УДК 519.872

ОБ ОДНОЙ МОДЕЛИ УПРАВЛЯЕМОЙ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ

А.В. Кузнецов, А.С. Мандель, А.Б. Токмакова

Институт проблем управления им. В.А. Трапезникова РАН, г. Москва

Исследована одна из моделей управляемых систем массового обслуживания. В качестве управляемого и динамически оптимизируемого параметра выступает число рабочих устройств обслуживания в многолинейной системе массового обслуживания, которое выбирается как функция от изменяющейся по времени интенсивности входящего потока требований. Процесс изменения во времени интенсивности входящего потока является дискретной по времени цепью Маркова с конечным числом состояний.

ВВЕДЕНИЕ

Рассматриваемая модель управляемой системы массового обслуживания (СМО) с переменным и выбираемым в динамике числом рабочих устройств обслуживания широко применяется в сфере анализа деятельности таких социально-экономических объектов, как системы продажи различного рода билетов, бронирования мест в гостиницах, а также торгово-производственные, транспортные и многие другие системы.

Для решения подобных задач необходимо располагать достаточно точной моделью управляемой системы, которая строится с помощью классических методов теории массового обслуживания [1, 2], и адекватным аппаратом решения соответствующих задач дискретной по времени оптимизации. Значительный вклад в создание инструмента решения подобных задач принадлежит Ричарду Беллману, который в конце 1950-х — начале 1960-х гг. предложил аппарат дискретного динамического программирования [3]. Основное достоинство моделей динамического программирования заключается в том, что во многих случаях они позволяют получать серьезные аналитические результаты относительно оптимальных управлений. Одним из примеров подобных результатов служит установление факта оптимальности двухуровневых стратегий управления запасами [4]. Применительно к марковским моделям систем метод Беллмана был исследован и развит Р. Ховардом [5], который назвал рассмотренный в его работах класс моделей (и соответствующие алгоритмы) марковскими процессами с доходами.

Интерес к управляемым СМО возникает в 1970-е гг. В связи с этим отметим фундаментальную работу В.В. Рыкова [6], в которой была приведена достаточно общая постановка задачи управления системами массового обслуживания и предложены подходы к ее решению. Было отмечено, что управлять в СМО можно дисциплиной обслуживания требований, выбором их структуры, воздействиями на входящий поток требований, изменением интенсивности обслуживания требований и рядом других параметров. Оригинальным вкладом в исследование возможности постановки и решения задачи управления так называемыми конфликтными СМО (т. е. системами, которые описываются моделями теоретико-игрового типа) стала работа Ю.И. Неймарка [7].

К концу 1980-х гг. интерес к управляемым СМО заметно возрос в связи с широким распространением принципиально нового класса прикладных систем, которые можно было интерпретировать и анализировать как системы массового обслуживания. Этим классом стали вычислительные и информационные компьютерные сети. Именно поэтому усилия специалистов в области теории массового обслуживания оказались направлены на разработку специфических моделей (в том числе и сетевых) массового обслуживания, рассчитанных на применение к указанному классу прикладных систем. Особо отметим вклад теоретиков, представляющих школы В.М. Вишневского и В.А. Каштанова [8, 9].

Таким образом, в достаточно продолжительном интервале времени, который исчисляется примерно 20 годами (с 1985 по 2005 г.), практически вне зоны внимания большинства специалистов по теории масового обслуживания оставались такие модели управляемых СМО, которые были бы пригодны для описания, анализа и оптимизации функционирования объектов социально-экономической природы, к числу которых можно относить различные системы продажи билетов, системы бронирования мест в гостиницах, торгово-закупочные, транспортные и другие системы. Специфика объектов социально-экономической природы заключается, прежде всего, в том, что при их описании значительную роль играет эконометрический анализ их функционирования, выражающийся в тщательном и корректном учете всех экономических показателей, которые необходимо принимать во внимание при оценке качества функционирования таких объектов.

