УДК.621.391.05

ИССЛЕДОВАНИЕ МОДЕЛИ НА ОСНОВЕ РЕКУРСИВНОГО АЛГОРИТМА ДЛЯ ОПРЕДЕЛЕНИЯ ОБЪЕМА КАНАЛЬНОГО РЕСУРСА ТРАФИКА В СЕРВИСАХ РЕАЛЬНОГО ВРЕМЕНИ

ЗИМИН И.В.

КГТУ им. И.Раззакова

izvestiya@ktu.aknet.kg

В статье исследуется базовая модель, которая является мультисервисным аналогом классической модели Эрланга. В анализируемой системе рассматривается процесс совместного использования канального ресурса звена мультисервисной сети произвольным числом пуассоновских потоков заявок, различающихся интенсивностью поступления, количеством ресурса, выделяемого для обслуживания одной заявки, и временем его занятия на передачу информации пользователя. Для исследуемой модели обслуживания заявок получены все основные результаты, которые ранее были получены для модели Эрланга и способствовали её широкому распространению среди инженеров, занимающихся проектированием сетей связи.

The article deals with the basic model, which is analogous to the classical model of multiservice Erlang. In the analyzed system deals with the process of sharing of resource managers multiservice network-layer any number of applications, varying flows for Poisson intensity, the amount of resources allocated to Service SA one application and the time of his classes on the transfer of user information. For the studied models of service applications received for all of the major results of the model were obtained for Erlang and contributed to its wide dissemination among the engineers involved in the design of Tei connection.

Введение. Для моделей мультисервисных систем связи схема занятия канального ресурса зависит от типа поступившей заявки. Выделение групп однородных событий, описывающих последовательность моментов поступления заявок, не приводит их к одному потоку, как это происходит в классических моделях теории телетрафика. Процесс обслуживания каждой группы заявок необходимо рассматривать отдельно. Таким образом, возникает класс много потоковых моделей. Привязка модели к реальным условиям обслуживания поступающих заявок происходит на этапе формализации, когда технические характеристики систем связи (скорость линий, доступность и т.д.) интерпретируются в терминах понятий, используемых при описании соответствующих моделей теории телетрафика.

Исследуемая модель. Основной областью использования исследуемой модели является определение необходимого объёма канального ресурса для передачи трафика сервисов реального времени.

Достаточность ресурса будет оцениваться долей потерянных заявок. В рамках предположений для вычисления этого показателя достаточно знать долю времени пребывания мультисервисной линии в состоянии с известным числом заявок каждого вида, находящихся на обслуживании.

Выбор показателей обслуживания заявок задаёт вид пространства состояний исследуемой модели и структуру случайного процесса, описывающего динамику их изменения. На рис. 1 для исследуемой модели показаны пространство состояний и соответствующая диаграмма переходов при $n = 2$.

Принято, что $bi = b_2 = 1$. Схема построения модели и введённые ограничения на изменения входных параметров позволяют утверждать, что процесс $r(t)$ — Марковский. Если дополнительно предположить, что все состояния, входящие в S, — сообщающиеся, т.е. из каждого состояния в любое другое и обратно можно попасть в результате некоторого числа поступлений заявок или освобождений канального ресурса, то можно сделать вывод о наличии для $r(t)$ стационарного режима [1].

Обозначим через $p(i_1,i_2,\ldots\ldots\ldots,i_n)$, стационарную вероятность состояния $(i_1,i_2,\ldots\ldots\ldots,i_n)$, где — число заявок k-го потока, находящихся на обслуживании в стационарном режиме, k=1,2,..., n.

В дальнейшем $p(\bullet)$ используется для обозначения нормированных значений вероятностей стационарных состояний исследуемых моделей, а $P(\bullet)$ — для обозначения ненормированных значений стационарных вероятностей.

Значения стационарных (предельных) вероятностей Марковского процесса $p(i_1,i_2,\ldots\ldots\ldots,i_n)$, имеют интерпретацию доли времени пребывания линии в состоянии $(i_1,i_2,\ldots\ldots\ldots,i_n)$.

Используя данную интерпретацию, дадим определение основным показателям QoS, необходимым для оценки достаточности канального ресурса линии.

Качество обслуживания заявок k-го потока определяется долей π_k потерянных заявок и средней величиной занятого канального ресурса.

Доля потерянных заявок k-го потока находится как доля времени пребывания процесса $r(t)$ в состояниях, когда приём поступившей заявки невозможен из-за нехватки свободного канального ресурса.

