

УДК 519.872

А.Н. Моисеев, С.П. Моисеева

ИССЛЕДОВАНИЕ ВХОДЯЩЕГО ПОТОКА ДЛЯ GRID-СИСТЕМЫ С АДАПТИРУЕМЫМ ВЫДЕЛЕНИЕМ ВЫЧИСЛИТЕЛЬНЫХ РЕСУРСОВ

В работе предложена математическая модель GRID-системы с адаптируемым выделением вычислительных ресурсов. Данная модель представлена в виде системы массового обслуживания с входящим потоком, имеющим несколько уровней интенсивности, и блоками обслуживания, соответствующими каждому такому уровню. Основное внимание сосредоточено на получении вероятностных характеристик входящего потока. Отдельно рассмотрена проекция входящего потока на один уровень интенсивности. В работе получены вероятностные характеристики потока для отдельных уровней интенсивности входящего потока, а также корреляционные характеристики разных режимов работы системы.

Ключевые слова: *GRID-система, системы массового обслуживания, марковский модулированный поток.*

Внимание к теории массового обслуживания в значительной степени обусловлено необходимостью применения результатов этой теории к важным практическим задачам, возникающим в связи с бурным развитием систем коммуникаций, эволюцией информационно-вычислительных комплексов, развитием автоматизированных систем управления.

Производительность вычислительных сетей связана с временными аспектами функционирования. При оценке производительности первостепенное значение имеет продолжительность вычислительных процессов. Случайный характер процессов формирования, обработки и передачи данных обуславливает необходимость применения стохастических моделей, в качестве которых широко используются модели, представляющие собой системы и сети массового обслуживания различных классов.

Следуя необходимости создания адекватных моделей различных явлений и систем, многие исследователи разработали схемы потоков событий, при помощи которых можно учитывать различные реальные факторы и, в частности, зависимость между поступающими требованиями. Д. Кокс [1] рассмотрел потоки однородных событий, интенсивность которых зависит от состояний управляющего потоком процесса. Позже были даны общие определения такого потока [2–4]. Одним из наиболее распространенных частных случаев марковских входящих потоков (МАР-потоков) является марковский модулированный пуассоновский поток событий (ММРР-поток). Исследованию ММРР-потока посвящены работы зарубежных и российских ученых [5, 6].

Как правило, все работы посвящены методам исследования процесса, характеризующего число наступивших событий в потоке за некоторое время. В то же время для решения практических задач представляет интерес исследование совокупности потоков различной интенсивности, определяемых состояниями управляющей цепи Маркова. Очевидно, что для оптимального функционирования вы-

числительных систем необходимо учитывать интенсивность входящих задач и предоставлять возможности для их быстрой обработки. То есть потоки заявок большей интенсивности должны обслуживаться быстрее. Если для обслуживания заявок, поступивших на разных уровнях интенсивности, применяются различные блоки серверов, то актуальной является также задача исследования проекций входящего ММРР-потока на каждый из таких блоков.

1. Постановка задачи

Пусть имеется GRID-система [7] – система распределенных вычислений на основе кластеров, где в качестве серверов используются обычные рабочие станции либо серверы, выполняющие повседневные задачи, не связанные с распределенными вычислениями. Таким образом, серверы выделяют под GRID-задачи лишь часть своих вычислительных ресурсов. Рассматривается ситуация, когда задачи в этой системе поступают на обработку с изменяющейся интенсивностью. Чтобы оптимизировать исполнение поступающих задач в условиях переменной интенсивности, владелец GRID-системы принял решение настроить всю систему таким образом, что при возрастании интенсивности поступления серверы будут выделять больше вычислительных ресурсов для исполнения GRID-задач, чтобы не создавать задержек в их выполнении. При уменьшении интенсивности поступления серверы уменьшают выделенные под GRID-задачи вычислительные ресурсы, чтобы возвратиться к решению повседневных задач.