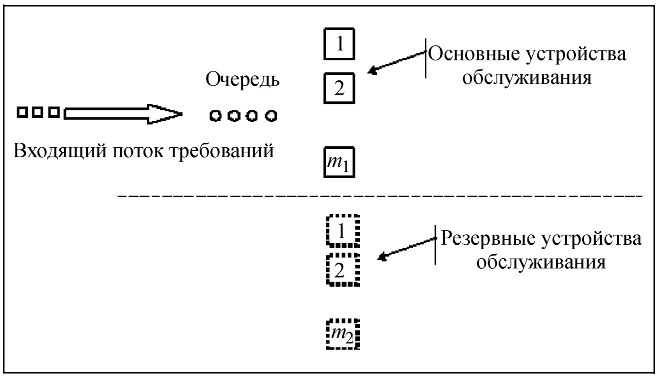
В настоящей статье рассматривается пример формирования многокомпонентного функционала экономической эффективности многолинейной СМО. С применением построенного функционала сформулирована и решена задача оптимизации функционирования рассматриваемой СМО.

1. ПОСТАНОВКА ЗАДАЧИ

Рассмотрим многолинейную СМО с набором подключаемых резервных устройств обслуживания, схематично представленную на рисунке.

Интенсивность λ входящего потока в фиксированные моменты времени $t_n = n\tau$, $n = 1, 2,..., N$, претерпевает скачкообразные изменения от λ_i к λ_j с вероятностями $p_{ij}(m_1)$[1], $i, j = 1, 2,..., L$ (в принципе возможен случай, когда переходные вероятности зависят также и от n, тогда вместо $p_{ij}(m_1)$ следует писать $p_{ij}(m_1, n)$). Таким образом, в каждом из интервалов длительности τ интенсивность входящего потока требований постоянна и описывается

[1] Введение зависимости вероятностей скачков интенсивности входящего потока от числа рабочих устройств m_1 связано с необходимостью учета эффекта влияния качества процесса обслуживания (в данном случае обусловленного временем ожидания клиентами начала обслуживания) на потребительский спрос, который и описывает интенсивность входящего потока.



Структура управляемой системы массового обслуживания

моделью случайной величины, принимающей значения из конечного множества $\Lambda = \{\lambda_1, \lambda_2,..., \lambda_L\}$ (считаем, что $\lambda_1 < \lambda_2 <...< \lambda_L$). Будем также считать, что $M\mu > \lambda_L$, где μ — это интенсивность обслуживания на одном основном (рабочем) устройстве, а $M = m_1 + m_2$ (в обозначениях рисунка).

В моменты времени t_n имеется возможность переводить устройства обслуживания из числа резервных в основные и обратно, т. е. осуществлять переключения резервных устройств в основные (включение) или обратные переключения: из основных — в резервные (отключение). Рассматриваемая модель массового обслуживания по своей идее весьма близка к предложенной в 60-е гг. прошлого столетия модели надежности, получившей название динамического резервирования [10, 11], хотя, как станет ясно из раздела 2 настоящей статьи, ее алгоритмическое решение качественно гораздо ближе к схеме двухуровневых стратегий управления запасами [4].

Будем рассматривать случай стационарных режимов в системе массового обслуживания, т. е. случай, когда в каждом из интервалов длительности τ успевает поступить (и пройти обслуживание) такое число требований, которое достаточно велико для того, чтобы в рассматриваемой СМО на каждом шаге устанавливался стационарный в вероятностном смысле режим функционирования. Практически для этого достаточно, чтобы

$$\mu\tau \gg 1. \tag{1}$$

Разумеется, в этом случае справедливо и $\lambda\tau \gg 1$.

Предположение стационарности накладывает достаточно жесткое ограничение на рассматриваемую модель.

