

УДК 621.391

## ПОДХОДЫ К АНАЛИЗУ ПОКАЗАТЕЛЕЙ КАЧЕСТВА СИГНАЛЬНОГО ОБМЕНА В СЕТЯХ NGN/IMS

Н. А. Куликов, аспирант СПбГУТ им. М. А. Бонч-Бруевича; kulikov@protei.ru

Качество обслуживания (QoS) является одной из основных характеристик любой сети связи, включая сеть, базирующуюся на архитектуре IMS. В процессе обработки различных сигнальных сообщений в узлах IMS-сети возникают задержки. При этом размер задержки варьируется в зависимости от типа запрашиваемой услуги. В статье предлагается комплексный подход к анализу величины задержки при рассмотрении различных случаев.

**Ключевые слова:** IMS, NGN, SIP, ТфОП, PSTN, качество обслуживания, СМО, кумулянтный анализ.

**Архитектура IMS.** IMS — это открытая сетевая архитектура, поддерживающая широкий спектр услуг на основе IP для сетей с коммутацией как каналов (CS), так и пакетов (PS). Управлять предоставлением услуг пользователям и одновременно модифицировать медиа-поток в рамках сессии позволяет протокол инициализации сессий SIP [1], являющийся основным сигнальным протоколом IMS. Кроме того, SIP — универсальный протокол для различных типов абонентского доступа.



Рис. 1. Упрощенная схема ядра IMS

Основой данной архитектуры является IMS-ядро, состоящее из набора специализированных модулей, отвечающих за различные функции при обслуживании абонентов [2]. Основные функциональные элементы ядра IMS-сети (рис. 1), занятые в процессе регистрации, установления соединения, маршрутизации пакетов RTP, тарификации вызовов к/от абонентов — это набор функций управления сессиями CSCF (Call Session Control Functions). Он разделяется между тремя типами модулей: проксирующим (P-CSCF), сессионным (S-CSCF) и модулем интерконнек-

ций (I-CSCF). Пользовательская база данных HSS (Home Subscriber Server) представляет собой централизованное хранилище информации о пользователях и услугах сети IMS.

**Качество обслуживания в IMS.** Одним из основных требований к IMS является поддержка качества обслуживания — QoS [3]. Анализ качества предоставления услуг предусматривает анализ качества «из конца в конец». Кроме того, должна учитываться функциональная и физическая структура IP-сети. В то время как требования к параметрам передачи IP-пакетов (задержка, джиттер, потери) проработаны достаточно хорошо, четких требований к показателям QoS в части задержек на предоставление пользовательских услуг для сетей IMS не разработано.

В международной рекомендации Y.1531 [4], посвященной вопросу оценки производительности IP-сетей и разработанной Международным Союзом Электросвязи, в качестве основной метрики рассматривается задержка предоставления основной услуги. В формулировках данного документа Call Setup Delay — это промежуток времени от начала установления соединения до получения подтверждения об успешном установлении сеанса. При этом не учитываются задержки, внесенные окончательным оборудованием (User Equipment — UE), а так же длительность передачи промежуточных сигнальных сообщений.

Более поздняя рекомендация Инженерного совета Интернета (IETF) [5] вводит большее число показателей, выделяя задержки, связанные с регистрацией абонента (Registration Request Delay — RRD), установлением голосового соединения (Session Request Delay — SRD) и его разрушением (Session Disconnect Delay — SDD). Однако указанные рекомендации не вводят численных значений для предлагаемых метрик.

Нормируемые значения определяются для гибридных сетей [6, 7] в случаях, когда устанавливаемое соединение проходит не только через сеть ТфОП/ISDN, но и через участок, построенный по IP-технологии. При этом допустимые значения нормируемых показателей менее жесткие, чем установленные [8] для однородной цифровой сети интегрального обслуживания.