Построим математическую модель описанной задачи в виде бесконечнолинейной системы массового обслуживания (СМО), на вход которой поступает поток событий с изменяющейся интенсивностью. Будем представлять его в виде ММРР-потока [6] с K состояниями, каждое из которых определяет собственную интенсивность событий λ_s ($s = \overline{1, K}$). Переходы между состояниями управляющей цепи задаются матрицей инфинитезимальных характеристик $\mathbf{Q} = \|q_{ij}\|$.

Основной целью исследования является получение вероятностных характеристик входящего потока. Отдельный интерес представляет проекция входящего потока на каждый из уровней интенсивности. Это актуально для случая, когда заявки, поступившие при разных интенсивностях (состояниях управляющей цепи входящего потока), обслуживаются на разных группах серверов, имеющих различную, заранее настроенную производительность обслуживания. В этом случае для одной такой группы ситуация будет выглядеть так, что пока управляющая цепь находится в соответствующем состоянии, заявки поступают в этот блок с определенной интенсивностью, когда же состояние управляющей цепи меняется, заявки начинают поступать в другие блоки, а на входе данного блока наступает время «тишины».

2. Входящий поток

Для начала получим характеристики входящего потока всей системы. Пусть $k(t)$ – состояние управляющей цепи входящего ММРР-потока в момент времени t , $n_1(t), \dots, n_K(t)$ – количество событий входящего потока, произошедших на каждом из K уровней интенсивности за время t . Рассмотрим многомерный случайный процесс $\{k(t), n_1(t), \dots, n_K(t)\}$. Обозначим через $P(k, n_1, \dots, n_K, t) = \mathbf{P}\{k(t) = k,$

$n_1(t) = n_1, \dots, n_K(t) = n_K$. Для этой функции можно записать следующее выражение:

$$P(k, n_1, \dots, n_K, t + \Delta t) = P(k, n_1, \dots, n_K, t)(1 - \lambda_k \Delta t)(1 + q_{kk} \Delta t) + \\ + P(k, n_1, \dots, n_k - 1, \dots, n_K, t)\lambda_k \Delta t + \sum_{\substack{v=1 \\ v \neq k}}^K P(v, n_1, \dots, n_K, t)q_{vk} \Delta t + o(\Delta t).$$

Выполнив несложные преобразования и перейдя к пределу при $\Delta t \rightarrow 0$, получим

$$\frac{\partial P(k, n_1, \dots, n_K, t)}{\partial t} = \\ = -P(k, n_1, \dots, n_K, t)\lambda_k + P(k, n_1, \dots, n_k - 1, \dots, n_K, t)\lambda_k + \sum_{\substack{v=1 \\ v \neq k}}^K P(v, n_1, \dots, n_K, t)q_{vk}.$$

Воспользуемся методом производящих (характеристических) функций [6]. Для этого умножим левую и правую части полученного уравнения на $e^{ju_1 n_1} \dots e^{ju_K n_K}$, где $j = \sqrt{-1}$, а u_1, \dots, u_K – некоторые переменные, и просуммируем по n_1, \dots, n_K от 0 до ∞ . Введя обозначение

$$H(k, u_1, \dots, u_K, t) = \sum_{n_1=0}^{\infty} \dots \sum_{n_K=0}^{\infty} e^{ju_1 n_1} \dots e^{ju_K n_K} P(k, n_1, \dots, n_K, t),$$

получим уравнение относительно функции $H(k, u_1, \dots, u_K, t)$:

$$\frac{\partial H(k, u_1, \dots, u_K, t)}{\partial t} = H(k, u_1, \dots, u_K, t)(e^{ju_k} - 1)\lambda_k + \sum_{v=1}^K H(v, u_1, \dots, u_K, t)q_{vk}. \quad (1)$$

Обозначим через $\mathbf{H}(u_1, \dots, u_K, t)$ вектор-строку, состоящую из компонент $H(1, u_1, \dots, u_K, t), \dots, H(K, u_1, \dots, u_K, t)$. Пусть $\mathbf{B}(u_1, \dots, u_K) = \text{diag} \left\{ (e^{ju_s} - 1)\lambda_s \right\} + \mathbf{Q}$.

