

Пример. На вход системы поступает произвольный (рекуррентный) поток вызовов с интенсивностью $\Lambda = 400$ Эрл. $S = 10$ для емкости системы $m = 410, 450$ и 500 линий (для исключения бесконечной очереди обязательно должно соблюдаться условие $m > \Lambda$). Длительность обслуживания постоянна и равна t .

При записи в обозначениях Кендалла эта модель имеет вид

$$GI/D/m/r = \infty.$$

Основные характеристики QoS:

- вероятность ожидания P_{so} ;
- средняя длительность ожидания t_w для задержанных вызовов;
- средняя длительность ожидания для любого вызова W ;
- средняя длину очереди Q .

Для каждой дисциплины обслуживания требуемые характеристики QoS могут быть различными и зависят они как от структурных параметров системы, так и от вида поступающего в нее потока требований на обслуживание. В случае простейшего потока вызовов, поступающего в полностью доступную систему с ожиданием, для расчета характеристик QoS при экспоненциальном времени обслуживания используется С-формула Эрланга, а при постоянном – метод Кроммелина.

Но, как показывают измерения параметров нагрузки, в мобильных сетях реальные потоки вызовов существенно отличаются от математической модели простейшего (пуассоновского) потока. Обобщением простейшего потока является модель потока Пальма, где распределение промежутков времени между вызовами одинаковое, но не строго экспоненциальное. При использовании такой модели произвольного потока параметры реальных потоков вызовов в большей степени соответствуют выбранной модели, однако точные формулы расчета систем с ожиданием для этого случая неизвестны.

Степень отличия реального потока от простейшего может быть определена через пик-фактор интенсивности нагрузки S , рассчитываемый по формуле

$$S = \sigma^2 / \Lambda, \quad (1)$$

где σ^2, Λ – соответственно дисперсия интенсивности нагрузки и ее математическое ожидание.

Для простейшего потока $S = 1$. В действительности на современных телекоммуникационных сетях $S = 2, 10$, что подтверждено статистическими данными практических измерений параметров нагрузки [3; 4]. Исследованиями установлено, что наибольшая степень согласия между реальными потоками вызовов и теоретическими законами распределения наблюдается при аппроксимации первых рекуррентным потоком Пальма с гиперэкспоненциальным распределением промежутков времени между вызовами (достаточно двух экспонент). При этом распределение вероятностей P_i случайной величины i описывается нормальным законом распределения:

$$P_i = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(i-\Lambda)^2}{2\sigma^2}}, \quad (2)$$

где i – количество вызовов в условную единицу времени (математическое ожидание случайной величины i и есть интенсивность нагрузки Λ).

Перечисленные характеристики QoS необходимо рассчитывать в указанной последовательности, поскольку последние из них однозначно определяются двумя первыми:

$$W = t_w \cdot P_{so}, \quad (3)$$

$$Q = \Lambda W. \quad (4)$$

Выражение (3) следует из того, что средняя длительность ожидания W для любого вызова (ожидающего и не ожидающего) является средним значением времени ожидания, отнесенным ко всем вызовам. Если известна средняя длительность ожидания t_w только задержанных

вызовов, то для нахождения W надо умножить значение t_w на вероятность $P_{>0}$, показывающую среднюю долю задержанных вызовов.

Выражение (4) полностью соответствует формуле Литтла: за W единиц времени ожидания в очередь поступит $\Lambda \cdot W$ вызовов.

Значительно труднее рассчитать $P_{>0}$ и t_w . Эти характеристики могут быть определены из функции распределения состояний системы P_j и распределения времени ожидания $P(t_w)$ начала обслуживания. Однако общего метода получения таких функций не существует, и выражения для них обычно не являются простыми.

С целью получения аналитического выражения для расчета t_w используем следующие известные результаты [5]:

- из С-формулы Эрланга следует, что в системе $M/M/m/r = \infty$ средняя длительность ожидания для задержанных вызовов при экспоненциальной длительности обслуживания $t_{экл} = 1/(m - \Lambda)$;

- из формулы Поллачека – Хинчина следует, что в системе $M/D/1/r = \infty$ средняя длительность ожидания для задержанных вызовов при постоянной длительности обслуживания $t_{плч} = t_{экл} / 2$.

Очевидно, что в искомом выражении для расчета t_w системы $GI/D/m/r = \infty$ должны учитываться также результаты:

- первый – потому что пуассоновский поток M есть частный случай рекуррентного GI ;
- второй – поскольку однолинейная система ($m=1$) является частным случаем многолинейной.