Прежде всего, должны быть выполнены соотношения $\rho_i = \lambda_i/\mu M < 1$, $\forall\lambda_i \in \Lambda$. Таким образом, система может устойчиво работать только с таким потоком, для которого выполнено условие

$$\lambda_i < \mu M, \quad \forall \lambda_i \in \Lambda. \qquad (2)$$

Пусть λ_{max} — максимальный элемент множества Λ (фактически, это значение λ_L). Тогда неравенство (2) можно переписать в виде

$$\lambda_{max} < \mu M. \qquad (3)$$

Далее, на каждом шаге существует нижняя граница допустимого числа рабочих устройств $\underline{m_{1p}^{(i)}}$, на которое можно переводить систему с расчетом на ее устойчивое функционирование (в смысле существования стационарного режима), и эта граница зависит от номера состояния i на соответствующем шаге.

Действительно, в состоянии i (по значению интенсивности входящего потока) должно выполняться условие стационарности $\rho_i = \lambda_i/\mu m_{1p}^{(i)} < 1$, откуда получаем

$$m_{1p}^{(i)} > \frac{\lambda_i}{\mu} \Rightarrow \underline{m_{1p}^{(i)}} = \frac{\lambda_i}{\mu} + 1, \quad \forall i = \overline{1, N}. \qquad (4)$$

Таким образом, число работающих в состоянии i устройств может принимать значения от $\underline{m_{1p}^{(i)}}$ до M.

Зададим компоненты затрат.

Пусть

c_1 — стоимость эксплуатации одного основного (рабочего) устройства обслуживания в единицу времени;

c_2 — стоимость содержания одного резервного устройства обслуживания в единицу времени (естественно, что $c_1 > c_2$, нередко $c_2 = 0$);

A_1 — цена переключения одного устройства из числа резервных в число основных (цена «включения»);

A_2 — цена переключения одного устройства из числа основных в число резервных (цена «отключения»);

d — стоимость единицы времени пребывания одного требования в очереди на обслуживание;

h — доход, связанный с окончанием обслуживания одного требования.

2. АЛГОРИТМ ДИНАМИЧЕСКОГО ПРОГРАММИРОВАНИЯ

Обозначим через $C^{(1)}$ (λ_i, m_1, m_{1p}) среднее значение прибыли на одном шаге (длительности τ), если в начале шага устанавливается значение интенсивности входящего потока, равное λ_i; число основных (рабочих) устройств, с которыми СМО подходит к этому шагу, равно m_1; и принимается управляющее решение о введении в действие момент начала шага m_{1p} основных устройств. Это означает сохранение прежнего числа работающих устройств обслуживания, если $m_1 = m_{1p}$; необходимость отключения $(m_1 - m_{1p})$ устройств, если $m_1 - m_{1p} > 0$; и необходимость включения $(m_{1p} - m_1)$ устройств, если $m_1 - m_{1p} < 0$.

Очевидно, что

$C^{(1)}(\lambda i, m_1, m_{1p})$ = средний доход за один шаг — средние одношаговые затраты.

Из предположения о стационарности следует, что выполняются все входящие требования, а значит, в любом случае средний доход за один шаг составляет $h\lambda_i$.

Средние затраты складываются из затрат на включение резервных (отключение основных) устройств; затрат, связанных с пребыванием требований в очереди; и затрат на эксплуатацию основных и резервных устройств.

Затраты на включение (отключение)

$$З_{\text{переключения}}(m_1, m_{1p}) = \begin{cases} A_1(m_{1p} - m_1), & \text{если } m_{1p} - m_1 > 0, \\ 0, & \text{если } m_{1p} = m_1, \\ A_2(m_1 - m_{1p}), & \text{если } m_{1p} - m_1 < 0. \end{cases}$$

Затраты, связанные с пребыванием требований в очереди на шаге длительности τ,

$$З_{\text{на очередь}} = \mathrm{E}\int_0^\tau dk(t, m_{1p}, \lambda_i)dt = d\int_0^\tau \mathrm{E}k(m_{1p}, \lambda_i)dt = d\bar{k}(m_{1p}, \lambda_i)\tau,$$