Тогда (1) в матричном виде запишется следующим образом:

$$\frac{\partial \mathbf{H}(u_1, \dots, u_K, t)}{\partial t} = \mathbf{H}(u_1, \dots, u_K, t)\mathbf{B}(u_1, \dots, u_K).$$

Решение этого уравнения записывается с помощью матричной экспоненты [8]

$$\mathbf{H}(u_1, \dots, u_K, t) = \text{Rexp} \{ \mathbf{B}(u_1, \dots, u_K) t \}, \quad (2)$$

где элементы вектор-строки $\mathbf{R} = \{R(1), \dots, R(K)\}$ находятся из начальных условий

$$R(k) = H(k, u_1, \dots, u_K, 0) = P \{ k(0) = k \}.$$

3. Проекция входящего потока

Рассмотрим теперь проекцию входящего потока на блок одной интенсивности, пусть это будет блок номер s . Положим в уравнении (1) все u_1, \dots, u_K , кроме u_s , равными нулю. Будем использовать обозначение $H(k, u_s, t) = H(k, 0, \dots, 0, u_s, 0, \dots, 0, t)$. Тогда уравнение (1) перепишется в виде

$$\frac{\partial H(s, u_s, t)}{\partial t} = H(s, u_s, t)(e^{ju_s} - 1)\lambda_s + \sum_{v=1}^K H(v, u_s, t)q_{vs} \text{ для } k = s;$$

$$\frac{\partial H(k, u_s, t)}{\partial t} = \sum_{v=1}^K H(v, u_s, t)q_{vk} \text{ для } k \neq s;$$

или в матричном виде

$$\frac{\partial \mathbf{H}(u_s, t)}{\partial t} = \mathbf{H}(u_s, t)\mathbf{B}_s(u_s),$$

где $\mathbf{H}(u_s, t)$ – вектор-строка с компонентами $H(k, u_s, t), k = \overline{1, K}$, а элементы матрицы $\mathbf{B}_s(u_s)$ равны элементам матрицы \mathbf{Q} за исключением элемента b_{ss} , который равен $(e^{ju_s} - 1)\lambda_s + q_{ss}$.

Решение этого уравнения имеет вид

$$\mathbf{H}(u_s, t) = \text{Rexp}\{\mathbf{B}_s(u_s)t\}.$$

Введем обозначение

$$h(u_s, t) = \sum_{k=1}^K H(k, u_s, t) = \mathbf{H}(u_s, t)\mathbf{E} = \text{Rexp}\{\mathbf{B}_s(u_s)t\}\mathbf{E},$$

где \mathbf{E} – вектор-столбец, состоящий из единиц. Тогда, учитывая, что

$$H(k, u_s, t) = \sum_{n_s=0}^{\infty} e^{ju_s n_s} P(k, n_s, t)$$

и, следовательно,

$$P(k, n_s, t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ju_s n_s} H(k, u_s, t) du_s,$$

получаем

$$P(n_s, t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ju_s n_s} h(u_s, t) du_s. \quad (3)$$

Здесь $P(n_s, t) = \mathbf{P}\{n_s(t) = n_s\}$ – вероятность того, что за время t в систему поступило n_s заявок с интенсивностью λ_s .

4. Основные вероятностные характеристики

Получим первый момент этого распределения. Про дифференцируем (1) по u_s и положим все $u_i = 0, i = \overline{1, K}$. Учитывая, что

$$\left. \frac{\partial H(k, u_1, \dots, u_K, t)}{\partial u_s} \right|_{u_1 = \dots = u_K = 0} = j \sum_{n_s=0}^{\infty} n_s P(k, n_s, t) = j m(k, s, t),$$