При переходе от ISDN-сети к архитектуре IMS, целиком построенной на технологии IP, характеристики предоставления услуг абонентам не должны ухудшаться как по среднему значению, так и по 95% квантилю. В связи с этим, целесообразно применить численные требования, приведенные в [8], к метрикам, определенным рекомендациями [4, 5]. Метрики должны оцениваться с учетом особенностей сигнального обмена в сетях IP-телефонии на основании принципов, установленных в [4]. Таким образом, для IMS-сети могут быть введены следующие характеристики QoS:

- Post-selection delay (PSD) — задержка на установление соединения, определяемая как промежуток времени между отправкой абонентом А начального сообщения INVITE до получения им сообщения о статусе абонента Б

(180 Ringing — в случае, если вызываемый абонент свободен; 486 Busy Here — если занят).

- Answer Signal Delay (ASD) — задержка сигнала об ответе абонента. Промежуток времени между отправкой ответного сообщения 200 OK вызываемым абонентом до получения им подтверждающего сообщения ACK.

- Call Release Delay (CRD) — задержка сигнала об отбое абонента между отправкой сообщения BYE до получения сообщения ACK.

Соответствующие метрики должны быть вычислены без учета длительности обработки сигнальных сообщений в оконечном оборудовании только на основании длительности передачи сообщений по сети и задержек на обработку в узлах IMS-ядра.

Таблица 1. Основные показатели QoS

Параметр	Среднее значение, с	Квантиль 95%, с
Post-selection delay	3,0	6,0
Answer signal delay	0,75	1,5
Call release delay	0,4	0,6

**Модель оценки времени предоставления услуги.** Пусть случайная величина длительности установления соединения в сети IMS обозначается  $T_{SRD}$ . В процедуре установления соединения она определяется, начиная с момента передачи вызывающим UE сообщения INVITE до момента принятия им уведомления 180 Ringing без учета времени обработки сообщений в самих UE [4].

Случайная величина  $T_{SRD}$  будет вычисляться как совокупность времени пребывания различных сигнальных сообщений в каждом из узлов сети ( $T_{P-CSCF}$ ,  $T_{I-CSCF}$ ,  $T_{S-CSCF}$ ,  $T_{HSS}$ ). Задержку на передачу сигнальных сообщений между самими модулями IMS-ядра будем считать пренебрежимо малой величиной, поскольку ядро IMS, как правило, располагается в рамках одного помещения и объединено высокоскоростной IP-сетью.

Само ядро IMS-сети можно рассматривать как многофазную систему массового обслуживания (СМО), состоящую из цепочки взаимонезависимых СМО, на вход которой поступают сигнальные сообщения с определенной интенсивностью  $\lambda$ . Общая величина параметров QoS будет определяться исходя из совокупной задержки заявок в каждой из фаз системы.

Например, величина  $T_{SRD}$  для успешного соединения будет иметь вид:

$$T_{SRD} = 2T_{INVITE}^{(P-CSCF)} + 2T_{INVITE}^{(S-CSCF)} + T_{INVITE}^{(I-CSCF)} + T_{INVITE}^{(HSS)} + 2T_{Ringing}^{(P-CSCF)} + 2T_{Ringing}^{(S-CSCF)} + T_{Ringing}^{(I-CSCF)}. \quad (1)$$

Процесс обработки сигнальных сообщений в ядре IMS-сети может быть проанализирован на основе сбора эмпирических данных. Обычно сведения о длительности пребывания заявки в системе формируются в виде гистограммы, получаемой в результате измерений. Если взять шаг гисто-

граммы  $\tau$ , то преобразование Лапласа-Стилтьеса функции распределения длительности пребывания заявок в системе  $S^*(s)$  будет определяться как

$$S^*(s) = e^{-z\tau} \sum_{i=0}^N p_i e^{-i\tau s}, \quad (2)$$

где  $z\tau$  — начало первого приращения функции  $S(t)$  относительно  $t=0$ ;  $p_i$  — приращение в точке  $i\tau$ ;  $N$  — номер последнего приращения.

