УДК 621.39

# Системы $M|G|1$ с групповым обслуживанием и их применение к анализу модели протокола управления потоковой передачей. Часть I

**Н. В. Першаков, К. Е. Самуйлов**

*Кафедра систем телекоммуникаций*
*Российский университет дружбы народов*
*ул. Миклухо-Маклая, 6, Москва, Россия, 117198*

В статье представлен обзор методов анализа систем массового обслуживания типа $M|G|1$ с групповым обслуживанием заявок, подготовленный в процессе исследований показателей качества функционирования протокола управления потоковой передачей (Stream Control Transmission Protocol, SCTP). В обзоре, выполненном с учётом хронологии наиболее известных публикаций, сформулированы имеющиеся на сегодняшний день результаты анализа основных вероятностных характеристик систем с обслуживанием группами переменной и фиксированной длины. Получены характеристики виртуального времени ожидания для анализа модели протокола SCTP.

**Ключевые слова:** протокол управления потоковой передачей, система массового обслуживания, обслуживание группами переменной длины, обслуживание группами фиксированной длины.

## 1. Введение

Данная статья состоит из двух частей, первая из которых представляет собой обзор методов анализа и расчёта характеристик систем массового обслуживания (СМО) типа $M|G|1$ с групповым обслуживанием заявок. Обзор был выполнен в процессе исследования конкретной прикладной задачи, постановка и решение которой представлены во второй части статьи[1] . Изначально авторы не ставили перед собой цели развития методов анализа СМО с групповым обслуживанием, предполагалось, что теоретические результаты в объёме, достаточном для решения прикладных вопросов, будут найдены в научных публикациях и останутся лишь известные специалистам по теории массового обслуживания чисто технические сложности, связанные с расчётом вероятностно-временных характеристик.

Вкратце, задача состоит в том, что для СМО с групповым обслуживанием требуется найти первые два начальных момента случайной величины (СВ) времени пребывания заявки (или группы заявок) в системе и несколько более «тонкую» характеристику — 95% квантиль этой СВ. Такого рода задача является достаточно традиционной для приложений в телекоммуникациях. Дело в том, что в международных отраслевых стандартах требования к показателям качества обслуживания (Quality of Service, QoS) представлены в виде технических норм для среднего значения, среднеквадратического отклонения и квантили, как правило, уровня 0,95, соответствующей СВ, например, задержки передачи сообщения по телекоммуникационной сети. Выходить за рамки предположений о пуассоновском входящем потоке в большинстве случаев не целесообразно, поскольку искомое решение должно быть найдено в виде, доступном для инженерных расчётов. В силу изложенного исследования были ограничены классом СМО, тип которых определён в заголовке статьи. Начальный обзор советских и более современных российских публикаций, в том числе переводных работ зарубежных авторов, не дал исчерпывающего ответа на поставленные вопросы, и поэтому было выполнено более детальное исследование, которое и излагается далее в весьма сжатом виде с учётом хронологии наиболее известных публикаций.

Статья поступила в редакцию 30 октября 2008 г.
Статья подготовлена при поддержке национального приоритетного проекта «Образование».

[1]Часть II статьи будет опубликована в одном из следующих номеров Вестника РУДН данной серии.

Системы с групповым обслуживанием находят применение для решения широкого спектра научно-прикладных задач. Разнообразие проблем, стоящих перед исследователями, привело к разработке самых различных модификаций таких СМО — с всевозможными типами входящих потоков и дисциплинами обслуживания, приоритетами, конечной очередью, ненадёжным прибором и др. Важной областью применения являются телекоммуникации, переживающие в настоящий момент очередной и весьма резкий скачок в своём развитии. Растущие требования пользователей и возможности современного оборудования сопровождаются появлением новых технологий, одной из которых стал протокол SCTP, обеспечивающий в IP-сетях надёжную передачу данных при оказании пользователям услуг реального времени. Исследования протокола, проведённые авторами данной статьи [1, 2], выявили возможность анализа его QoS-параметров посредством применения систем с групповым обслуживанием, но имеющиеся на сегодняшний день теоретические результаты исследований рассматриваемого класса СМО не полностью обеспечивают прикладные вычислительные аспекты. Применение этих результатов к решению поставленной задачи требует развития существующих и разработки новых моделей и методов анализа вероятностных характеристик, что показано во второй части данной статьи.

Статья структурирована следующим образом. В разделе 2 представлен обзор методов исследования, а в разделе 3 на основе этого обзора проведён анализ современного состояния основных результатов для СМО типа $M|G|1$ с групповым обслуживанием. В разделе 4 получены вероятностно-временные характеристики, необходимые для исследования важнейших показателей качества обслуживания протокола SCTP.