где E — символ оператора математического ожидания, $k(t, m_{1p}, \lambda_i)$ — число требований в очереди к СМО с параметрами m_{1p} и λ_i в момент времени t, а $\bar{k}(m_{1p}, \lambda_i)$ — средняя длина очереди в СМО с параметрами m_{1p} и λ_i. Интегрирование в силу предположения о стационарности ведется не в текущем интервале $[t_n, t_{n+1}]$, а в интервале $[0, \tau]$. Для стационарного случая имеем [1]:

$$\bar{k}(m_{1p}, \lambda_i) = \pi_0 \frac{(m_{1p}\rho_i)^{m_{1p}}\rho_i}{m_{1p}!(1-\rho_i)^2},$$

где $\rho_i = \dfrac{\lambda_i}{\mu m_{1р}}$, а $\pi_0 = \left[\sum\limits_{i=0}^{m_{1р}} \dfrac{(m_{1р}\rho_i)^i}{i!} + \dfrac{(m_{1р}\rho_i)^{m_{1р}}}{m_{1р}!(1-\rho_i)}\right]^{-1}$ — стационарная вероятность того, что в системе нет требований.

Параметр d характеризует косвенные издержки рассматриваемой СМО, связанные с тем, что увеличение длительности времени пребывания требований в очереди снижает предпочтения по обращению в данную СМО клиентов с требованиями на обслуживание. Можно построить математическую модель, которая связывала бы уровень утраты предпочтений со средним временем пребывания требований в очереди в случае его увеличения, и на основе этой модели сформировать обоснованную оценку значения коэффициента d. В настоящей работе задача построения подобной модели не рассматривается. Ограничимся замечанием, что в большинстве случаев значение параметра d на 1—2 порядка меньше значения введенного выше параметра h.

Затраты на эксплуатацию основных и резервных устройств за один шаг составляют:

$$З_{\text{эксплуатации}}(m_{1р}) = c_1 m_{1р} + c_2(M - m_{1р}).$$

Таким образом, средняя чистая прибыль за шаг составит:

$$C^{(1)}(\lambda_i, m_1, m_{1р}) =$$
$$= h\lambda_i - З_{\text{переключения}} - З_{\text{на очередь}} - З_{\text{эксплуатации}}.$$

Пусть $П_s^*(\lambda_i, m_1)$ — максимальное значение средней прибыли на интервале, который начинается за s шагов до конца периода планирования $[0, T]$, $s + n = N$, при значении λ_i интенсивности входящего потока и m_1 включенных (до принятия управляющего решения о включении $m_{1р}$ устройств) основных устройствах. Ниже выводятся уравнения динамического программирования для функционала $П_s^*(\lambda_i, m_1)$ с учетом условия (4).

Очевидно, что за один шаг до конца периода планирования при случайном значении интенсивности входящего потока λ_i значение введенного выше функционала $П_s^*(\lambda_i, m_1)$ запишется как

$$П_1^*(\lambda_i, m_1) = \max_{\underline{m}_{1р}^{(i)} \le m_{1р} \le M} C^{(1)}(\lambda_i, m_1, m_{1р}), \quad (5)$$

где $\lambda_i \in \Lambda$.

За s шагов до конца периода планирования имеем:

$$П_s^*(\lambda_i, m_1) = \max_{\underline{m}_{1р}^{(i)} \le m_{1р} \le M} \Big\{ C^{(1)}(\lambda_i, m_1, m_{1р}) +$$
$$+ \sum_{j=1}^{L} p_{ij}(m_{1р})\, П_{s-1}^*(\lambda_j, m_{1р}) \Big\}, \quad (6)$$

где $\forall s = \overline{2, N-1}$, $\lambda_i, \lambda_j \in \Lambda$ — значения интенсивностей входящего потока в моменты времени за s и за $s - 1$ шагов до конца периода планирования соответственно.