Для анализа показателей QoS удобно прибегнуть к механизмам кумулянтного анализа [9, 10]. Для нахождения кумулянтов требуется определить моменты каждого распределения, которые выражаются через производную от выражения (2) соответствующего порядка при  $s \rightarrow 0$ .

Поскольку процесс предоставления услуги в IMS-сети связан с обработкой сигнальных сообщений на различных стадиях, то аппроксимирующую функцию длительности задержки сообщения для всего IMS-ядра можно выразить через суммарные кумулянты  $k$ -го порядка [10]:

$$\chi_k^\eta = \sum_{i=1}^N \chi_k^{\xi_i}, \quad (3)$$

где  $\eta = \xi_1 + \xi_2 + \dots + \xi_N$  — суммирование кумулянтов по всем интересующим фазам СМО. Этот метод применим, как при вычислении параметров многофазной системы, состоящей из модулей ядра IMS-сети, так и при определении параметров QoS, получаемых из последовательного прохождения нескольких типов сигнальных SIP-сообщений через ядро IMS.

**Аппроксимация эмпирических данных.** Нахождение математического ожидания нормируемых показателей QoS путем прямого суммирования первых кумулянтов каждого распределения не составляет существенной сложности. Нахождение квантиля распределения множества случайных величин сопряжено с трудностями. В связи с этим прибегнем к аппроксимации собранных эмпирических данных.

Проанализируем возможность использования методов аппроксимации для анализа многофазной СМО, состоящей из нескольких взаимонезависимых простейших СМО. По классификации Кендалла-Башарина соответствующие модели обозначаются как  $M/M/1/\infty/f_0^0$ . Символ  $M$  в первой и второй позиции указывает на экспоненциальное распределение входящего потока  $A(t)$  и длительности обслуживания  $B(t)$ , цифра 1 говорит об анализе однолинейной СМО. Размер буфера полагается бесконечным, а запись  $f_0^0$  означает, что в системе используется дисциплина обслуживания FIFO.

Будем считать, что известны интенсивности входящего потока заявок  $\lambda$  и интенсивности обслуживания  $\mu_k$  в каждой СМО, при этом все  $\mu_k = 1$ . Характер функции  $S(m, t)$  зависит от нагрузки на систему и количества стадий на маршруте. На рис. 3 приведены графики плотности распределения вероятности для различного числа фаз при интенсивности нагрузки  $\rho = 0,5$ .

Из характера кривых видно, что при росте числа фаз огибающая ПРВ меняет свой характер: она все более при-

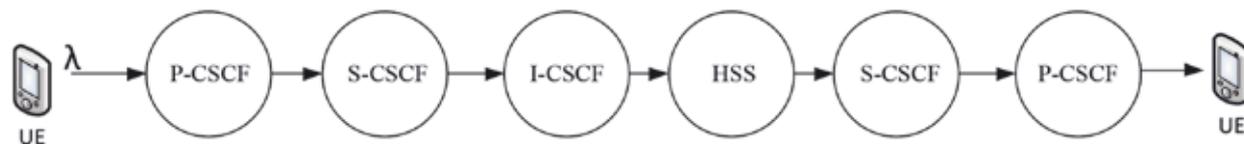


Рис. 2. Модель ядра IMS в виде многофазной системы

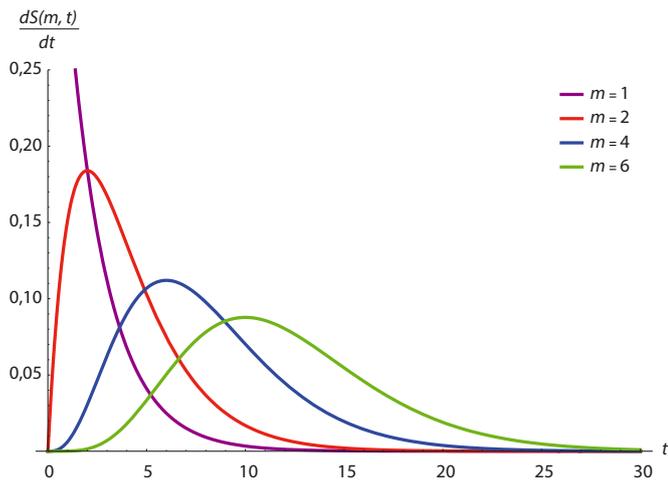


Рис. 3. Производные функции  $S(m, t)$  для многофазной СМО

ближается к плотности Гауссова закона распределения случайной величины.