## 2. Обзор методов исследования

Можно считать, что начало исследованиям СМО с групповым обслуживанием положено в работах Н. Бейли и Ф. Даунтона [3, 4], опубликованных в середине 1950-х гг. В этот период, в рамках мероприятий, нацеленных на улучшение здравоохранения Великобритании, возникла задача установления зависимости времени пребывания нового пациента в так называемом “списке ожидания” от потребности населения в медицинском обслуживании. Обратившись к её решению, Н. Бейли впервые рассмотрел систему с пуассоновским входящим потоком заявок (пациентов) и переменной, ограниченной сверху длиной группы заявок (максимальное количество пациентов, которых специалист может принять во время работы клиники). Впоследствии, в монографии М. Шодри и Дж. Темплетона [5] в классификации Башарина-Кендалла эта система была кодирована как $M|G^B|1$, где индекс «$B$» обозначает обслуживание заявок группами максимальной длины $B$ и имеет мнемонический смысл Bulk (группа). Далее в статье, ввиду отсутствия единого принятого обозначения данного параметра, при записи аналитических выражений мы будем использовать символ «$K$», но соответствующие СМО — кодировать без изменений, как приведено в рассматриваемых работах.

Для анализа системы $M|G^B|1$ Н. Бейли воспользовался методом вложенных цепей Маркова (ЦМ), где в качестве точек регенерации были выбраны моменты, непосредственно предшествующие окончанию обслуживания групп заявок. Производящую функцию (ПФ) $Q^-(z)$ стационарного распределения $q_i^-$ числа заявок в очереди он получил в работе [3] в следующем виде

$$Q^-(z) = \frac{\beta\left(\lambda - \lambda z\right) \sum\limits_{i=0}^{K-1} q_i^- \left(z^i - z^K\right)}{\beta\left(\lambda - \lambda z\right) - z^K}, \tag{1}$$

где $\beta\left(\cdot\right)$ — преобразование Лапласа-Стилтьеса (ПЛС) функции распределения (ФР) времени обслуживания $B\left(\cdot\right)$, в качестве которой было рассмотрено распределение $\chi^2$ (хи-квадрат) с чётным числом степеней свободы и его частные случаи — экспоненциальное и детерминированное. Для этих распределений Н. Бейли получил среднее значение и дисперсию числа заявок в очереди, выраженные через корни характеристического уравнения $z^K = \beta\left(\lambda - \lambda z\right)$, лежащие как внутри

$|z| < 1$ (детерминированное обслуживание), так и вне $|z| > 1$ (экспоненциальное обслуживание) единичного круга.

Ф. Даунтон, в продолжение исследований Н. Бейли, нашёл связь ПФ $Q^{-}(z)$ с ПЛС ФР $W(x)$ времени ожидания начала обслуживания заявки $\omega(t)$ при условии того, что в момент её поступления прибор занят,

$$w(s|\omega(t) > 0) = \frac{1-\beta(s)}{sm_B\beta(s)}Q^{-}\left(1-\frac{s}{\lambda}\right), \tag{2}$$

где $m_B = \int\limits_0^\infty x\mathrm{d}B(x)$ — среднее время обслуживания. В [4] Ф. Даунтон получил выражения для среднего значения и дисперсии времени ожидания, предложил метод вычисления квантилей, так же, как и Н. Бейли, в качестве дисциплины обслуживания им было рассмотрено распределение $\chi^2$.

Исследования систем с групповым обслуживанием заявок были продолжены многими учёными. Дж. Миллер [6] впервые рассмотрел СМО с групповым поступлением и групповым обслуживанием, в которой длины групп заявок, поступающих в систему или на обслуживающий прибор, являются дискретными случайными величинами. Для произвольных распределений входящего потока и времени обслуживания он нашёл условие эргодичности очереди. При пуассоновском поступлении заявок методом вложенных ЦМ Дж. Миллер исследовал время ожидания и период занятости, а для случая экспоненциального обслуживания — характеристики стационарного распределения очереди.

В приложении к задаче анализа работы пассажирских лифтов, городских автобусов, так называемой «задаче перевозки» (Transportation Problem), Н. Жайсвал в [7] предложил СМО, в которой длина группы заявок, поступающей на обслуживание, зависит от числа заявок, уже обслуживаемых прибором, кроме того, прибор может обслуживать группы нулевой длины. Для исследования этой системы он также использовал метод вложенных ЦМ. Подробное описание этого метода, как одного из основных для анализа СМО с групповым обслуживанием, содержится в работе Ю. Бхата [8].

П. Финч в [9] исследовал СМО с конечной очередью и экспоненциальным обслуживанием группами фиксированной длины. Проводя аналогии с системой $E_m|M|1$ с эрланговским входящим потоком и экспоненциальным обслуживанием, он методом вложенных ЦМ получил выражения для ПФ распределения числа заявок в системе.

Дж. Гойял в [10] методом ПФ исследовал систему $\mathrm{H}_m|M|1$ с гиперэкспоненциальным входящим потоком и экспоненциальным обслуживанием группами переменной длины. Для системы с конечной очередью он нашёл вероятность простоя прибора, а для случая бесконечной очереди – выражения для стационарного по времени распределения числа заявок в очереди и средней длины очереди.