Для системы $GI/D/m/r = \infty$ показано [6], что t_w больше $t_{экл}$ в $S/2$ раз при емкости системы, близкой к $m = \Lambda$. Это хорошо согласуется с приведенными известными соотношениями: учтены как отличие рекуррентного потока от пуассоновского через пик-фактор S , так и то, что средняя длительность ожидания при постоянной длительности обслуживания вдвое меньше, чем при экспоненциальной, но отнесено это к характерной точке $m = \Lambda$.

Однако с ростом емкости m системы коэффициент $k=2$ убывает со скоростью $k(m) \approx (m + \Lambda)/m$. По результатам имитационного моделирования системы $GI/D/m/r = \infty$ замечено, что точность расчета t_w повышается при замене этой зависимости выражением

$$k(m) \approx (m + \Lambda + 1 + \Lambda/m) / m \quad (5)$$

... Окончательно формула для расчета t_w системы $GI/D/m/r = \infty$ принимает вид

$$t_w \approx S \frac{1}{m - \Lambda} \cdot \frac{m}{m + \Lambda + 1 + \Lambda/m} = \frac{S}{(m + 1) [1 - (\Lambda/m)^2]} \quad (6)$$

Для расчета $P_{>0}$ применим такие доводы: вероятность ожидания $P_{>0}$ равна вероятности того, что вновь приходящий вызов застает все m линий занятыми:

$$P_{>0} = \sum_{j=m}^{\infty} P_j = 1 - \sum_{j=0}^{m-1} P_j, \quad (7)$$

где j – состояние системы (для линий $0 \leq j < m$, для очереди $m < j \leq \infty$).

В условиях неограниченного количества линий ($m = \infty$) вызовы обслуживаются без потерь. При постоянной длительности обслуживания t , когда нет потерь, свойства потока освобождений совпадают со свойствами потока поступления вызовов, так как происходит только сдвиг по времени на величину t между моментом поступления вызова и моментом окончания его обслуживания. При этом состояния системы обслуживания полностью определяются свойствами потока вызовов, а функции распределения количества вызовов в системе P_j и количества поступающих вызовов P , за время t полностью совпадают.

Например, подставив в (7) значения, получаемые из (2), при $m=5$ получаем $P_{11}=0$

При конечном числе m и неограниченном количестве мест ожидания ($r=\infty$) вызовы также обслуживаются без явных потерь. В этом случае вызовы, поступающие после занятия всех линий системы, попадают в очередь на ожидание и в случае освобождения хотя бы одной из m занятых линий тут же подаются из очереди на обслуживание. Теперь на линии системы поступают вызовы из первичного потока с интенсивностью Λ и из очереди с интенсивностью $\Lambda \cdot P_{11} \cdot t_m$, поскольку ожидающие в очереди вызовы образуют дополнительный поток с интенсивностью $\Lambda \cdot P_{11}$ и каждый ожидающий вызов ожидает в среднем в течение времени t_m . Поэтому общая интенсивность нагрузки на линии увеличится до $\Lambda_2 = \Lambda + Q$, так как согласно (3) $t_m \cdot P_{11}$ и есть средняя длительность ожидания W , а согласно (4) $\Lambda \cdot W$ и есть средняя длина очереди Q .

В этих условиях функция распределения количества вызовов в системе (на обслуживании и в очереди) или состояний системы P_j отличается от функции распределения количества поступивших вызовов P_j .

Результаты имитационного моделирования представлены на рис. 2 (кривые 1, 2, 3) в виде распределения состояний системы при поступающем в нее произвольном потоке вызовов с заданными параметрами.