Справедливость данного утверждения вытекает из свойств пуассоновского потока, которому подчиняются моменты поступления заявок k-го потока.

Обозначим через U_k множество соответствующих состояний. Оно включает в себя состояния $(i_1,i_2,\ldots\ldots\ldots,i_n)\in S$, удовлетворяющие условию:

$$i_1 b_1 + i_2 b_2 + \ldots\ldots + i_n b_n > v - b_k \qquad (1)$$

Занумеруем $b_к$ так, чтобы выполнялись неравенства: $b_1 \leq b_2 \leq \ldots \leq b_n$

ABBYY PDF Transformer 2.0 Click here to buy www.ABBYY.com

Построенная модель хорошо известна в теории телетрафика и её исследованию посвящено множество работ [2,3]. Она имеет ряд отличительных особенностей, значительно упрощающих вычисление введённых показателей качества обслуживания заявок. Начнём со свойства мультипликативности значений стационарных вероятностей $p(i_1, i_2, \dots, i_n)$. Выполнение данного свойства позволяет представить $p(i_1, i_2, \dots, i_n)$ в виде следующего соотношения

$$p(i_1, i_2, \dots, i_n) = \frac{1}{N} \frac{\left(\frac{\lambda_1}{\mu_1}\right)^{i_1}}{i_1!} \frac{\left(\frac{\lambda_2}{\mu_2}\right)^{i_2}}{i_2!} \cdots \frac{\left(\frac{\lambda_n}{\mu_n}\right)^{i_n}}{i_n!}, (i_1, i_2, \dots, i_n) \in S. \quad (2)$$

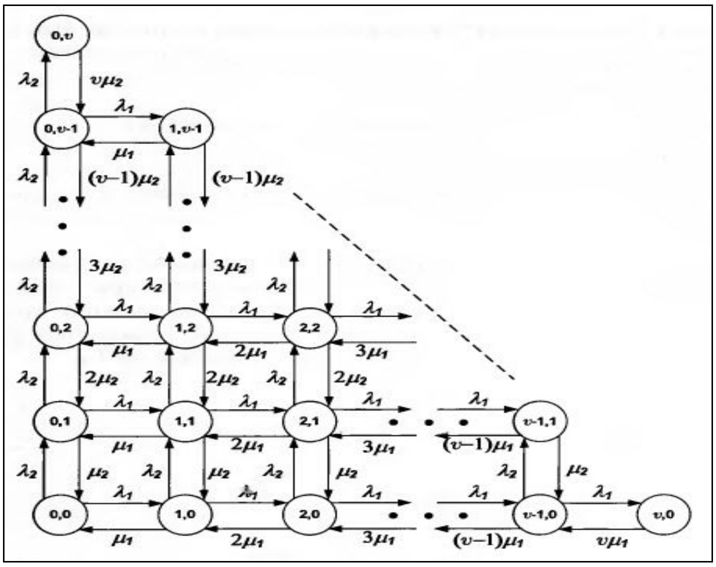

Рис.1. Пример пространства состояний и диаграмма переходов для базовой модели мультисервисной линии

В приведённом выражении символ N означает нормировочную константу. Она определяется из равенства

$$N = \sum_{(i_1, i_2, \dots, i_n) \in S} \frac{\left(\frac{\lambda_1}{\mu_1}\right)^{i_1}}{i_1!} \frac{\left(\frac{\lambda_2}{\mu_2}\right)^{i_2}}{i_2!} \cdots \frac{\left(\frac{\lambda_n}{\mu_n}\right)^{i_n}}{i_n!}$$

Свойство мультипликативности легко установить в ситуации, когда объём канального ресурса линии не ограничен, т.е. $v = \infty$. Действительно, в данном частном случае значение $i_k(t)$ числа заявок k-го потока, находящихся в момент времени t на обслуживании, описывается процессом рождения и гибели с интенсивностями рождения λ_k и гибели $i_k \mu_k, k = 1, 2, \cdots, n$. В силу неограниченности ресурса линии все n процессов не зависят друг от друга. Для k-го потока стационарная вероятность $p_k(i_k)$ нахождения i_k заявок на обслуживании определяется из равенства:

$$p_k(i_k) = e^{-\frac{\lambda_k}{\mu_k}} \frac{\left(\frac{\lambda_1}{\mu_1}\right)^{i_k}}{i_k!}, \qquad i_k = 0, 1, \ldots, \quad k = 1, 2, \cdots, n.$$