М. Ньютс [11] рассмотрел более общую в сравнении с введённой Н. Бейли СМО, в которой длина группы является переменной с нижней и верхней границей числа заявок, поступающих на обслуживание. Для обозначения такого типа обслуживания М. Ньютс ввёл термин «General bulk service rule», в его работе получены характеристики длины очереди, как в дискретные, так и в произвольные моменты времени, распределение периода занятости, а также предельные теоремы для числа обслуженных заявок за длительный промежуток времени. Впоследствии, обратившись к решению задачи М. Ньютса, Дж. Медхи в [12] провёл анализ времени ожидания заявок в этой системе.

В монографии [13] Дж. Коэн систематизировал имеющиеся к тому моменту результаты исследований одноканальных СМО, а также изложил свои достижения в данной области. В главе III «Some variants of the single server queue» этой монографии второй раздел посвящён анализу СМО, определённых Дж. Миллером, Н. Жайсвалом, М. Ньютсом и др., с произвольными распределениями входящего потока и обслуживания. Системы были кодированы как $G|G|1$ с указанием типа группового обслуживания или решаемой задачи. Дж. Коэн также рассмотрел систему, в которой группа заявок может поступать на прибор в момент времени, когда обслуживание текущей группы ещё не завершилось (Accessible Batches). Второе издание данной книги датируется 1982 г.

Практически одновременно с работой Дж. Коэна появилась монография Т. Саати, через два года переведённая на русский язык [14]. В главе VII книги достаточно подробно изложены результаты работ Н. Бейли, Ф. Даунтона, Дж. Миллера.

Результаты этих и многих других работ вошли в уже упоминавшуюся монографию, посвящённую исследованию СМО с групповым поступлением и обслуживанием заявок, подготовленную Дж. Темплетоном в соавторстве с М. Шодри [5]. В главе IV «Bulk-service queues» изучены системы с ординарным поступлением заявок и групповым обслуживанием: $M|G^B|1$ с переменной длиной группы заявок, $M|G^B|1|R$ с конечной очередью, $M|G^k|1$ с фиксированной длиной группы, $GEOM|G^B|1$ с геометрическим распределением входящего потока и $M|M^B|1|RB$ с экспоненциальным обслуживанием и накопителем ровно на $R$ групп заявок.

В [5] система $M|G^B|1$ исследована с помощью метода введения дополнительной переменной — прошедшего времени обслуживания группы заявок. Найдены выражения для ПФ $Q(z)$ стационарного распределения длины очереди по времени, для вложенной ЦМ $Q^-(z)$ в виде (1), а также ПЛС $w(s)$ виртуального времени ожидания в виде (2). Таким образом, показана эквивалентность результатов, полученных методом введения дополнительной переменной [5], и методом вложенных ЦМ, применённым Н. Бейли [3] и Ф. Даунтоном [4]. В частном случае эрланговского обслуживания проведено доказательство единственности представления ПФ $Q^-(z)$, а также найдена связь вероятности $P_0$ простоя прибора с вероятностью $q_0^-$ отсутствия заявок в очереди по вложенной ЦМ в виде

$$P_0 = \frac{q_0^-}{q_0^- + \rho}, \tag{3}$$

где $\rho = \lambda m_B$ — нагрузка на систему.

Для СМО $M|G^k|1$ с помощью ЦМ, вложенной по моментам непосредственно после окончания обслуживания групп заявок, в [5] получена ПФ $Q^-(z)$ в виде (1), что свидетельствует о взаимосвязи этой СМО с системой $M|G^B|1$, и ПЛС времени ожидания группы заявок

$$w_G(s) = \frac{Q^-\left(1 - \frac{s}{\lambda}\right)}{\beta\,(s)}. \tag{4}$$

Системы $M|G^B|1|R$ и $GEOM|G^B|1$ в [5] также исследованы методом вложенных ЦМ. В обоих случаях найдена ПФ распределения длины очереди и, кроме того, для СМО $GEOM|G^B|1$ — первые два начальных момента времени ожидания. Для СМО $M|M^B|1|RB$, частного случая системы $M|G^B|1|R$, обсуждаются некоторые детали анализа её вероятностных характеристик.

С момента выхода в печать монографии М. Шодри и Дж. Темплетона многие авторы обращались к исследованию СМО с групповым обслуживанием заявок, привнеся огромное количество модификаций и новшеств. Поэтому, для краткости несколько нарушая хронологический порядок, укажем лишь наиболее характерные работы, результаты которых необходимы для решения задачи исследований.

