC». – М.: РУДН. 22-26 апреля 2013 г. – 55-57 с.

COMPARING SYSTEMS WITH ONE AND TWO THRESHOLDS FOR LOAD CONTROL OF SIP-SERVER

Talanova M.O., Shcherbakova A.E.

Peoples' Friendship University of Russia, matalanova@gmail.com, aewerbakova@gmail.com

We developed model of hysteretic overload control with one and two thresholds. The values of characteristics were obtained for compare the two systems.

Key words: SIP-server, hysteretic load control, LBOC.