

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ФУНКЦИОНИРОВАНИЯ ОПЕРАТОРСКОГО ЦЕНТРА

Ваняшин С.В.

Поволжская Государственная Академия телекоммуникаций и информатики

Операторские центры (центры обслуживания вызовов – ЦОВ) получают все большее признание в качестве эффективного инструмента взаимодействия компаний с клиентами [1]. Большинство организаций, работающих с клиентами, а это могут быть как частные компании, так и правительственные организации или экстренные службы, переоснащают свою инфраструктуру, встраивая в нее от одного до нескольких ЦОВ. Для многих компаний, таких как телекоммуникационные, страховые, торговые, коммерческие банки и банки кредитно-карточной системы, гостиницы и т.д. ЦОВ являются незаменимым средством связи и эффективным менеджмента взаимоотношения с клиентами.

В последние несколько лет происходит интеграция телефонной и компьютерной индустрии, которая привела к появлению многофункциональных ЦОВ (так называемые контакт-центры), обладающих несравненно большим набором услуг и возможностями. Контакт-центры обеспечивают обработку вызовов (запросов), передаваемых как по телефонным, так и по пакетным сетям (Интернету, сетям IP-телефонии). Потребность в расширении спектра предоставляемых информационных услуг постоянно растет почти во всех сферах применения ЦОВ (рис. 1).

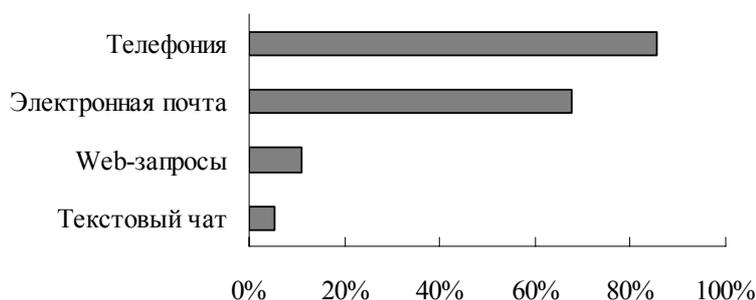


Рис. 1 – Каналы, по которым клиенты обращались в ЦОВ [2]

Введение новых услуг позволяет операторам контакт-центра работать более эффективно, что увеличивает общую производительность и добавляет гибкость в управлении ЦОВ. Происходит это путем равномерного распределения нагрузки в течение всего рабочего времени. В периоды низкой занятости от основной работы – обслуживания телефонных звонков, операторы ЦОВ могут обрабатывать трафик из пакетных сетей. Однако, это вносит и свои затруднения – ведь телефонные звонки имеют самый высокий приоритет и должны быть обслужены в течение нескольких секунд или минут. Иначе будет расти количество отказов от обслуживания, что приведет к потере клиентов. И наоборот, ответы на электронную почту, факс или текстовый чат могут быть отложены на некоторое время. Данная проблема известна в западной литературе как интеграция вызовов (call blending). С математической точки зрения возникает потребность в анализе новых моделей с целью их эффективного применения на практике.

Рассмотрим модель с двумя видами нагрузки – традиционной телефонной и нагрузкой из пакетных сетей в виде электронных писем, текстового чата и веб-обслуживания (низкоприоритетная нагрузка). Распределение интервалов между вызовами подчиняется закону распределения Пуассона. Интенсивность поступающей телефонной нагрузки λ . Промежутки времени обслуживания независимо и экспоненциально распределены, а интенсивности обслуживания телефонной нагрузки и нагрузки из пакетных сетей равны μ_1 и μ_2 соответственно. При высокой нагрузке на центр новые входящие вызовы помещаются в очередь. Имеются две неограниченные очереди как для телефонной, так и для низкоприоритетной нагрузки, дисциплина обслуживания которых FCFS (первым пришел – первым обслуживаешься). Общее количество операторов – N .

В таком операторском центре возникает потребность максимально увеличить количество обслуженных запросов из сетей с пакетной передачей данных, но так, чтобы качество обслуживания традиционных телефонных звонков при этом не страдало. В общем случае, качество обслуживания телефонных вызовов задается их временем ожидания в очереди (как правило берется величина 20 сек.). При этом необходимо найти оптимальное соотношение между операторами, которые занимаются только обслуживанием телефонных вызовов и операторами, которые обслуживают как телефонные, так и низкоприоритетные заявки.