Одним из распространенных способов аппроксимации является ряд Эджворта [9], позволяющий представить исследуемую функцию  $S(t)$  через нормальное распределение  $\Phi(t)$  и его производные  $\Phi^{(k)}(t)$ . Однако, если реальная функция значительно отличается от нормальной, то подобная аппроксимация становится непригодной, поскольку:

- сумма членов ряда Эджворта может привести к отрицательным значениям частот особенно на «хвостах» распределений;
- с увеличением числа суммируемых членов ряда Эджворта приближение может ухудшаться [11].

При малом числе фаз функция  $S(m, t)$  значительно отличается от нормальной, поэтому прибегнем к оценке квантиля методом Корниша-Фишера [12], который так же базируется на принципах кумулянтного анализа, но не прибегает к построению функции распределения вероятности.

**Нахождение квантиля для цепочки СМО типа М/М/1.** Предлагаемая техника использует первые моменты распределения и позволяет дать достаточно точную оценку квантиля распределений, существенно отличающихся от нормального. Значение аппроксимированного  $\alpha$ -квантиля определяется следующим образом:

$$x_\alpha = M + \frac{(\sigma^2 + 1)z_\alpha}{2} + \frac{(z_\alpha^2 - 1)\kappa_3}{6} + \frac{(z_\alpha^3 - 3z_\alpha)\kappa_4}{24} - \frac{(2z_\alpha^3 - 5z_\alpha)\kappa_3^2}{36}, \quad (4)$$

где  $z_\alpha$  —  $\alpha$ -квантиль нормального распределения;  $\kappa_3$  и  $\kappa_4$  — третий и четвертый кумулянт распределения;  $\sigma$  — среднеквадратичное отклонение;  $M$  — математическое ожидание.

Для упрощения вычислений удобно произвести нормализацию выражения (4), переписав его через нормализованный квантиль:

$$x'_\alpha = (x_\alpha - M) / \sigma. \quad (5)$$

После нормализации математическое ожидание становится равным нулю, а среднеквадратическое отклонение — единице. В результате выражение (4) будет иметь вид

$$x'_\alpha = z_\alpha + \frac{(z_\alpha^2 - 1)\beta_1}{6} + \frac{(z_\alpha^3 - 3z_\alpha)\beta_2}{24} - \frac{(2z_\alpha^3 - 5z_\alpha)\beta_1^2}{36}; \quad (6)$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}; \quad \beta_2 = \frac{\mu_4^2}{\mu_2^4}, \quad (7)$$

где  $\beta_1$  — коэффициент асимметрии;  $\beta_2$  — коэффициент эксцесса. Затем получаемое значение должно быть трансформировано обратно к значению квантиля исходного ненормализованного распределения.

В основе аппроксимации Корниша-Фишера лежит следующий принцип: квантиль аппроксимированного распределения можно рассматривать как аппроксимацию квантиля реального распределения при условии, что множество моментов реального и аппроксимированного распределения согласованы. Квантиль приближенного распределения выражается в виде асимптотического ряда, являющегося функцией соответствующего квантиля распределения Гаусса. Слагаемые ряда — полиномиальные функции соответствующего квантиля нормального распределения, а коэффициенты — кумулянты анализируемого распределения.

Функция распределения длительности пребывания заявок в системе, состоящей из набора СМО типа М/М/1, имеет вид [13]:

$$S(m, t) = 1 - e^{-(1-\rho)\mu t} \sum_{k=0}^{m-1} \frac{[(1-\rho)\mu t]^{m-k-1}}{(m-k-1)!}. \quad (8)$$

Сравним величины 95% квантиля, вычисленные по данной формуле и с помощью метода аппроксимации при коэффициенте загрузки системы  $\rho = 0,5$ .

