

Управление пакетным трафиком в телематических устройствах с ограниченным буфером при использовании абсолютного приоритета и вероятностного выталкивающего механизма.

О.И. Заяц, В.С. Заборовский, В.А. Мулюха, А.С. Вербенко

Введение

Реальные потоки данных в современных компьютерных сетях весьма сложны по своей структуре. На практике пакетный трафик с достаточной степенью точности можно считать суперпозицией нескольких независимых однородных потоков данных. Эти составляющие трафика могут различаться по многим признакам: скоростью передачи через каналы связи, пропускной способностью используемых виртуальных соединений, алгоритмами обработки в узлах коммутации, доступным объемом буферной памяти, допустимым уровнем потерь, способом защиты информации, экономическими показателями трафика и т.п.

Применяемые в инженерной практике сетевые технологии чрезвычайно разнообразны, причем сильно опережают в своем развитии уровень их теоретического обоснования и разработку адекватных аналитических моделей. Стохастический характер процессов функционирования компьютерных сетей и структура формируемых виртуальных соединений предопределяет целесообразность использования для их анализа методов и моделей теории массового обслуживания (ТМО). Однако анализ сетевых процессов на основе однопотоковых и однофазных моделей систем массового обслуживания (СМО), характерный для раннего этапа теоретического анализа сетевых взаимодействий, в настоящее время уже не отражает должным образом особенностей современных компьютерных сетей. Упомянутые выше модели позволяют получить лишь самые общие качественные характеристики процессов сетевого взаимодействия, но не могут использоваться для оценки пропускной способности и задержек в каналах связи при передаче мультимедиа трафика, образованного данными от различных источников информации. В этих условиях только многопоточковые модели СМО позволяют точнее описать сложные процессы передачи трафика в современных компьютерных сетях. При этом повышается не только детальность, точность и надежность результатов, но и открываются возможности решения принципиально новых задач, важнейшими из которых являются задачи управления.

В настоящей статье изучается простейшая многопоточковая модель, учитывающая два потока данных. В этой модели механизм управления основан на использовании, во-первых, абсолютного приоритета в обслуживании пакетного трафика, сформированного из команд управления, и, во-вторых, вероятностного выталкивающего механизма, которым снабжена буферная память телематических устройств управления. В результате пакеты приоритетного типа с некоторой вероятностью $0 \leq \alpha \leq 1$ способны выталкивать из буфера телематического устройства управления неприоритетные пакеты. С помощью выбора величины α появляется возможность адаптивной настройки алгоритма управления трафиком, что позволяет эффективно перераспределять доступную для различных виртуальных соединений пропускную способность сети, управляя тем самым взаимной пропускной способностью приоритетных и неприоритетных потоков данных.

Заметим, что в классической теории СМО хорошо изучены лишь системы с детерминированным выталкиванием ($\alpha=1$) и без выталкивания ($\alpha=0$). Произвольные значения α изучались только в системах с относительным приоритетом, который менее употребителен для рассматриваемого класса задач управления. Полученные результаты показывают, что использование параметра α в алгоритме управления является чрезвычайно простым и эффективным способом оптимизации сетевых взаимодействий. Модели, основанные на использовании СМО с абсолютным приоритетом и вероятностным выталкивающим механизмом являются достаточно точными и позволяют

получить на их основе все необходимые для оценки эффективности управления характеристики каналов связи, а затем в процессе космического эксперимента идентифицировать параметры используемых моделей.

1. Функциональная схема организации информационного взаимодействия.

Наглядным примером широких возможностей использования такого механизма, является задача управления удаленным робототехническим объектом, результаты функционирования которого, наряду с данными телеметрии и видеопотоком, передаются с помощью широкополосных сетей общего доступа. В этом случае команды управления передаются с использованием транспортного протокола TCP, а результаты наблюдения представляют собой поток видеоданных (см. рис.1). В данном космическом эксперименте пропускная способность сети составила: для приоритетных TCP-пакетов управления и телеметрии –100Кбит/с, а для фонового видеопотока UDP пакетов – 1,2Мбит/с, так что отношение коэффициентов загрузки [1,5] для двух указанных потоков составило 1 к 12.

В рассматриваемом примере использование приоритетного механизма управления в комбинации с выталкивающим механизмом, позволяют сбалансировать такие показатели функционирования сети, как вероятность потери команд управления и качество получаемой картины видеонаблюдения для различных состояний сетевой среды. Параметр α может варьироваться и с целью изменения задержек пакетов в контуре обратной связи системы управления роботом.

Данная задача важна для дальнейшего развития интерактивного управления удаленными динамическими объектами, в случае когда сетевая среда является составной частью контура обратной связи, где уровень потерь и задержек является важным параметром, характеризующим эффективность системы управления в целом.

