

ЗАДАЧА ОПТИМАЛЬНОГО РАСПРЕДЕЛЕНИЯ РЕСУРСА МЕЖДУ СОСТАВЛЯЮЩИМИ ЦЕНТРА ОБСЛУЖИВАНИЯ ВЫЗОВОВ

Червяков Н.И., Листова Н.В.

Выделены основные составляющие центра обслуживания вызовов в виде канальной, операторской и технической подсистем, каждая из которых представлена соответствующей моделью массового обслуживания. Осуществлена постановка задачи максимизации вероятности обслуживания заявки при ограниченных финансовых ресурсах за счет выбора оптимального числа обслуживающих элементов в указанных подсистемах. Определены возможность и алгоритм применения метода динамического программирования для решения данной задачи.

Введение

Индустрия центров обслуживания вызовов (ЦОВ) или контакт-центров, являющихся эффективным инструментом взаимодействия компаний с клиентами, развивается стремительно как по производительности, так и по экономическим показателям [1-2]. Несмотря на широкий размах отрасли за рубежом, где доходы от использования ЦОВ исчисляются миллиардами долларов, в России многие компании, взаимодействующие с большим числом клиентов, не имеют собственных ЦОВ, опасаясь больших расходов на их содержание [3].

Оптимальным решением в этой ситуации является аутсорсинг ЦОВ, т.е. аренда на определенный период некоторых из его функций [4]. При этом возникает задача оперативного расчета количества арендуемых каналов связи, числа операторов, обрабатывающих заявки клиентов, и номенклатуры технического персонала, удовлетворяющих заявки по устранению возникших проблемных ситуаций, с учетом выделяемых финансовых ресурсов на аренду ЦОВ. Решение данной задачи сводится к максимизации прибыли от использования ЦОВ, косвенным показателем которой выступает вероятность удовлетворения заявок клиентов.

Несмотря на неоднородность структуры ЦОВ для расчета показателей их функционирования использовались простейшие аналитические модели, построенные на базе формул Эрланга [5-6]. Однако применение в качестве целевых функций известных моделей ЦОВ в виде однофазных систем массового обслуживания (СМО) затрудняет решение задачи. Поэтому необходи-

мо рассмотреть ЦОВ как сложную систему, состоящую из трех подсистем: канальной (КПс), операторской (ОПс) и технической (ТПс). Учитывая последовательность прохождения и обработки заявок клиентов через указанные подсистемы, моделью ЦОВ является трехфазная СМО с различными частными моделями на каждом этапе обслуживания.

Модель канальной подсистемы ЦОВ

Канальная подсистема ЦОВ содержит n каналов связи, за которыми закреплен один телефонный номер дозвона в клиент-центр. Предполагается, что звонки абонентов равномерно распределены на исследуемом интервале времени T (свойство стационарности потока заявок) и вероятность появления двух и более звонков в один и тот же момент времени ничтожно мала (свойство ординарности). Если к тому же интенсивность звонков в будущем не зависит от их количества в прошлом, то их можно рассматривать как простейший поток заявок λ_1 , а функционирование КПс исследовать с помощью Марковских процессов.

Каждый поступивший звонок переводит КПс из j -го состояния $S_j^{(1)}$ в состояние $S_{j+1}^{(1)}$, где $j = 0, 1, 2, \dots, n$, при этом состояние $S_0^{(1)}$ соответствует случаю незанятости всех каналов, состояние $S_1^{(1)}$ – занятости одного канала (остальные свободные) и т.д., а состояние $S_n^{(1)}$ – занятости всех каналов. Среднее время занятости канала $\bar{\tau}_1$ определяет интенсивность обслуживания заявки в первой системе $\mu_1 = 1/\bar{\tau}_1$. Так как при занятости всех каналов заявка на обслуживание не принимается, то моделью первой подсистемы ЦОВ является n -канальная СМО с отказами, для которой вероятность отказа заявки на обслуживание определяется по первой формуле Эрланга [7]:

$$P_{отк1} = \left(\lambda_1^n / \mu_1^n n! \right) \cdot \left[\sum_{j=0}^n (\lambda_1 / \mu_1)^j / j! \right], \quad (1)$$