Для совместных вероятностей получаем следующее выражение:

$$p(i_1, i_2, \ldots, i_n) = e^{-\sum_{k=1}^{n} \frac{\lambda_k}{\mu_k}} \frac{\left(\frac{\lambda_1}{\mu_1}\right)^{i_1}}{i_1!} \frac{\left(\frac{\lambda_2}{\mu_2}\right)^{i_2}}{i_2!} \cdots \frac{\left(\frac{\lambda_n}{\mu_n}\right)^{i_n}}{i_n!} \tag{3}$$

$i_k = 0, 1, \ldots, \qquad k = 1, 2, \ldots, n.$

Оно следует из независимости процессов обслуживания заявок отдельных потоков. Таким образом, в данном частном случае мы установили свойство мультипликативности для вероятностей совместного обслуживания заявок. Чтобы показать справедливость (2) в общем случае, заметим, что для модели с неограниченным канальным ресурсом процессы $i_k(t), k = 1, 2, \ldots, n$ независимы и обладают свойством обратимости. Обратимые Марковские процессы обладают рядом замечательных характеристик, важнейшей из которых является сохранение свойства обратимости и значений (с точностью до нормировки) стационарных вероятностей после выполнения некоторых видов урезания используемого пространства состояний. Процедура урезания связана с возможностью изменения характера переходов анализируемого обратимого процесса. Приведём запись этого утверждения. Для обратимого Марковского процесса $r(t)$, определённого на счётном пространстве состояний Ω и имеющего стационарное распределение $\pi(j),\ j \epsilon \Omega$, совершим урезание используемого пространства состояний до размеров $S \subset \Omega$. Далее переопределим процесс $r(t)$, положив равными нулю интенсивности перехода из состояний, принадлежащих множеству S, в состояния, принадлежащие множеству $\Omega \backslash S$. Если построенный таким образом Марковский процесс $r*(t)$ обладает свойством неприводимости на пространстве S, то процесс $r*(t)$ будет также и обратим, а его стационарные вероятности $\pi*(j)$ определяются из соотношений:

$$\pi*(j) = \frac{\pi(j)}{\sum_{s \in S} \pi(s)}, \quad j \in S \qquad (4)$$

Чтобы проверить наличие свойства неприводимости у Марковского процесса, использованного для описания анализируемой модели мультисервисной линии, достаточно проверить, что из произвольного состояния, принадлежащего пространству состояний S, можно попасть в любое другое состояние из множества S и обратно в результате поступления некоторого числа заявок и окончания их обслуживания. В рассматриваемом случае все условия сформулированного утверждения выполняются. Следовательно, для стационарных вероятностей модели мультисервисной линии с ограниченной скоростью передачи справедливы равенства (2), что и доказывает требуемый результат.

Свойство мультипликативности, а также используемая схема распределения канального ресурса дают возможность построить эффективные рекурсивные алгоритмы расчёта стационарных характеристик модели. Следует отметить, что свойство мультипликативности в форме (2) сохраняется и при изменении характера функции распределения времени обслуживания заявки *k*-го потока. Для этого достаточно потребовать, чтобы для всех заявок *k*-го потока вид функции распределения был одинаков, а времена обслуживания заявок не зависели бы друг от друга и от поступающих потоков. Доказательство данного утверждения осуществляется ап-проксимацией произвольного распределения длительности обслуживания эрланговских распределений или распределений фазового типа, а для них справедливость соотношений (2) доказывается преобразованием системы уравнений равновесия.

Таким образом, расчётные алгоритмы, разработанные на основе (2), можно использовать и в ситуации, когда для заявки *k*-го потока время занятия канального ресурса на передачу информации имеет детерминированное распределение (т.е. постоянно), эрланговское распределение, распределение фазового типа. Устойчивость мультипликативного представления по отношению к функции распределения времени обслуживания заявки сохраняется и при некотором ослаблении ограничений на модель входного потока и структуру анализируемой сети связи. Для упрощения вида последующих выражений перейдём к записи интенсивностей поступающих заявок в Эрлангах, т.е. в числе заявок, поступающих за среднюю продолжительность обслуживания одной заявки, которая далее будет принята за единицу. Введём соответствующие обозначения

$$a_k = \frac{\lambda_k}{\mu_k}, \quad k = 1, 2, \dots, n. \qquad (5)$$

Для исследуемой модели значение a_k задаёт интенсивность предложенного трафика *k*-го потока заявок. Соответствующая характеристика определяется как среднее число заявок *k*-го потока, находящихся на обслуживании в отсутствии потерь. Величина a_k определяет потенциальное число соединений для рассматриваемого потока заявок.