получим

$$\frac{\partial m(k, s, t)}{\partial t} = R(k)\lambda_k + \sum_{v=1}^K m(v, s, t)q_{vk} \text{ для } k = s; \quad (4)$$

$$\frac{\partial m(k, s, t)}{\partial t} = \sum_{v=1}^K m(v, s, t)q_{vk} \text{ для } k \neq s. \quad (5)$$

Здесь функция $m(k,s,t)$ имеет смысл среднего числа заявок уровня s , поступивших за интервал длины t , в то время, когда состояние управляющей цепи равно k . Тогда $m(s,t) = \sum_{k=1}^K m(k,s,t)$ есть среднее число заявок уровня s , поступивших за интервал времени длины t (первый момент). Суммируя уравнения системы (4), (5) и учитывая, что $\sum_{k=1}^K q_{vk} = 0$, получаем дифференциальное уравнение относительно этой функции:

$$\frac{dm(s,t)}{dt} = R(s)\lambda_s .$$

С учетом начального условия $m(s,0) = 0$ это уравнение имеет следующее решение:

$$m(s,t) = R(s)\lambda_s t . \tag{6}$$

Найдем второй начальный момент распределения (3). Продифференцируем (1) по u_s и u_i , где s и i – некоторые числа из диапазона от 1 до K . Учтем, что

$$\left. \frac{\partial^2 H(k, u_1, \dots, u_K, t)}{\partial u_s^2} \right|_{u_1 = \dots = u_K = 0} = j^2 \sum_{n_s=0}^{\infty} n_s^2 P(k, n_s, t) = j^2 m_2(k, s, t)$$

и

$$\left. \frac{\partial^2 H(k, u_1, \dots, u_K, t)}{\partial u_s \partial u_i} \right|_{u_1 = \dots = u_K = 0} = j^2 \sum_{n_s=0}^{\infty} \sum_{n_i=0}^{\infty} n_s n_i P(k, n_s, n_i, t) = j^2 r(k, s, i, t) \text{ для } s \neq i .$$

Здесь функции $m_2(k,s,t)$ и $r(k,s,i,t)$ имеют смысл соответственно второго начального момента и смешанного момента для числа заявок, поступивших за время t на уровнях интенсивности s и i при условии, что управляющая цепь находилась в состоянии k . В результате получаем следующие системы дифференциальных уравнений. Относительно второго момента:

$$\frac{\partial m_2(k, s, t)}{\partial t} = 2m(k, s, t)\lambda_k + R(k)\lambda_k + \sum_{v=1}^K m_2(v, s, t)q_{vk} \text{ для } k = s; \tag{7}$$

$$\frac{\partial m_2(k, s, t)}{\partial t} = \sum_{v=1}^K m_2(v, s, t)q_{vk} \text{ для } k \neq s. \tag{8}$$

Относительно смешанного момента:

$$\frac{\partial r(k, s, i, t)}{\partial t} = m(k, s, t)\lambda_k + m(k, i, t)\lambda_k + \sum_{v=1}^K r(v, s, i, t)q_{vk} \text{ для } k = s; \tag{9}$$

$$\frac{\partial r(k, s, i, t)}{\partial t} = \sum_{v=1}^K r(v, s, i, t)q_{vk} \text{ для } k \neq s. \tag{10}$$

Функция $m_2(s,t) = \sum_{k=1}^K m_2(k, s, t)$ – это второй начальный момент распределения вероятностей (3) числа заявок, пришедших на уровне интенсивности s за время t , а

$r(s, i, t) = \sum_{k=1}^K r(k, s, i, t)$ – смешанный момент числа заявок, пришедших за время t

на уровнях интенсивности s и i . Суммируя (7), (8) и (9), (10) по k от 1 до K , получаем дифференциальные уравнения относительно этих искомым функций:

$$\frac{\partial m_2(s, t)}{\partial t} = 2m(s, t)\lambda_s + R(s)\lambda_s; \quad (11)$$

$$\frac{\partial r(s, i, t)}{\partial t} = [m(s, t) + m(i, t)]\lambda_s. \quad (12)$$

Здесь функции $m(s, t)$ определяются из (6). Решения этих уравнений при начальных условиях $m_2(s, 0) = 0$ и $r(s, i, 0) = 0$ могут быть легко получены для каждого частного случая.