Вероятность того, что заявка будет обслужена (относительная пропускная способность КПс) определяется из выражения (1) [7, 8]:

$$q_1 = 1 - P_{омк1} = 1 - \left(\lambda_1^n / \mu_1^n n! \right) \cdot \left[\sum_{j=0}^n (\lambda / \mu)^j / j! \right]. \quad (2)$$

Модель операторской подсистемы ЦОВ

Если в течение заданного времени T на вход КПС поступило N заявок, то на выходе первой подсистемы будет $N \cdot q_1 = M$ заявок. Следовательно, интенсивность обработанных заявок, поступающих на операторную подсистему ЦОВ:

$$\lambda_2 = M/T = q_1 N/T = \lambda_1 \cdot q_1. \quad (3)$$

Интенсивность обслуживания заявки оператором μ_2 , при этом, как правило, число операторов m гораздо меньше числа обслуживаемых каналов связи n , но даже при занятости всех операторов (состояние $S_m^{(2)}$ второй подсистемы) «необслуженные» каналы ждут своей очереди. В этом случае моделью ОПс является m -канальная СМО с ограниченной очередью, для которой относительная пропускная способность [7]

$$q_2 = 1 - \left(\frac{\lambda_2}{\mu_2} \right)^n \cdot p_0^{(2)} / m^{n-m} m!, \quad (4)$$

где предельная вероятность состояния отсутствия заявок во второй подсистеме

$$p_0^{(2)} = \left\{ 1 + \frac{\lambda_2}{\mu_2} + \dots + \left(\frac{\lambda_2}{\mu_2} \right)^m / m! + \dots + \left(\frac{\lambda_2}{\mu_2} \right)^{m+1} \times \right. \\ \left. \times \left[1 - \left(\frac{\lambda_2}{\mu_2 m} \right)^{n-m} \right] / m \cdot m! \left(1 - \frac{\lambda_2}{\mu_2 m} \right) \right\}^{-1}.$$

Модель технической подсистемы ЦОВ

После уяснения содержания заявки абонента оператор делает запрос в техническую подсистему, соответствующую уровню выполнения заявки. Интенсивность поступления заявок в ТПС:

$$\lambda_3 = q_2 \lambda_2 = q_1 q_2 \lambda_1. \quad (5)$$

Частной моделью ТПС является СМО, содержащая k однотипных каналов обслуживания (технических работников одинаковой квалификации) и характеризующаяся экспоненциальным временем обслуживания со средним значением $\bar{\tau}_{ia}$ или, что эквивалентно, с простейшим потоком обслуживания с интенсивностью $\mu_3 = 1/\tau_{ia}$, независимо от типа обслуживаемой заявки. Дисциплины ожидания в очереди длиной r и обслуживания – беспriorитетные.

Так как значение $\bar{\tau}_{об}$ значительно больше, чем в предыдущих подсистемах, то заявки покидают систему с интенсивностью $\nu = 1/\bar{\tau}_{дон}$, где $\bar{\tau}_{дон}$ – математическое ожидание допустимого времени пребывания заявки в системе, являющееся случайной величиной $\tau_{дон}$ с экспоненциальной плотностью вероятности. Уходы заявок возможны либо из очереди, если время ожидания $t_{ожс} > \tau_{дон}$, либо из канала обслуживания, если $t_{ожс} \leq \tau_{дон}$, что отражается в двух потоках уходов с интенсивностями соответственно $\nu_{ожс}$ и $\nu_{об}$. Таким образом, в ТПС потери заявок возможны либо в форме отказа вследствие переполнения СМО, либо в форме ухода нетерпеливых заявок из системы.