Таким образом, свойство мультипликативности (2) приобретает вид:

$$p(i_1, i_2, \dots, i_n) = \frac{1}{N} \frac{a_1^{i_1}}{i_1!} \frac{a_2^{i_2}}{i_2!} \cdots \frac{a_n^{i_n}}{i_n!}, \quad (i_1, i_2, \dots, i_n) \in S. \qquad (6)$$

$$N = \sum_{(i_1, i_2, \dots, i_n) \in S} \frac{a_1^{i_1}}{i_1!} \frac{a_2^{i_2}}{i_2!} \cdots \frac{a_n^{i_n}}{i_n!}$$

Полученные выражения можно использовать для оценки π_k, m_k.

$$\pi_k = \frac{1}{N} \sum_{(i_1, i_2, \dots, i_n) \in U_k} \frac{a_1^{i_1}}{i_1!} \frac{a_2^{i_2}}{i_2!} \cdots \frac{a_{1n}^{i_n}}{i_n!} \qquad (7)$$

$$m_k = \frac{b_k}{N} \sum_{(i_1, i_2, \dots, i_n) \in S} \frac{a_1^{i_1}}{i_1!} \frac{a_2^{i_2}}{i_2!} \cdots \frac{a_n^{i_n}}{i_n!} i_k$$

Соотношения (7) при $n = 1$ активно используются при проектировании сетей. Приведём соответствующие расчётные выражения. Пусть для простоты $b_1 = 1$. Поскольку поток заявок один, то здесь и далее при записи параметров и характеристик моносервисной модели не использовать нижний индекс для обозначения номера потока и примем $a = \frac{\lambda}{\mu}$. Получаем $S = (0,1,\ldots,v)$, $U = (v)$. Выражения (7) приобретают вид:

$$\pi = \frac{1}{N}\frac{a^v}{v!} = \frac{\frac{a^v}{v!}}{1+a+\frac{a^2}{2!}+\cdots+\frac{a^v}{v!}}, \tag{8}$$

$$m = \frac{1}{N}\sum_{i=1}^{v}\frac{a^i}{i!}i = a(1-\pi)$$

Выражение для доли потерянных заявок π представляет собой формулу Эрланга [2]. Для записи π используется специальное обозначение $E(v,a)$.

Аналитические выражения (2) и (6) обычно не применяются для вычислений, поскольку число неизвестных вероятностей $p(i_1, i_2, \ldots, i_n)$ быстро увеличивается, начиная со сравнительно небольших значений v и n. Примем для простоты, что $b_k = 1, k = 1,2,\ldots,n$. Тогда общее число состояний в пространстве S можно найти из выражения:

$$C_v^m = \frac{(v+1)(v+2)\cdots(v+n)}{n!} \approx \frac{e^n}{\sqrt{2\pi n}}\left(\frac{v}{n}\right)^n$$

оценка в правой части приведённого соотношения получена с использованием формулы Стирлинга, что при $n = 10, v = 100$ число состояний превосходит значение $2 \cdot 10^{13}$, а при $n = 10$, $v = 1000$ значение $2 \cdot 10^{23}$.

Тем не менее, наличие выражений (2) и (6) значительно упрощает процедуру расчёта введённых показателей качества обслуживания поступающих заявок.

Рекурсивный алгоритм. Эффективный алгоритм оценки $\pi_k, m_k, k = 1,2,\ldots,n$ основан на использовании значений вероятностей пребывания $r(t)$ во множестве состояний, куда входят состояния $(i_1, i_2, \ldots, i_n) \in S$ удовлетворяющие условию:

$$i_1 b_1 + i_2 b_2 \cdots i_n b_n = i$$

Значение i меняется от 0 до v и показывает, сколько единиц ресурса мультисервисной линии используется всеми поступающими потоками заявок. Пространство состояний S можно представить как объединение всех взаимно непересекающихся подмножеств $S_i, i = 0,1,\cdots,v$. Таким образом, получаем:

$$S = \cup_{i=09}^{v} S_i \tag{9}$$

Определим $p(i)$из равенства:

$$p(i) = \sum_{(i_1, i_2, \dots, i_n) \in S_i} p(i_1 i_2, \dots, i_n), \quad i = 0, 1, \cdots, v.$$

Введённые показатели качества обслуживания заявок $k-$го потока могут быть найдены, если известны только значения $p(i)$.