Возвращаясь к анализу СМО, введённой Н. Бейли, А. Янсен и Дж. ван Леуварден в [15] получили распределение длины очереди по вложенной ЦМ $\{x_n^-\}$, зависящей от корней характеристического уравнения $z^K = A(z)$, в виде

$$q_j^- = \frac{1}{\beta_0}\left[\sum_{i=1}^{j}\left(\delta_{i,K} - \beta_i\right) q_{j-i}^- - \frac{K-\rho}{\prod_{i=1}^{K-1}\left(1 - z_i\right)} \sum_{i=0}^{\min(j,K)} \beta_{j-i} r_i\right], \tag{5}$$

где $\delta_{i,K}$ — символ Кронекера, $r_i$ — коэффициенты разложения произведения $\prod\limits_{i=0}^{K-1}(z - z_i)$ в ряд по степеням $z$, и вид самих корней с помощью разложения в ряд Фурье

$$z_k = \sum_{l=1}^{\infty} c_l w_k^l, \quad k = \overline{0, K-1}, \tag{6}$$

где $w_k = e^{2\pi k i/K}$, а коэффициенты $c_l$ имеют представление

$$c_l = \frac{1}{2\pi i l} \int\limits_{|z|\leqslant 1} \frac{\beta^{l/K}(\lambda - \lambda z)}{z^l} \mathrm{d}z, \quad l \geqslant 1. \tag{7}$$

В этой работе ПФ $Q^-(z)$ (1) была записана в виде

$$Q^-(z) = (K - \rho) \frac{(1-z)\beta(\lambda - \lambda z)}{\beta(\lambda - \lambda z) - z^K} \prod_{i=1}^{K-1} \frac{z - z_i}{1 - z_i}. \tag{8}$$

Для обслуживания группами фиксированной длины, введённого П. Финчем, в случае бесконечной очереди Дж. Дшалалов [16] нашёл связь ПФ $P(z)$ стационарного по времени распределения числа заявок в системе и ПФ $Q^-(z)$

$$P(z) = \frac{1 - z^K}{(1-z)K} Q^-(z). \tag{9}$$

В [17] Л. Тадж, кодировав эту СМО как $M|G^r|1$, получил стационарные вероятности числа заявок в очереди, а также среднюю длину очереди в терминах корней характеристического уравнения, лежащих как внутри $|z| < 1$, так и вне $|z| > 1$ единичного круга. Анализ данной СМО в той или иной модификации продолжен в ряде работ, например, в [18] исследована система с ненадёжным прибором, кодированная как $M|G^K|1|FS$ (Fixed Size, FS).

В конце 1980-х–начале 1990-х г. в связи с развитием компьютерных технологий все большее число работ направлено на применение численных методов для расчёта параметров производительности систем с групповым обслуживанием и на решение конкретных прикладных задач: множественный доступ с разделением по времени (Time Division Multiple Access, TDMA); передача пакетов в компьютерных сетях; функционирование спутниковой коммутационной станции; передача данных в сетях с асинхронным режимом (Asynchronous Transfer Mode, ATM); агрегация фреймов в беспроводных локальных сетях (Wireless Local Area Network, WLAN); инкапсуляция и агрегация пакетов в сетях передачи данных; оптимизация управления питанием в сетях WLAN стандарта IEEE 802.11 и т.д.

В отечественной литературе анализу систем с групповым обслуживанием уделено достаточно мало внимания. Г. Ш. Цициашвили в [19] исследовал коммутационные эффекты в системах с групповым и приоритетным обслуживанием, в рамках простейших моделей построил количественные оценки и дал их содержательную интерпретацию. В [20] им же был проведён анализ системы $G|G|1|\Phi$ с групповым обслуживанием. В [21] П.П. Бочаров и С. В. Аль-Натор исследовали одноканальную систему конечной ёмкости с произвольным рекуррентным потоком и групповым марковским обслуживанием при специальной дисциплине, в работе найдено стационарное распределение очереди в произвольные моменты времени, моменты потока поступления заявок и выходящего потока. В. В. Чаплыгин в [22] рассмотрел однолинейную СМО с групповым обслуживанием типа $G|BMSP|1|r$ с накопителем конечной или бесконечной ёмкости. Для этой системы методом вложенных ЦМ он нашёл стационарные распределения основных характеристик обслуживания.

# 3. Современное состояние основных результатов исследований

Подводя итог вышесказанному, напомним основные вехи в истории исследований систем с групповым обслуживанием:

– первые работы Н. Бейли и Ф. Даунтона, посвящённые анализу СМО $M|G^B|1$ (1954 г.);
– монографии Дж. Коэна «The single server queue» и Т. Саати «Элементы теории массового обслуживания и её приложения», в которых авторы обобщили разрозненные результаты (1969 г.);
– первая монография по СМО с групповым поступлением и обслуживанием М. Шодри и Дж. Темплетона «A first course in bulk queues» (1983 г.);
– разработка эффективных методов и алгоритмов вычисления вероятностных характеристик СМО с групповым обслуживанием А. Янсен и Дж. ван Леуварден (2005 г.).