Таблица 2. Ошибка аппроксимации Корниша-Фишера многофазной системы, состоящей из  $m$  СМО М/М/1

$m$ — число фаз	Квантиль, определенный по формуле (8)	Аппроксимированный квантиль	Ошибка аппроксимации, %
1	5,99146	6,03431	0,71
2	9,48773	9,51189	0,25
3	12,5916	12,6084	0,13
4	15,5073	15,5202	0,08
5	18,307	18,3175	0,06
6	21,0261	21,0349	0,04
7	23,6848	23,6925	0,03
8	26,2962	26,303	0,03
9	28,8693	28,8753	0,02
10	31,4104	31,4159	0,02
15	43,773	43,7767	0,01
20	55,7585	55,7611	0,00

Ошибка при аппроксимации Корниша-Фишера не превышает 1% даже для одной системы М/М/1, функция распределения длительности пребывания в которой существенно отличается от нормального распределения. При увеличении числа фаз ошибка аппроксимации ожидается уменьшается. Этот результат важен, поскольку элементы, входящие в состав ядра IMS-сети (CSCF, AS и т.д.), представляют собой аппаратно-программные комплексы, построенные на базе центрального процессора (Central Processing Unit — CPU).

Полагая входящий поток сигнальных сообщений простейшим [13], большинство модулей можно рассматривать

как СМО типа М/М/1. Однако такой подход сложно применить к подсистеме HSS, представляющей собой абонентскую базу данных. Поскольку она взаимодействует не только с CPU, но и с накопителями данных, а так же выполняет различные операции, согласно установленному расписанию, HSS необходимо рассматривать как СМО типа М/Г/1 [14, 15].

**Нахождение квантиля для СМО типа М/Д/1.** Для оценки возможности применения аппроксимации Корниша-Фишера для систем с произвольным процессом обслуживания рассмотрим СМО с экспоненциальным входящим потоком и постоянным временем обслуживания. Такая модель характерна тем, что для распределений  $B(t)$  с коэффициентом вариации меньше единицы, системы М/Д/1 и М/М/1 позволяют вычислить нижнюю и верхнюю границы для интересующих параметров [16]. По классификации Кендалла-Башарина модель обозначается как М/Д/1/∞/f<sub>0</sub><sup>0</sup>. Символ D во второй позиции указывает на постоянный (детерминированный) характер длительности обслуживания  $B(t)$ . СМО такого типа входят в ряд классических моделей, активно исследуемых с начала прошлого века.

Функция распределения длительности пребывания заявки в СМО типа М/Д/1 может быть записана следующим образом [17]:

$$W(t) = (1 - \rho) \sum_{k=0}^{\lfloor \mu t \rfloor} \frac{\left[ \lambda \left( \frac{k}{\mu} - t \right) \right]^k}{k!} e^{-\lambda \left( \frac{k}{\mu} - t \right)}, \quad (9)$$

где символ  $\lfloor \cdot \rfloor$  означает, что берется целое значение от результата деления.

Рассмотрим одну систему типа М/Д/1 при различной интенсивности входящей нагрузки  $\lambda$ , полагая интенсивность обслуживания  $\mu = 1$ . Сравним величины 95% квантиля, вычисленные по формуле (9) и с помощью метода аппроксимации Корниша-Фишера.

**Таблица 3. Ошибка аппроксимации Корниша-Фишера СМО М/Д/1 при различной нагрузке системы**

$\rho$ — нагрузка системы	Квантиль, определенный по формуле	Аппроксимированный квантиль	Ошибка аппроксимации, %
0,1	1,54	1,47	4,3
0,2	1,86	1,79	3,8
0,3	2,07	2,12	2,43
0,4	2,54	2,52	0,46
0,5	3,04	3,06	0,42
0,6	3,81	3,83	0,40
0,7	5,07	5,12	0,81
0,8	7,57	7,66	1,07
0,9	14,89	15,09	1,32

Ошибка расчета квантиля для одной системы М/Д/1 не превышает 5%. При малых и больших значениях  $\rho$ , функция распределения вероятностей длительности ожидания для системы М/Д/1 существенно отличается от нормального распределения, приводя к увеличению ошибки аппроксимации.