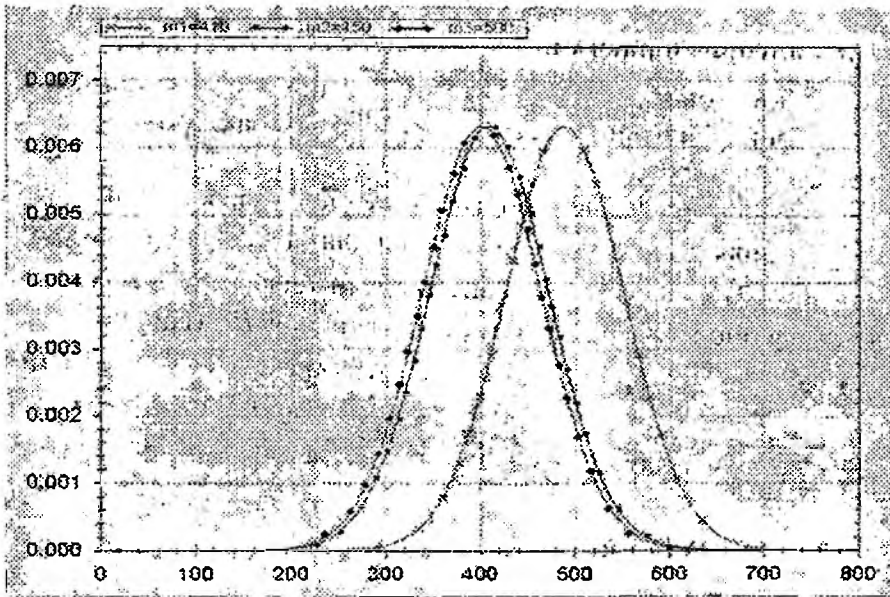


Рис 2

Очевидно, что уже при $m = 500$ (кривая 3) функция распределения достаточно симметрична, что позволяет целиком аппроксимировать ее нормальным законом распределения. Приведенный график показывает, что при уменьшении емкости m системы разброс отдельных значений функции распределения состояний системы от среднего или математического ожидания увеличивается. Отсюда следует, что дополнительный поток вызовов из очереди не только увеличивает общую интенсивность нагрузки Λ_2 , но и ее дисперсию σ . Для данного примера выполнена аппроксимация состояний системы нормальным законом распределения с параметрами $\Lambda_2 = \Lambda + Q$ и $\sigma_2 = \sigma + Q/2$ на участке, обусловленном пределами суммирования в выражении (7), т. е. от 0 до $m-1$.

График, представленный на рис. 3 (кривые 1, 2, 3), подтверждает приемлемую точность аппроксимации функции распределения состояний системы P_j нормальным законом распределения (сплошная линия) с параметрами:

$$\Lambda_2 = \Lambda + Q; \quad (8)$$

$$\sigma_2 = \sigma + Q/2. \quad (9)$$

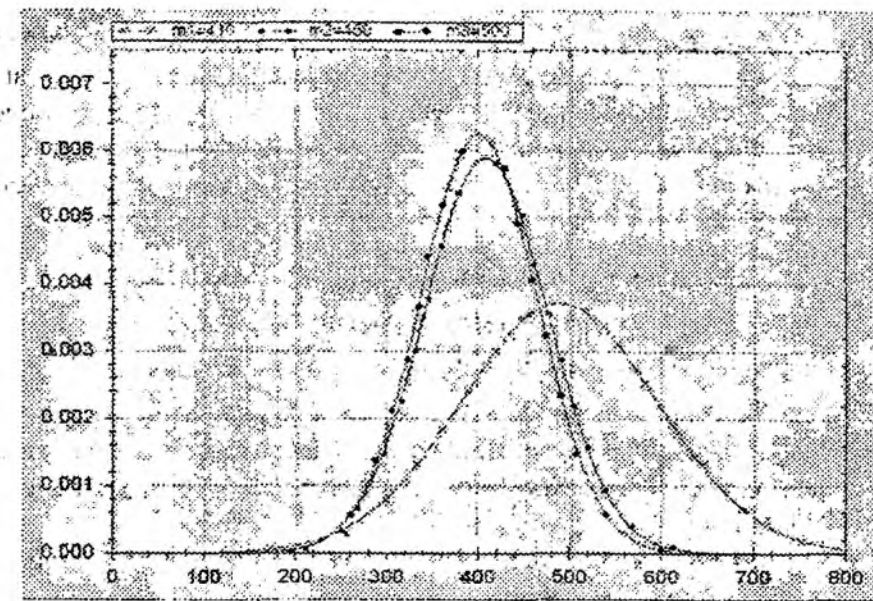


Рис 3

Итерационный алгоритм расчета

Из приведенных доводов и результатов имитационного моделирования (рис 2, 3) вытекает простой итерационный алгоритм расчета основных характеристик качества обслуживания системы $GI/D/m/r = \infty$:

- согласно (6) по заданным характеристикам потока вызовов Λ и S при емкости m системы определяется средняя длительность ожидания для задержанных вызовов t_w ,
- в соответствии с (7) и (2) по характеристикам интенсивности нагрузки Λ и ее дисперсии σ^2 определяется предполагаемая вероятность ожидания $P_{\text{до}}$ (как бы для случая, когда вызовы из очереди не возвращаются в систему и не увеличивают нагрузку на нее),
- с использованием найденных значений t_w и $P_{\text{до}}$ по формулам (3) и (4) определяются предполагаемые значения средней длительности ожидания W для всех вызовов и средней длины очереди Q ,
- с учетом вычисленных по (8) и (9) значений интенсивности нагрузки Λ_2 и среднеквадратичного отклонения σ_2 , в соответствии с (2) и (7) определяется уточненная вероятность ожидания P_w (с учетом влияния дополнительной нагрузки на линии системы из очереди в этом случае длина очереди более реальна, поскольку вызовы, не уходящие из системы немедленно, приводят к ее росту),
- с использованием уточненного значения вероятности ожидания P_w , согласно (3) и (4) уточняются значения средней длительности ожидания для всех вызовов W и средней длины очереди Q .