Стоит отметить, что в модели не допускается приоритетное прерывание обслуживания в связи с появлением в системе телефонных заявок, иначе данная проблема была бы не актуальна – при отсутствии в системе телефонных вызовов все операторы были бы заняты обслуживанием заявок из пакетных сетей, а при поступлении нового звонка сразу же переключались на его обслуживание. Обслуженная нагрузка из пакетных сетей была бы равна: $\mu_1(N - \lambda_1/\mu_1)$.

Обслуживание низкоприоритетного трафика должно инициироваться только в тех случаях, когда в очереди нет телефонных заявок, иначе время ожидания будет расти. Математический аппарат приведен в [3, 4]. Обозначим за c количество операторов, которые занимаются обслуживанием двух видов нагрузки. Запишем выражения для основных характеристик качества обслуживания.

Вероятность ожидания телефонных заявок будет определяться выражением:

$$C_{(c)}(N) = \sum_{y=0}^c \sum_{x=N-y}^{\infty} q_{x,y} = \sum_{y=0}^c \sum_{i=y}^c \frac{K_{y,i-y} z_i^{i-y}}{1 - z_i},$$

где x – количество телефонных заявок на обслуживании и в очереди;
 y – количество низкоприоритетных заявок на обслуживании;

$q_{x,y} = \sum_{i=y}^c K_{y,i-y} z_i^{x-(N-i)}$ – стационарные вероятности перехода системы из одного состояния в другое;

$$z_i = \frac{((N-i)\mu_1 + \lambda + i\mu_2) - \sqrt{((N-i)\mu_1 + \lambda + i\mu_2)^2 - 4\lambda(N-i)\mu_1}}{2(N-i)\mu_1}, \text{ для } i = 0, \dots, c;$$

$$K_{y,i} = \frac{(y+1)\mu_2}{i(\mu_1 + \mu_2) - i\mu_1 z_{y+i}} K_{y+1,i-1}, \quad 0 \leq y < c, \quad 1 \leq i \leq c - y.$$

Время ожидания телефонных заявок будет определяться выражением:

$$EW_{(c)}^q = \frac{1}{\mu_1} \sum_{x,y} [(x+y) - N]^+ q_{x,y}, \quad \text{где } [x]^+ = \max\{x, 0\}.$$

И, наконец, обслуженная низкоприоритетная нагрузка:

$$\xi_{(c)} = \mu_2 \sum_{x,y} y q_{x,y}.$$

На рис. 2 показано поведение функций времени ожидания телефонных заявок в очереди и обслуженная низкоприоритетная нагрузка в единицу времени в зависимости от значения c . Расчеты проводились при $N = 10$ и $\lambda = 3$ для разнообразной низкоприоритетной нагрузки в соответствии с наиболее востребованными на сегодняшний день услугами: электронная почта (e-mail), веб-обслуживание (web) и текстовый чат в реальном масштабе времени (chat). При этом среднее время обслуживания различных низкоприоритетных заявок различается: для электронной почты оно равно около десяти минут ($\mu_2 = 0,1$), для веб-обслуживания около четырех минут ($\mu_2 = 0,25$) и для текстового чата около одной минуты ($\mu_2 = 1$). При этом на обслуживание традиционных телефонных звонков затрачивается в среднем около двух минут ($\mu_1 = 0,5$). Для сравнения приведены результаты моделирования ЦОВ без применения интеграции вызовов, т.е. когда операторы, обслуживающие низкоприоритетную нагрузку, не прибегают к обслуживанию телефонных вызовов. В этом случае значение c равно количеству операторов, обслуживающих только нагрузку из пакетных сетей. Тогда значения времени ожидания телефонных вызовов рассчитываются по хорошо известной формуле Эрланг-С [3], а величина обслуженной низкоприоритетной нагрузки получается путем перемножения количества операторов, её обслуживающих, на интенсивность обслуживания низкоприоритетных заявок.