$$\pi_k = \frac{1}{N} \sum_{(i_1, i_2, \dots, i_n) \in U_k} \frac{a_1{}^{i_1}}{i_1!} \frac{a_2{}^{i_2}}{i_2!} \cdots \frac{a_{1n}{}^{i_n}}{i_n!} =$$

$$= \frac{1}{N} \sum_{i=v-b_k+1}^{v} \sum_{(i_1, i_2, \dots, i_n) \in S_i} \frac{a_1^{i_1}}{i_1!} \frac{a_2^{i_2}}{i_2!} \cdots \frac{a_{1n}^{i_n}}{i_n!} = \sum_{i=v-b_k+1}^{v} p(i) \qquad (10)$$

Для оценки m_k можно использовать равенство:

$$m_k = a_k b_k (1 - \pi_k) \qquad (11)$$

Из равенств (10) и (11) следует, что для вычисления π_k, m_k, достаточно знать величину вероятностей $p(i), i = 0, 1, \dots, v$. Построим алгоритм оценки значений $p(i), i = 0, 1, \dots, v$, воспользовавшись для этого результатами [4]. В частности, покажем, что нормированные величины искомых вероятностей $p(i), i = 0, 1, \dots, v$ связаны рекуррентными соотношениями следующего вида:

$$p(i) = \frac{1}{i} \sum_{k=1}^{n} a_k b_k p(i - b_k) I(i - b_k \geq 0), \qquad i = 1, 2, \cdots, v, \qquad (12)$$

где I(•) — индикаторная функция. Отметим, что выражение (12) выполняются и для ненормированных значений $P(i)$.

Для доказательства (12) преобразуем $ip(i)$ к виду:

$$ip(i) = \sum_{(i_1, i_2, \dots, i_n) \in S_i} ip\,(i_1, i_2, \dots, i_n) = \sum_{(i_1, i_2, \dots, i_n) \in S_i} \left(\sum_{k=1}^{n} b_k i_k\right) p\,(i_1, i_2, \dots, i_n) =$$

$$= \sum_{k=1}^{n} b_k \sum_{(i_1, i_2, \dots, i_n) \in S} p\,(i_1, i_2, \dots, i_n) i_k \qquad (13)$$

Воспользовавшись мультипликативным представлением стационарных вероятностей (6), получим выражение для:

$$\sum_{(i_1, i_2, \dots, i_n) \in S_i} p\,(i_1, i_2, \dots, i_n) i_k$$

через $p(i—bk)$,соответствующее выражение выглядит следующим образом:

$$\sum_{(i_1,i_2,\dots,i_n)\in S_i} p\,(i_1,i_2,\dots,i_n) i_k = \frac{1}{N}\sum_{(i_1,i_2,\dots,i_n)\in S_i} \frac{a_1{}^{i_1}}{i_1!}\frac{a_2{}^{i_2}}{i_2!}\cdots\frac{a_{1n}{}^{i_n}}{i_n!}\times i_k =$$

$$= \frac{a_k}{N}\sum_{(i_1,i_2,\dots,i_n)\in S_i} \frac{a_1^{i_1}}{i_1!}\cdots\frac{a_k^{i_k-1}}{(i_k-1)!}\cdots\frac{a_n^{i_n}}{i_n!} I(i_k>0) =$$

$$= a_k \sum_{(i_1,i_2,\dots,i_n)\in S_i} p(i_1,\dots,i_k-1,\dots,i_n) I\,(i_k>0) =$$

$$= a_k \sum_{(i_1,i_2,\dots,i_n)\in S_{i-b_k}} p(i_1,i_2,\dots,i_n) I(i-b_k\geq 0) = a_k p(i-b_k) I\,(i-b_k\geq 0). \quad (14)$$

Подставив (14) в (13), получаем рекуррентную формулу (12).

Найдём выражение (11) для оценки m_k через значение доли потерянных заявок $k-$го потока. Для этого воспользуемся соотношениями (9), (14).

$$m_k = \sum_{(i_1,i_2,\dots,i_n)\in S} p(i_1,i_2,\dots,i_n) i_k b_k = \quad (15)$$

$$= b_k \sum_{i=0}^{v} \sum_{(i_1,i_2,\dots,i_n)\in S_i} p(i_1,i_2,\dots,i_n) i_k =$$

$$= a_k b_k \sum_{i=0}^{v} p(i-b_k) I(i-b_k\geq 0) =$$

$$= a_k b_k \sum_{i=0}^{v-b_k} p(i) = a_k b_k \left(1-\sum_{i=v-b_k+1}^{v} p\,(i)\right) = a_k b_k (1-\pi_k)$$

Используя соотношения (12), нетрудно выразить значения ненормированных вероятностей $P(i)$ через ненормированное значение вероятностей какого-либо одного состояния, например, через $P(0)$. Далее после нормировки находятся вероятности $p(i), i = 0,1,\dots,v$, а с ними и величины характеристик $\pi_k, m_k, k = 1,2,\dots,n$. Перечислим шаги соответствующего рекурсивного алгоритма.