Для приведенного на рис 2 и 3 примера имитационного моделирования согласно предложенному алгоритму выполнен расчет, результаты которого сведены в таблицу, из которой видно, что относительная ошибка расчета Q и W определяется суммарной точностью расчета t_w и $P_{\text{до}}$ и, вместе с тем, ошибка расчета всех характеристик QoS остается в пределах $\pm 10\%$ (вторая и третья итерации). Расчеты, выполненные для данного примера методом Кроммелина, дают занижение результата в пределах 50 – 95 % (с ростом m увеличивается ошибка)

Выводы

1. Предложена методика с анализом основных характеристик качества обслуживания мобильных сетей 3G на платформе IMS с возможностью развития мультимедийных услуг новых поколений (Quad Play).

Параметры QoS	Λ и σ	Λ_2 и σ	Λ_2 и σ_{2-10}
		$M = 410$	
$P_{0,1}$	0.899	0.894	0.958
t_w	0.505	0.181	0.181
W	0.454	0.162	0.174
Q	181.482	66.196	66.377
		$M = 450$	
$P_{0,1}$	0.917	0.914	0.931
t_w	0.106	0.087	0.087
W	0.097	0.08	0.08
Q	38.738	31.138	31.138
		$M = 500$	
$P_{0,1}$	0.963	0.96	0.977
t_w	0.055	0.062	0.062
W	0.053	0.06	0.059
Q	21.364	24.504	24.078

Примечание: Величины W и t_w выражены в единицах средней длительности обслуживания.

2. В предлагаемом примере с применением имитационного моделирования и разработанным итерационным алгоритмом расчета основных характеристик качества обслуживания системы IMS установлено, что реализация алгоритма в большом диапазоне варьирования параметров качества дает всегда несколько заниженную оценку вероятности ожидания услуг, однако при этом относительная ошибка не превышает 10% от общего времени ожидания.

По результатам расчета построены графики (рис. 2, 3) распределения состояний системы $GI/D/m/r = \infty$. Очевидно, что уже при $m = 500$ (рис. 2, кривая 3) функция распределения достаточно симметрична, что позволяет целиком аппроксимировать ее нормальным законом распределения.

Приведенные графики показывают, что при уменьшении емкости m системы разброс отдельных значений функции распределения состояний системы от среднего или математического ожидания увеличивается. Отсюда следует, что дополнительный поток вызовов из очереди не только увеличивает общую интенсивность нагрузки Λ_2 , но и ее дисперсию σ .

Построены графики, которые представляют собой аппроксимацию функций распределения состояний системы (рис. 3), выполненную нормальным законом распределения с параметрами $\Lambda_2 = \Lambda + Q$ и $\sigma_2 = \sigma + Q/2$ на участке, обусловленном пределами суммирования от 0 до $m-1$. Графики, подтверждают приемлемую точность аппроксимации функции распределения состояний системы P_j нормальным законом распределения с заданными параметрами.

Список литературы: 1. Гольдштейн И.Б. Softswitch. Санкт-Петербург: БХВ, 2006. 2. Костин А.А., Шустров А.К. SLA – еще одно средство в конкурентной борьбе // Вестник связи. 2001. №8. С. 54-66. 3. Ложковский А.Г., Захарченко Н.В., Горохов С.М. Экспериментальная оценка модели потока вызовов на современных телефонных сетях // Наук. пр. ОНАЗ им. О.С. Попова. 2001. № 2. С. 40-43. 4. Ложковский А.Г. Исследование параметров телефонной нагрузки на сотовой сети мобильной радиосвязи // Тр. УНИИРТ. 2001. М.3. С. 10-14. 5. Кленцрок Л. Теория массового обслуживания: Пер. с англ. М.: Машиностроение, 1979. 432 с. 6. Ложковский А.Г. Исследование системы обслуживания с ожиданием и рекуррентным потоком вызовов // Наук. пр. ОНАЗ им. О.С. Попова. 2004. С. 66-69.

Харьковский национальный
университет радиоэлектроники

Поступила в редакцию 10.09.2008