Говоря об эволюции методов исследований, необходимо заметить, что в большинстве работ используются тесно связанные между собой методы вложенных ЦМ и ПФ. Хорошо известно, что относительная их простота нивелируется ограниченными возможностями анализа характеристик системы. Производящая функция позволяет вычислить лишь среднее значение, дисперсию длины очереди и, используя формулу Литтла, среднее значение времени ожидания начала обслуживания. Для нахождения дисперсии последнего приходится прибегать к дополнительному анализу, например, виртуального времени ожидания. Метод введения дополнительной переменной лишён указанного недостатка, его конечной целью является выражение для совместного ПЛС числа заявок в очереди и времени ожидания, которое позволяет получить все перечисленные характеристики. Тем не менее, высокая сложность реализации этого метода для систем с групповым обслуживанием обуславливает весьма редкое его применение.

К настоящему моменту существует три основных типа систем с пуассоновским поступлением заявок, групповым обслуживанием одним прибором и бесконечной очередью:

– $M|G^B|1$ с обслуживанием группами переменной длины, ограниченной числом $B$ заявок;
– $M|G^k|1$ с обслуживанием группами фиксированной длины, ограниченной числом $k$ заявок;
– $M|G^{a,b}|1$ с обслуживанием группами переменной длины с нижней $a$ и верхней $b$ границами числа заявок.

Последняя СМО при значениях параметров $a = 1$, $b = B$ или $a = b = k$ очевидно приводится к системе с переменной или фиксированной длиной группы заявок соответственно.

В табл. 1 сведены имеющиеся на сегодняшний день результаты для указанных систем, а в примечаниях к ней — работы, в которых они были получены. Выбор СМО $M|G^B|1$ и $M|G^k|1$ обоснован тем, что именно они и являются предметом анализа в рамках поставленной задачи (см. вторую часть статьи), а СМО $M|G^{a,b}|1$ — как системы, непосредственно обобщающей системы $M|G^B|1$ и $M|G^k|1$. В качестве искомых характеристик нас, в первую очередь, будут интересовать среднее значение и дисперсия времени ожидания начала обслуживания, а также вспомогательные характеристики длины очереди в дискретные и произвольные моменты времени.

Как видно из таблицы, все три рассматриваемые СМО изучены в сравнительно одинаковой степени. Методом вложенных ЦМ, если не указано иное, получены ПФ стационарного распределения длины очереди в дискретные и произвольные моменты времени, ПЛС ФР времени ожидания начала обслуживания. Исключение составляет СМО $M|G^k|1$, для которой последняя характеристика найдена только для группы заявок, о чем говорит обозначение «$\pm$» в соответствующей ячейке таблицы. Система $M|G^{a,b}|1$ изучена в той же степени, поэтому результаты её исследования не дополняют результаты, полученные для СМО $M|G^B|1$ и $M|G^k|1$.

Таким образом, нерешёнными остаются задачи нахождения следующих характеристик:

– средней длины очереди для СМО $M|G^B|1$;
– дисперсии длины очереди в дискретные моменты времени для СМО $M|G^k|1$;

Таблица 1

**Вероятностно-временные характеристики СМО с групповым обслуживанием**

| Характеристика | СМО | | |
|---|---|---|---|
| | $M\|G^B\|1$ | $M\|G^k\|1$ | $M\|G^{a,b}\|1$ |
| ПФ стационарного распределения длины очереди для вложенной ЦМ | $+^1$ | $+^5$ | $+^9$ |
| среднее значение | $+^2$ | $+^6$ | $+^9$ |
| дисперсия | $+^2$ | – | – |
| ПФ стационарного по времени распределения длины очереди | $+^3$ | $+^7$ | $+^9$ |
| среднее значение | – | $+^6$ | $+^9$ |
| дисперсия | – | – | – |
| ПЛС ФР времени ожидания начала обслуживания | $+^4$ | $\pm^8$ | $+^9$ |
| среднее значение | – | – | – |
| дисперсия | – | – | – |

Таблица 2

**Примечания к табл. 1**

| | |
|---|---|
| [1] [Baily 1954], [Tadj 2003] | [5] [Chaudhry, Templeton 1983] |
| [2] [Denteneer, Janssen, van Leeuwaarden 2005] | [6] [Tadj 2003] |
| [3] [Chaudhry, Templeton 1983] получена методом введения дополнительной переменной | [7] [Dshalalow, Tadj 1992], [Lee, Lee, Chae 1996] |
| [4] [Downton 1955] | [8] [Chaudhry, Templeton 1983] получена для группы заявок |
| | [9] [Neuts, 1967] |

– дисперсии длины очереди в произвольные моменты времени для СМО $M|G^B|1$ и $M|G^k|1$;
– среднего значения и дисперсии времени ожидания начала обслуживания для СМО $M|G^B|1$ и $M|G^k|1$.