1. Положим значение $P(0) = 1$.

2. Выразим значения вероятностей $P(i), i = 1,2,\dots,v$ через $P(0)$, используя соотношение

$$P(i) = \frac{1}{i}\sum_{k=1}^{n} a_k b_k P(i-b_k) I(i-b_k\geq 0) \quad (16)$$

и последовательно увеличивая величину i, от 1 до v. При каждом фиксированном i значения выражений $\mathrm{P(i - b_k)I(i - b_k > 0)}$, $k = 1,2,\dots,n$, участвующих в записи правой части суммы (16), либо уже представлены через $P(0)$ (для $\mathrm{i - b_k \geq 0}$), либо равны 0 (для $\mathrm{i - b_k < 0}$).

Находим величину нормировочной константы:

$$N = \sum_{i=0}^{v} P(i)$$

Определяем нормированные значения вероятностей *p(i):*

$$p(i) = \frac{P(i)}{N}, \quad i = 0, 1, \dots , v.$$

Находим величину введённых показателей качества совместного обслуживания заявок для каждого из n анализируемых потоков:

$$\pi_k = \sum_{i=v-b_k+1}^{v} p(i), \qquad m_k = a_k b_k (1 - \pi_k), \quad k = 1, 2, \dots, n.$$

Вывод. Оценим вычислительную сложность реализации построенного алгоритма количеством выполненных арифметических операций.

Основные усилия тратятся на вычисление $\mathrm{P(i)}$ в соответствии с шагом 2. При этом выполняется $\mathrm{O(n)}$ арифметических операций.

Общее число операций, которые необходимо совершить для вычисления всех вероятностей, оценивается величиной $\mathrm{O(nv)}$.

Так как n — константа, то отсюда следует, что время счёта увеличивается линейно с ростом объёма канального ресурса линии.

Таким образом, для каждого из n потоков решена задача определения доли потерянных заявок, а также среднего значения канального ресурса линии, занятого на их обслуживание.

Поскольку появилось эффективное средство оценки показателей качества обслуживания заявок, то тем самым значительно упростилось проблема нахождения минимального объема канального ресурса линии, обеспечивающего заданные (нормированные) значения характеристик приема заявок на обслуживание.

Литература

1. Зимин И.В., Алымкулов С. А. Исследование модели оценки канального ресурса для сервисов реального времени // Теоретический и прикладной научно-технический журнал, Бишкек, 2013г.
2. Зимин И.В. Учебник: «Управление трафиком в сетях и системах телекоммуникаций» // Издательство «Алтын Принт», Бишкек 2012г.;
3. Степанов С.Н. Материалы курса лекций «Основы теории моделирования сетей и систем телекоммуникаций» // Московский технический университет связи и информатики, Москва, 2008г.
4. Лагутин В.С., Степанов С.Н. Трафик мультисервисных сетей связи //Издательство «Радио и связь», Москва, 2000г.

5. Зимин И.В., Баянкина Е.В. Практическое применение задач управления сетевыми ресурсами в телекоммуникационных сетях // Журнал Проблемы автоматики и управления №2, Институт автоматики Национальной Академии наук КР, материалы международной конференции «Проблемы управления и информационных технологий», Кыргызстан Бишкек, 2010г. С. 31-35.

6. Зимин И.В.Методы теории массового обслуживания в задачах моделирования перспективных телекоммуникационных систем // ГОУ «Восточно - Сибирский государственный технологический университет» Материалы конференции часть 1 десятой Всероссийской научно-технической конференции «Теоретические и прикладные вопросы современных информационных технологий», Российская Федерация город Улан-Удэ, июль 2009г. С. 66-72

7. Степанов С.Н. Основы телетрафика мультисервисных сетей // Издание ЭКОТРЕНДЗ, Москва, 2010г.;

8. Степанов С.Н.; Иверсин В.Б. Способы уменьшения объема вычислений при расчете моделей систем связей с потерями, основанные на игнорировании маловероятных состояний // Проблемы передачи информации. 2001. Том. 37. Выпуск. 3.;