**Заключение.** Проведенный анализ показывает, что аппроксимация методом Корниша-Фишера дает приемлемые результаты для различных типов СМО. Причем анализ многофазных систем показывает, что аппроксимация улучшается с ростом числа фаз. Возможность применения разложения Корниша-Фишера для определения квантиля в системах М/М/1 и М/Д/1 позволяет использовать данный подход для исследования эмпирических функций, полученных при исследовании IMS-сети.

На практике не всегда есть доступ к анализу взаимодействия между узлами, входящими в состав IMS-ядра. Распределение длительности обслуживания  $B(t)$  в отдельных модулях, представляющих собой закрытые аппаратно-программные комплексы, не всегда можно определить на практике. Решая практические задачи анализа QoS, можно рассматривать ядро IMS как единую систему и анализировать распределение длительности пребывания заявок  $S(t)$  в ней. Суммируя кумулянты распределений длительности пребывания соответствующих сообщений в сети IMS, можно определить результирующие показатели качества предоставления услуг. Проведение такого анализа позволяет установить нижний предел совокупной интенсивности обслуживания многофазной системы, позволяющий гарантировать соблюдение заданных параметров QoS.

#### ЛИТЕРАТУРА

1. **Rosenberg J. et al.** SIP: Session Initiation Protocol. RFC 3261, IETF.— June 2002.
2. **3GPP.** Network architecture. TS 23.002 (V12.4.0).— March 2014.
3. **3GPP.** Policy and charging control architecture. TS 23.203 (V12.4.0).— March 2014.
4. **ITU-T Recommendation Y.1531,** 2007. SIP-based Call Processing Performance.
5. **IETF Request for Comments: 6076,** 2011. Basic Telephony SIP End-to-End Performance Metrics
6. **ITU-T Recommendation Y.671,** 2000. Post-selection delay in PSTN/ISDN networks using Internet telephony for a portion of the connection.
7. **ITU-T Recommendation Y.1530,** 2007. Call processing performance for voice service in hybrid IP networks.
8. **ITU-T Recommendation E.721,** 1999. Network grade of service parameters and target values for circuit-switched services in the evolving ISDN.
9. **Крамер Г.** Математические методы статистики.— М.: Мир, 1975.
10. **Малахов А.Н.** Кумулянтный анализ случайных негауссовских процессов и их преобразование.— М.: Сов. радио, 1978.
11. **Кендалл М., Стьюарт А.** Теория распределений.— М.: Наука, 1966.
12. **Cornish E. A., Fisher, R. A.** Moments and cumulants in the specification of distributions / *Revue de l'Institut International de Statistique*.— 1937.— № 5.— P. 307–320.
13. **Fischer M. J., Fowler T. B.** Fractals, Heavy-Tails, and the Internet / *Mitretek Technology Summaries*, Sigma.— 2001.
14. **Kihl M. et al.** Performance modeling of databases in a Telecommunication Service Management system / *IARIA 7th International Conference on Digital Telecommunications (ICDT 2012)*.— 2012.
15. **Arzuaga E., Kaeli D. R.** An M/G/1 Queue Model for Multiple Applications on Storage Area Networks. <http://www.ece.neu.edu/groups/nucar/publications/CAECW08arzuaga.pdf>
16. **Соколов А.Н., Соколов Н.А.** Однолинейные системы массового обслуживания.— СПб.: Изд-во «Теледом» ГОУВПО СПбГУТ, 2010.
17. **Franx G. J.** A Simple Solution for the M/D/c Waiting Time Distribution. Universiteit van Amsterdam.— 16 November 1998.

Получено 10.11.14