Резюмируя вышесказанное, необходимо сделать несколько замечаний. Во-первых, несмотря на то, что выражения для перечисленных средних значений не были обнаружены в рассмотренных работах, их вычисление, очевидно, при наличии соответствующих представлений ПФ и ПЛС не вызывает больших затруднений. Кроме того, авторы не претендует на абсолютную полноту настоящего литературного обзора в связи с достаточно серьёзным вниманием многих учёных-математиков к системам с групповым обслуживанием с момента их появления в 1954 г.

Во-вторых, большинство рассмотренных работ носят сугубо теоретический характер. Процесс исследования в этих работах, как правило, завершается на этапе нахождения выражений для ПФ длины очереди или ПЛС времени ожидания. Важная для телекоммуникационных приложений задача нахождения дисперсии в работах практически не рассматривается. Тем не менее, её решение является весьма трудоёмким. Непосредственное использование имеющихся результатов для вычисления искомых вероятностных характеристик весьма затруднительно, что ставит задачу развития и модификации существующих методов. Проще говоря, в отдельных случаях для достижения цели приходится заново воспроизводить теоретические результаты с целью представления их в удобном для вычислений виде.

Отдельно следует отметить задачу вычисления квантилей вероятностных характеристик, имеющую важное прикладное значение. За исключением работы Ф. Даунтона [4], в которой была сделана попытка поиска точного аналитического решения для расчёта квантилей времени ожидания, авторы не обращаются к данной задаче ввиду её высокой степени сложности. Даже в простейших случаях для СМО типа $M|M|1$ вычисление квантилей является нетривиальной задачей [1]. Таким образом, на сегодняшний день её решение производится либо с помощью приближенных формул, либо с применением численных методов или статистического моделирования.

## 4. Характеристики виртуального времени ожидания

Как было показано в разделе 2, основные результаты исследований виртуального времени ожидания $\omega(t)$ в СМО $M|G^B|1$ и $M|G^k|1$ изложены в работах [4,5,14]. Обратимся к вычислению недостающих характеристик времени ожидания (табл. 3).

Воспользуемся ПЛС $w(s)$ ФР времени ожидания $W(x)$ и заметим, что общее время ожидания заявки можно представить как сумму временных интервалов, показанных на рис. 1:

– остаточного времени формирования группы, т.е. времени с момента поступления заявки до момента освобождения прибора (для СМО $M|G^B|1$, если формирующаяся группа единственная в очереди) или момента поступления $K$-й, последней заявки в группу;
– времени ожидания начала обслуживания сформированной группы, т.е. времени с момента поступления $K$-й заявки в группу до момента поступления группы заявок на обслуживающий прибор.

Обозначим $W_C(x)$ (Customer, C) ФР остаточного времени формирования группы, а $W_G(x)$ (Group, G) — ФР времени ожидания начала обслуживания группы. Тогда ФР $W(x)$ является свёрткой ФР $W_C(x)$ и $W_G(x)$, а соответствующее ПЛС $w(s)$ имеет вид

$$w(s) = w_C(s)\, w_G(s), \tag{10}$$

и справедлива следующая теорема, которую мы формулируем без доказательства.

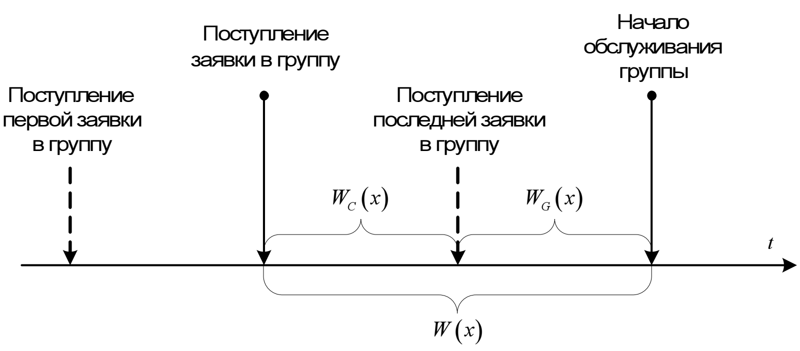


Рис. 1. **Диаграмма времени ожидания заявки**

**Теорема 1.** *Если $\rho < K$ и $|z| \leqslant 1$, тогда для СМО с переменной длиной группы*

$$w(s) = \frac{K-\rho}{q_0^- + \rho} \frac{1-\beta(s)}{\beta(s) - \left(1-\frac{s}{\lambda}\right)^K} \prod_{i=1}^{K-1} \left(1 - \frac{s}{\lambda(1-z_i)}\right). \tag{11}$$

Вычисляя соответствующие производные от ПЛС $w(s)$ в нуле, найдём среднее значение и дисперсию времени ожидания начала обслуживания.