Уравнения (5) и (6) используются для расчета программы оптимального управления $m_{1р}^*(i, s)$, где $1 \leqslant s \leqslant N - 1$, $i = 1, 2,..., L$.

3. ОБСУЖДЕНИЕ. МОДЕЛИ И АЛГОРИТМЫ

Следует признать, что при случайном характере входящего потока требований построение долгосрочных стратегий управления вряд ли окажется достаточно эффективным. Исходя из этого, разумно рассматривать такие значения N, которые отвечают реальным длительностям периода планирования. В зависимости от прикладной задачи это могут быть сутки, неделя, месяц или, самое большее, квартал.

Также естественно предполагать, что число устройств обслуживания в рассматриваемой СМО достаточно велико для того, чтобы удавалось обслужить все входящие требования. При этом в прикладном плане следует считать, что издержки содержания одного резервного устройства обслуживания достаточно малы по сравнению с издержками содержания рабочего устройства. Иначе говоря, полагается, что $c_2 \ll c_1$.

Предположение о стационарности потока на каждом временном интервале накладывает серьезные ограничения на допустимое множество значений переменных и заметно сужает область перебора (неравенства (3), (4)). Отметим дополнительно, что в отдельных случаях указанное условие может оказаться настолько действенным, что выстроенная в результате программа оптимального управления будет в значительной степени определяться именно этим условием, а не процедурой максимизации целевой функции. Иначе говоря, набор решений будет почти однозначно определяться вектором $\underline{m}_{1р}^{(i)}$, $i = 1, 2,..., L$.

ЗАКЛЮЧЕНИЕ

Предложена и исследована новая модель управления многолинейной системой массового обслуживания, с помощью которой можно анализировать и оптимизировать функционирование объектов социально-экономической природы различных классов: систем продажи и бронирования билетов и мест в гостиницах, торгово-закупочных систем, систем общественного транспорта, билинговых систем и ряда других.

Выписан и покомпонентно проанализирован функционал эффективности функционирования рассматриваемой многолинейной СМО. На основе проведенного анализа построена модель выбора оптимальных решений. Обсуждаются возможности практического использования предложенной модели и ограничения на ее применения.

ЛИТЕРАТУРА

1. *Гнеденко Б.В., Коваленко И.Н.* Введение в теорию массового обслуживания. — М.: Наука, 1966.
2. *Саати Т.Л.* Элементы теории массового обслуживания и ее приложения — М.: Сов. радио, 1971.
3. *Беллман Р.* Динамическое программирование. — М.: ИИЛ, 1960. — 400 с.
4. *Хедли Дж., Уайтин Т.* Анализ систем управления запасами. — М.: Наука, 1969. — 512 с.
5. *Ховард Р.А.* Динамическое программирование и марковские процессы. — М.: Сов. радио. 1984. — 190 с.
6. *Рыков В.В.* Управляемые системы массового обслуживания // Теория вероятностей. Математическая статистика. Теоретическая кибернетика. — 1975. — Т. 12. — С. 43—153.
7. *Неймарк Ю.И.* Динамические системы и управляемые процессы. — М.: Наука, 1978.
8. *Вишневский В.М.* Теоретические основы проектирования компьютерных сетей. — М.: Техносфера, 2003.
9. *Захаров П.П.* Разработка автоматизированного программного комплекса для исследования качества и эффективности функционирования моделей технических систем и управляемых систем массового обслуживания: Дис. канд. техн. наук. — М.: МГИЭМ, 2006.
10. *Райкин А.Л.* Элементы теории надежности для проектирования технических систем. — М.: Сов. радио, 1967. — 256 с.
11. *Мандель А.С., Райкин А.Л.* Формирование оптимального плана включения запасных элементов / Автоматика и телемеханика. — 1967. — № 5. — С. 55—63.

☎ *(495) 334-89-69, e-mail: manfoon@ipu.ru*

Статья представлена к публикации членом редколлегии Ф.Т. Алескеровым. ❑