Заключение

Таким образом, в работе проведено исследование входящего потока для модели GRID-системы с адаптируемым выделением вычислительных ресурсов, представленной в виде системы массового обслуживания с входящим ММРР-потокм и блоками обслуживания различной производительности. Получено выражение для характеристической функции (2) многомерного распределения числа заявок, поступивших на каждом из уровней интенсивности, распределение вероятностей (3) числа заявок одного уровня, а также первый момент (6) и общий вид систем обыкновенных дифференциальных уравнений для вычисления второго начального (11) и смешанного (12) моментов этого распределения. Полученные результаты могут быть использованы на практике при построении GRID-систем с соответствующей инфраструктурой.

ЛИТЕРАТУРА

1. Cox D.R. The analysis of non-Markovian stochastic processes // Proc. Cambr. Phil. Soc. 1955. V. 51. N 3. P. 433–441.
2. Serfozo R. Processes with conditional independent increments // Appl. Prob., 1972. V. 9. P. 303–315.
3. Neuts M.F. A versatile Markovian arrival process // J. Appl. Prob. 1979. V. 16. P. 764–779.
4. Lucantoni D.M., Meier-Hellsten K.S., Neuts M.F. A single-server queue with server vacations and a class of non-renewal arrival processes // Adv. Appl. Prob. 1990. No. 22. P. 676–705.
5. Дудин А.Н., Клименок В.И. Системы массового обслуживания с коррелированными потоками. Минск: БГУ, 2000. 175 с.
6. Назаров А.А., Моисеева С.П. Метод асимптотического анализа в теории массового обслуживания. Томск: Изд-во НТЛ, 2006. 112 с.
7. Линец М. Грид – масштабируемый распределенный компьютеринг [Электронный ресурс] // gridclub.ru: Интернет-портал по грид-технологиям. URL: <http://gridclub.ru/library/publication.2007-07-19.5491913210/view> (дата обращения: 31.05.12).
8. Bhatia R. Matrix Analysis. Graduate Texts in Mathematics. V. 169. New York: Springer, 1997. 368 p.

Моисеев Александр Николаевич

Моисеева Светлана Петровна

Томский государственный университет

E-mail: amoiseev@ngs.ru; smoiseeva@mail.ru

Поступила в редакцию 24 мая 2012 г.

Moiseev Alexander N., Moiseeva Svetlana P. (Tomsk State University). **Investigation of input flow for the GRID-system with adaptive providing of computing resources.**

Keywords: GRID-system, queuing system, Markov-modulated Poisson process.

We consider mathematical model of the GRID-system with adaptive providing of computing resources. The model is represented as queuing system with input MMPP-flow and server blocks which service intervals depend on input modulating process state.

We obtained an expression for characteristic function of multidimensional distribution for number of arrivals at each modulating process state. Input flow projection on single modulating process state was particularly considered. It was shown that probability distribution of events number n_s arriving in the flow at modulating process state s during period t is defined as

$$P(n_s, t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ju_s n_s} h(u_s, t) du_s,$$

where $h(u_s, t) = \mathbf{R} \exp\{\mathbf{B}_s(u_s)t\}\mathbf{E}$, \mathbf{R} is row vector of stationary distribution of modulating process state, \mathbf{E} is unit column vector, $\mathbf{B}_s(u_s)$ is a matrix with elements which are equal to elements of infinitesimal matrix \mathbf{Q} for modulating process except the single element b_{ss} which is equal to $(e^{ju_s} - 1)\lambda_s + q_{ss}$; λ_s is input flow intensity at state s , q_{ss} is element of the matrix \mathbf{Q} .

First moment and general form of differential equations system for second and mixed moments of the considered distribution are obtained in the paper also. Results of the paper can be applied to GRID-system construction practice.