**Следствие 1.** Если $\rho < K$ и $|z| \leqslant 1$, тогда

$$m_W = \frac{1}{q_0^- + \rho} \left[m_{X^-} m_B + \frac{f}{\lambda}\right], \tag{12}$$

$$\begin{aligned} \sigma_W^2 = {} & \tfrac{1}{q_0^- + \rho}\left[\tfrac{1}{3}\lambda m_B^{(3)} - 2\rho\sigma_B^2 + \tfrac{m_B}{\lambda}\left(\sigma_{X^-}^2 - m_{X^-}\right)\right] + \\ & + \tfrac{1}{\lambda^2\left(q_0^- + \rho\right)^2}\left[q_0^- m_{X^-}\left(\rho m_{X^-} + 2f\right) - f^2\right], \end{aligned} \tag{13}$$

где $f = \frac{\rho^2}{2}\left(C_B^2 - 1\right)$, а $m_{X^-}$ и $\sigma_{X^-}^2$ — среднее значение и дисперсия длины очереди для вложенной ЦМ, которую можно обнаружить, например, в [23], $m_B^{(n)}$ и $C_B$ — $n$–й начальный момент и коэффициент вариации распределения времени обслуживания $B(x)$.

В СМО $M|G^k|1$ с фиксированной длиной группы число заявок, поступающих на прибор, равно $K$ с вероятностью единица и не зависит от моментов окончания обслуживания. Поэтому ПЛС $w_G(s)$ ФР $W_G(x)$ имеет вид

$$w_G(s) = \frac{Q^-\left(1-\frac{s}{\lambda}\right)}{\beta(s)}. \tag{14}$$

При поступлении $r$-й заявки для завершения формирования группы в систему должно поступить ещё $K-r$ заявок пуассоновского потока. Поскольку эти события для $r = \overline{1,K}$ равновероятны, то остаточное время формирования группы распределено по закону Эрланга с параметрами $(K-r, \lambda)$, т.е.

$$W_C(x) = \frac{1}{K}\sum_{r=1}^{K} E_{K-r}(x), \tag{15}$$

откуда соответствующее ПЛС представляется в виде

$$w_C(s) = \frac{1}{K}\sum_{r=1}^{K}\left(\frac{\lambda}{\lambda+s}\right)^{K-r}. \tag{16}$$

К сожалению рассуждения, приводящие к выражению (10), для СМО $M|G^k|1$ несправедливы, так как в отличие от СМО $M|G^B|1$ здесь очевидным образом присутствует зависимость между остаточным временем формирования группы и временем ожидания начала обслуживания группы. Действительно, в СМО $M|G^k|1$ при поступлении заявки в систему, занятую обслуживанием, её время ожидания будет являться не суммой двух указанных случайных величин, независимых для СМО $M|G^B|1$, а максимумом из них. Однако учитывая свойство математического ожидания суммы случайных величин, формулы (15) и (16) пригодны для получения среднего значения времени ожидания заявки. Окончательно для СМО $M|G^k|1$ имеем,

$$m_W = \frac{\lambda m_B^{(2)} - m_B(K-1)}{2(K-\rho)} + \frac{1}{\lambda}\sum_{i=1}^{K-1}\frac{1}{1-z_i}, \tag{17}$$

где значения $z_i, i = \overline{1, K-1}$ могут быть вычислены по формуле (6). Задача вычисления дисперсии $\sigma_W^2$ до сих пор остаётся не решённой.

Учитывая полученные в данном разделе результаты, табл. 1 относительно виртуального времени ожидания можно переписать следующим образом (табл. 3).

Таблица 3

**Дополнение к табл. 1**

| Характеристика | СМО | |
|---|---|---|
| | $M\|G^B\|1$ | $M\|G^k\|1$ |
| ПЛС ФР времени ожидания начала обслуживания | (11) | (14), (16)* |
| среднее значение | (12) | (17) |
| дисперсия | (13) | – |

∗ для СМО $M|G^k|1$ ПЛС ФР $W(x)$ найдено отдельно для остаточного времени формирования и времени ожидания группы заявок.

## 5. Заключение

В статье выполнен обзор систем типа $M|G|1$ с групповым обслуживанием заявок, исследованы и уточнены методы анализа времени ожидания начала обслуживания в СМО с переменной и фиксированной длиной группы, найдены аналитические выражения для среднего значения и дисперсии.

Полученные результаты являются важной частью исследования протокола управления потоковой передачей и позволяют проводить анализ показателей качества его обслуживания. С помощью данного анализа может быть решена задача выбора оптимальных параметров протокола для обеспечения надлежащего функционирования приложений реального времени.