Вероятность отказа $P_{омк3}$ определяется как вероятность нахождения системы в состоянии $S_{k+r}^{(3)}$ [8]. Тогда вероятность обслуживания заявки в ТПС:

$$q_3 = 1 - P_{омк3} = 1 - P_{k+r}^{(3)} = 1 - \left(\frac{\lambda_3}{\mu_3} \right)^{k+r} p_0^{(3)} \times \\ \times \left\{ k! \left(1 + \frac{\nu_{об}}{\mu_3} \right)^k \cdot \prod_{i=1}^r \left[k \left(1 + \frac{\nu_{об}}{\mu_3} \right) + \frac{i \nu_{ожс}}{\mu_3} \right] \right\}^{-1}, \quad (6)$$

где вероятность отсутствия заявок в ТПС [8]:

$$p_0^{(3)} = \left\{ 1 + \sum_{i=1}^k \frac{(\lambda_3/\mu_3)^i}{i! (1 + \nu_{об}/\mu_3)^i} + \frac{(\lambda_3/\mu_3)^k}{k! (1 + \nu_{об}/\mu_3)^k} \times \right. \\ \left. \times \sum_{l=1}^r \prod_{i=1}^l \frac{\lambda_3/\mu_3}{[k(1 + \nu_{об}/\mu_3) + i \nu_{ожс}/\mu_3]} \right\}^{-1}.$$

Формулировка задачи распределения ресурса в ЦОВ

Заявки клиентов будут полностью обслужены в ЦОВ, то есть произойдет событие A , если они будут обслужены во всех подсистемах, что соответствует событиям A_1 , A_2 и A_3 с соответствующими вероятностями q_1 , q_2 и q_3 . Известно [9], что вероятность произведения нескольких событий

$$P(A) = P(A_1 A_2 A_3) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 A_2),$$

где $P(A_d | A_1 \dots A_{d-1})$ – вероятность события A_d при условии, что все предыдущие события имели место. Данные условные вероятности легко получить при выражении интенсивностей входных заявок в ОПс и ТПС с учетом зависимостей (3) и (5). Тогда полная вероятность обслуживания заявки в ЦОВ $P_{обсл} = P(A)$:

$$P_{обсл} = q_1(\lambda_1, n) \cdot q_2(\lambda_1 q_1, m) \cdot q_3(\lambda_1 q_1 q_2, k). \quad (7)$$

При этом каждый из частных показателей определяется по соответствующим моделям СМО вида (2), (4) и (6) и имеет прямую зависимость от количества обслуживающих элементов (ОЭ) – каналов n , операторов m и техников k . Однако рост количества ЭО ограничен суммарным ресурсом C_Σ , выделяемым на аренду каналов связи $\sum_{j=1}^n c_j$, оплату труда операторов $\sum_{b=1}^m c_b$ и технического персонала $\sum_{l=1}^k c_l$.

Обозначим числом x_d количество обслуживающих элементов в d -ой подсистеме ЦОВ (в данном случае $d=1, 3$); $q_d(x_d)$ – вероятность обслуживания заявки в d -ой подсистеме; $\tilde{n}_d(x_d)$ – стоимость ОЭ в количестве x_d в d -й подсистеме. Полученные при расчете в соответствии с моделями (2), (4) и (6) данные, содержащие сведения о показателях обслуживания заявок q_d и стоимостях подсистем c_d , заносятся в таблицу 1, где Λ – некоторое число дискретных значений распределяемых ресурсов (причем $n, m, k \leq \Lambda$).

Таблица 1. Вероятности обслуживания заявок q_d и стоимости c_d подсистем ЦОВ при различных количествах ОЭ x_d

x_d	Подсистема ЦОВ					
	КПс ($d=1$)		ОПс ($d=2$)		ТПс ($d=3$)	
1	$q_1(1)$	$\tilde{n}_1(1)$	$q_2(1)$	$\tilde{n}_2(1)$	$q_3(1)$	$\tilde{n}_3(1)$
2	$q_1(2)$	$\tilde{n}_1(2)$	$q_2(2)$	$\tilde{n}_2(2)$	$q_3(2)$	$\tilde{n}_3(2)$
...
Λ	$q_1(\Lambda)$	$\tilde{n}_1(\Lambda)$	$q_2(\Lambda)$	$\tilde{n}_2(\Lambda)$	$q_3(\Lambda)$	$\tilde{n}_3(\Lambda)$