Без применения интеграции вызовов при заданных параметрах мы имеем приемлемое время ожидания телефонных заявок в очереди (21,5 сек.), если на обслуживание низкоприоритетного трафика адресовать не более двух операторов. При этом обслуженная низкоприоритетная нагрузка равна 0,5 (для примера бралось значение для веб-обслуживания $\mu_2 = 0,25$). В случае с интеграцией вызовов при $c = 2$ время ожидания очень мало (3 – 4 сек.), а обслуженная НП нагрузка равна 0,029. При больших значениях c мы получаем выигрыш в обслуженной низкоприоритетной нагрузке при приемлемых значениях времени ожидания телефонных вызовов. Если мы возьмем время ожидания равным 21,5 сек., тогда обслуженная низкоприоритетная нагрузка в случае веб-обслуживания увеличится на 62,4% (значение 0,812). Аналогично, зафиксировав обслуженную низкоприоритетную нагрузку на уровне 0,5, мы можем максимально снизить время ожидания (до 9,5 сек.). В общем случае – задавая качество обслуживания (время ожидания не превышает некоторой величины) мы можем найти оптимальное соотношение между операторами, которое удовлетворяет заданным требованиям, но при этом обслуженный низкоприоритетный трафик будет максимально возможным.

Из рис. 2 видно, что среднее время ожидания заявок в телефонной очереди увеличивается медленно, в то время как обслуженная низкоприоритетная нагрузка изменяется практически по линейному закону. Это справедливо в тех случаях, когда интенсивность обслуживания низкоприоритетных заявок сравнима, или больше по величине интенсивности обслуживания телефонных заявок. Применительно к ЦОВ, в этом случае, можно сказать, что эффективность работы операторов будет возрастать по линейному закону, в то время как время ожидания телефонных заявок будет расти медленно.

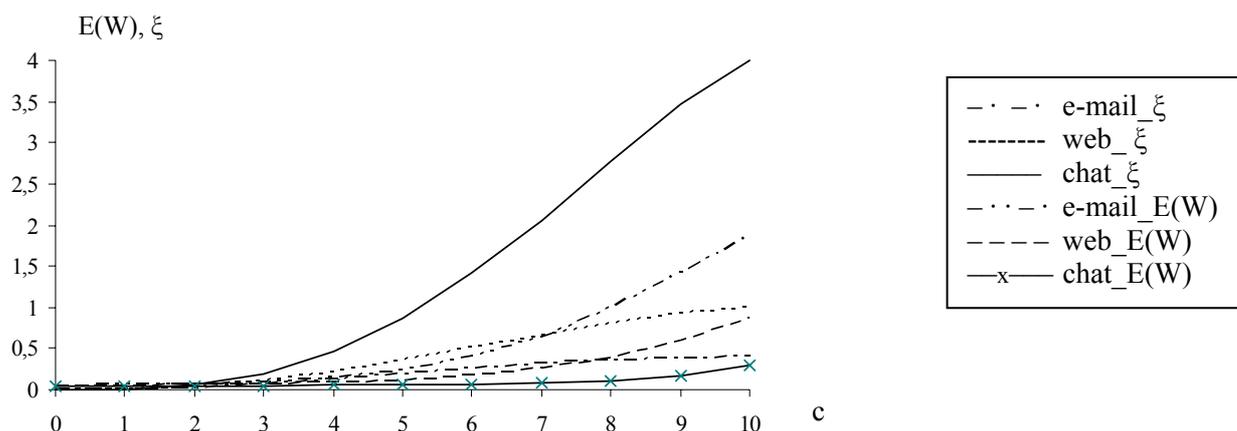


Рис. 2 – Время ожидания телефонных вызовов и обслуженная нагрузка из пакетных сетей в зависимости от параметра c

Литература

1. Росляков А.В., Самсонов М.Ю., Шибаета И.В. Центры обслуживания вызовов (Call centre). – М.: Эко-Трендз, 2002. – 272 с.
2. Солонин В. В Россию приходят передовые технологии IP-телефонии // CNews Analytics – 2004. [<http://cnews.ru/reviews/ip2003/part2/ip-pbx.shtml>].
3. Таха Х.А. Введение в исследование операций / Пер. с англ. Изд. 6-е – М.: Издательский дом «Вильямс», 2001. – 912 с.
4. Bhulai S., Koole G.M. A queueing model for call blending in call centers // In proceedings of the 39th IEEE CDC, – 1421-1426, 2000.