## Литература


1. *Самуйлов К. Е., Першаков Н. В., Гудкова И. А.* Построение и анализ моделей системы с групповым обслуживанием заявок // Вестник РУДН. Серия «Математика. Информатика. Физика». — 2007. — Т. 3–4. — С. 45–52.
2. *Chukarin A. V., Pershakov N. V.; Samouylov K. E.* Performance of Sigtran-Based Signaling Links Deployed in Mobile Networks // Proc. of the 9th International Conference on Telecommunications (ConTEL 2007, Zagreb, Croatia). — 2007. — Pp. 163–166.
3. *Bailey N. T. J.* On Queueing Processes with Bulk Service // Journal of the Royal Statistical Society, Ser. B. — 1954. — Vol. 16, No 1. — Pp. 80–87.
4. *Downton F.* Waiting Times in Bulk Service Queues // Journal of the Royal Statistical Society, Ser. B. — 1955. — Vol. 17, No 2. — Pp. 256–261.
5. *Chaudhry M. L., Templeton J. G. C.* A First Course in Bulk Queues. — New York: Wiley, 1983.
6. *Rupert G., Miller J.* A Contribution to the Theory of Bulk Queues // Journal of the Royal Statistical Society. Ser. B. — 1959. — Vol. 21, No 2. — Pp. 320–337.
7. *Jaiswal N.* A Bulk-Service Queueing Problem with Variable Capacity // Journal of the Royal Statistical Society. Ser. B. — 1961. — Vol. 23, No 1. — Pp. 143–148.
8. *Bhat U. N.* Imbedded Markov Chain Analysis of Single-Server Bulk Queues // Journal of the Australian Mathematical Society. — 1964. — Vol. 4, No 2. — Pp. 244–263.
9. *Finch P. D.* On the Transient Behavior of a Queueing System with Bulk Service and Finite Capacity // Annals of Mathematical Statistics. — 1962. — Vol. 33, No 3. — Pp. 973–985.
10. *Goyal J. K.* Queues with Hyper-Poisson Arrivals and Bulk Exponential Service // Metrika. — 1967. — Vol. 11, No 1. — Pp. 157–167.

11. *Neuts M. F.* A General Class of Bulk Queues with Poisson Input // Annals of Mathematical Statistics. — 1967. — Vol. 38, No 3. — Pp. 759–770.
12. *Medhi J.* Waiting Time Distribution in a Poisson Queue with a General Bulk Service Rule // Management Science. — 1975. — Vol. 21, No 7. — Pp. 777–782.
13. *Cohen J.* The Single Server Queue. — Amsterdam: North Holland, 1969.
14. *Саати Т. Л.* Элементы теории массового обслуживания и ее приложения / пер. с англ. под ред. И. Н. Коваленко. — М.: Советское Радио, 1971.
15. *Janssen A. J. E. M., van Leeuwaarden J. S. H.* Analytic Computation Schemes for the Discrete-Time Bulk Service Queue // Queueing Systems. — 2005. — Vol. 50, No 2–3. — Pp. 141–163.
16. *Dshalalow J. H., Tadj L.* A Queueing System with a Fixed Accumulation Level, Random Server Capacity and Capacity Dependent Service Time // Mathematics and Mathematical Sciences. — 1992. — Vol. 15, No 1. — Pp. 189–194.
17. *Tadj L.* Explicit Solution of a Quorum Queueing System // Stochastic Analysis and Applications. — 2003. — Vol. 21, No 3. — Pp. 703–717.
18. *Lee H., Lee S., Chae K.* A Fixed-Size Batch Service Queue with Vacations // Applied Mathematics and Stochastic Analysis. — 1996. — Vol. 9, No 2. — Pp. 205–219.
19. *Цициашвили Г. Ш.* Коммутационные эффекты в системах с групповым и приоритетным обслуживанием. — Владивосток: ДВО АН СССР, 1991.
20. *Цициашвили Г. Ш.* Асимптотический анализ системы $G|G|1|\Phi$ с групповым обслуживанием. — Владивосток: Дальнаука, 1996.
21. *Бочаров П. П., Аль-Натор С. В.* Анализ однолинейной системы конечной емкости с марковским групповым обслуживанием // Вестник РУДН. — 1996. — № 1.
22. *Чаплыгин В. В.* Система массового обслуживания $G|BMSP|1|r$ // Информационные процессы. — 2003. — Т. 3, № 2. — С. 97–108.
23. *Denteneer D., Janssen A. J. E. M., van Leeuwaarden J. S. H.* Moment Inequalities for the Discrete-Time Bulk Service Queue // Mathematical Methods of Operations Research. — 2005. — Vol. 61, No 1. — Pp. 85–108.

UDC 621.39

# $M|G|1$ Queues with Batch Service and its Application to the Stream Control Transmission Protocol Performance Analysis. Part I

**N. V. Pershakov, K. E. Samouylov**

*Telecommunication Systems Department*
*Peoples' Friendship University of Russia*
*6, Miklukho–Maklaya str., Moscow, Russia, 117198*

In this paper we present a survey of $M|G|1$ batch service queues which was prepared under performance evaluation of the Stream Control Transmission Protocol (SCTP). The survey is organized in chronological order of the known publications and contains available until now results for probability measures of queues in which jobs are served in batches of fixed and variable size. Some probability measures of virtual waiting time are obtained for SCTP protocol performance analysis.