Задача распределения ресурса между подсистемами ЦОВ формулируется следующим образом: требуется найти вектор параметров $X^* = (x_1^*, x_2^*, x_3^*)$ такой, что показатель ЦОВ

$$P_{обсл} = q_1(x_1) \cdot q_2(x_2) \cdot q_3(x_3) \rightarrow \max_X \quad (8)$$

при условии, что $\left(\sum_{j=1}^n c_j + \sum_{b=1}^m c_b + \sum_{l=1}^k c_l \right) \leq C_\Sigma$.

Решение задачи методом динамического программирования

Учитывая мультипликативность целевой функции (8), решим оптимизационную задачу с помощью метода динамического программирования,

используя процедуру обратной прогонки (начиная с последней подсистемы – ТПс) [7]. При этом d -ый шаг ставится в соответствие d -й подсистеме, состояние z_d выражает суммарные затраты на содержание обслуживающих элементов подсистем с номерами $d, d+1, \dots, \pi$ (в рассматриваемом частном случае $\pi=3$), а вариант решений x_d представляет собой число ОЭ в d -й подсистеме.

Обозначим f_d максимальное значение вероятности обслуживания заявки в цепочке из последовательного соединения подсистем с номерами $d, d+1, \dots, \pi$ при заданной стоимости ОЭ. Тогда рекуррентное соотношение, связывающее f_d для значений индекса d , может быть представлено в виде:

$$f_\pi(z_\pi) = \max_{x_\pi} \{q_\pi(x_\pi)\}; \dots;$$

$$f_d(z_d) = \max_{x_d} \{q_d(x_d) \cdot f_{d+1}[z_d - c_d(x_d)]\} \quad (9)$$

Промежуточные результаты расчета для d -го шага заносятся в таблицу 2, где значения условного выигрыша $f_d(z_d)$ определяются в соответствии с выражением (9) для каждой строки при распределении ресурса в количестве z_d , а условное оптимальное решение

$$x_d^*(z_d) = \arg \max_{x_d} \{q_d(x_d) \cdot f_{d+1}[z_d - c_d(x_d)]\}.$$

Таблица 2. Форма представления результатов промежуточных расчетов условных (по ресурсу) оптимальных параметров на d -ом шаге динамического программирования

z_d	$q_d(x_d) \times$ $\times f_{d+1}[z_d - c_d(x_d)]$			$f_d(z_d)$	$x_d^*(z_d)$
	$x=1$...	$x=\Lambda$		
$z_{d \min}$					
...					
$z_{d \max}$					

При этом для π -го шага величина $f_{d+1} = 1$, а для первого шага переменная z_1 соответствует величине ресурса, распределяемого между всеми подсистемами, то есть $z_1 = C_\Sigma$.

Для решения задачи необходимо предварительно определить границы значений каждой из переменных состояния $z_d, d=1, \pi$.

Для третьего шага наименьшее значение $z_{3 \min} = c_3(x_3 = 1)$ равно затратам на содержание

одного техника. Наибольшее значение z_3 выбирается из условия обеспеченной работоспособности центра обслуживания вызовов при минимальном числе элементов в первой и второй подсистеме (то есть при $x_1 = 1$, $x_2 = 1$). При этом $z_{3\max} = C_\Sigma - [c_1(x_1 = 1) + c_2(x_2 = 1)]$, где $c_1(x_1 = 1)$ и $c_2(x_2 = 1)$ – затраты на аренду одного канала связи и заработную плату одному оператору.

Переменная z_2 представляет собой часть ресурса, который должен быть распределен между второй и третьей подсистемами. Поэтому $z_{2\min} = c_2(x_2 = 1) + c_3(x_3 = 1)$. Максимальное значение ресурса, которое может быть выделено для второй и третьей подсистем, определяется минимальными затратами на первую подсистему:

$$z_{2\max} = C_\Sigma - c_1(x_1 = 1).$$

Значение условного выигрыша на первом шаге (в первой подсистеме – КПс) определяет искомое решение (8), то есть $f_1(C_\Sigma) = \max P_{обсл}$, а также оптимальное число каналов в КПс: $x_1^*(C_\Sigma) = n$. Значения остальных составляющих искомого вектора $X^* = (x_1^*, x_2^*, x_3^*)$ определяются на этапе прямой прогонки, который разворачивается в последовательности, обратной первому этапу (9), используя готовые рекомендации, представленные в таблицах вида 2 для различных шагов предварительного расчета.

Выводы

Центр обслуживания вызовов представляет собой сложную систему, состоящую из последовательно соединенных канальной, операторской и технической подсистем, поэтому для ее моделирования целесообразно использовать многофазную СМО. Частными моделями составляющих ЦОВ являются: для КПс – n -канальная СМО с отказами; для ОПс – m -канальная СМО с ограниченной очередью; для ТПс – k -канальная СМО с ограниченным временем ожидания.

Определение числа обслуживающих элементов в подсистемах ЦОВ (соответственно n , m и k)

осуществляется при решении оптимизационной задачи распределения ограниченного ресурса, а целевой функцией является максимизация произведения условных вероятностей обслуживания заявки в подсистемах центра.

Использование метода динамического программирования предполагает дискретизацию распределяемого ресурса C_Σ , число градаций которого зависит от стоимости единицы ОЭ в каждой подсистеме и влияет на объем вычислительных операций на первом этапе решения задачи оптимизации.

Литература

1. Росляков А.В., Самсонов М.Ю., Шиббаева И.В. Центры обслуживания вызовов (Call Centre). М.: Эко-Трендз, 2002. – 272 с.
2. Степанова И.В. Развитие концепции построения центров обслуживания вызовов // Электро-связь. № 2, 2006. – С. 16-17.
3. Гольдштейн Б.С., Шурыгина С.Б. Еще раз об экономике контакт-центров // Сети и системы связи. № 4, 2003. – С. 12-14.
4. Самолюбова А.Б. Call Center на 100%: Практическое руководство по организации центра обслуживания вызовов. М.: Альпина Бизнес Букс, 2004. – 309 с.
5. Росляков А.В., Ваняшин С.В., Решодько А.А. Сравнительный анализ математических моделей центров обслуживания вызовов // Электро-связь. № 9, 2004. – С. 32-34.
6. Зарубин А.А. Call- и контакт-центры: эволюция технологий и математических моделей // Вестник связи. №8, 2003. – С. 85-89.
7. Вентцель Е.С. Исследование операций: задачи, принципы, методология. М.: Высшая школа, 2001. – 208 с.
8. Вероятностные методы в вычислительной технике. Под ред. А.Н. Лебедева и Е.А. Чернявского. М.: Высшая школа, 1986. – 312 с.
9. Вентцель Е.С., Овчаров Л.А. Теория вероятностей и ее инженерные приложения. – М.: Высшая школа., 2000. – 480 с.

УДК 681.513

МЕТОД АНАЛИЗА ПРОЦЕССА УПРАВЛЕНИЯ УРОВНЕМ ОБСЛУЖИВАНИЯ СИСТЕМЫ УПРАВЛЕНИЯ УСЛУГАМИ СВЯЗИ

Мочалов В.П., Яковлев С.В.

Рассматривается процесс управления уровнем обслуживания системы управления услугами связи. Предлагается метод построения и анализа модели процесса управления уровнем обслуживания, основанный на формализме раскрашенных сетей Петри. Приводятся результаты имитационного моделирования.

Введение

Внедрение новых технологий, позволяющих поставлять на телекоммуникационный рынок все большее количество услуг связи, заставляет мировое телекоммуникационное сообщество взгля-