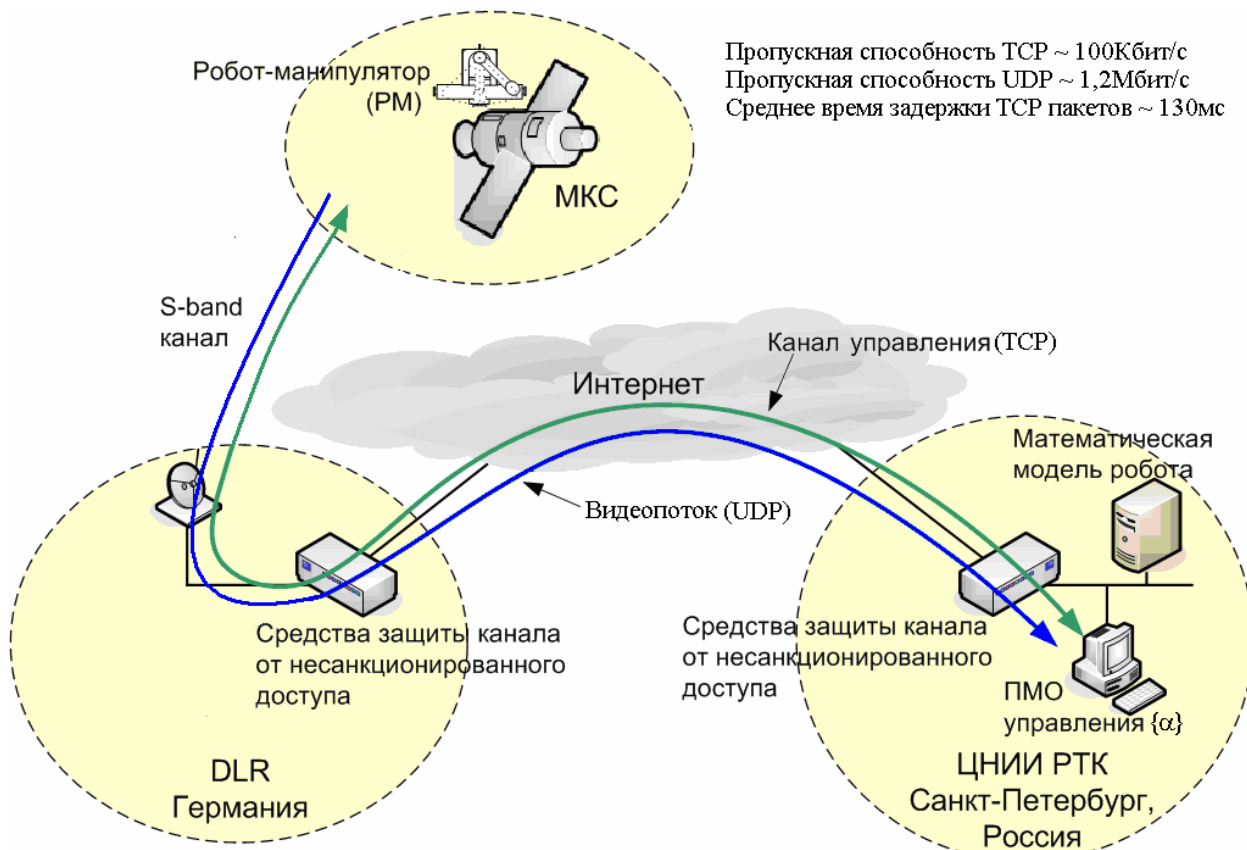


Рис. 1 Схема космического эксперимента «Контур»

2. Математическая постановка задачи

Рассмотрим одноканальную СМО ограниченной емкости k ($1 < k < \infty$), на вход которой поступают два независимых простейших потока заявок (пакетов) интенсивностью λ_1 и λ_2 . Допустим, что длительность обслуживания (обработки любого пакета) распределена по одному и тому же показательному закону. Все эти длительности независимы в совокупности. Интенсивность обслуживания обозначается через μ .

Пакетам первого типа предоставлен абсолютный приоритет в обслуживании. Пока все приоритетные требования не будут обработаны, заявки второго типа остаются в очереди. Вновь поступившие приоритетные требования прерывают обработку неприоритетных заявок и вытесняют их в накопитель (если есть свободные места ожидания), либо за пределы системы (если свободных мест нет). Снятые с обслуживания пакеты присоединяются к очереди неприоритетных требований.

Накопитель является общим, свободные места ожидания полностью доступны для любого подошедшего требования. Динамически формируются две отдельные очереди пакетов, их суммарная длина ограничена емкостью накопителя ($k-1$). В отличие от типовых приоритетных СМО [1,2] рассматриваемая система снабжена вероятностным выталкивающим механизмом. Приоритетный пакет, заставший все места ожидания занятыми в момент обработки другого приоритетного пакета, с заданной вероятностью α вытесняет из накопителя один из менее приоритетных пакетов и занимает его место. Вытесненный пакет теряется.

Схема описанной СМО приведена на рис.2. Через $a_i(\tau)$ и $b_i(x)$ обозначены законы распределения, соответственно, интервала между требованиями и времени обслуживания для i -го потока пакетов. В отсутствие выталкивающего механизма ($\alpha = 0$) и при детерминированном выталкивании ($\alpha = 1$) аналогичные системы изучал Г.П.Башарин [3]. Сама концепция вероятностного выталкивающего механизма применительно к сетевым и телекоммуникационным проблемам предложена в [4], где этот механизм сочетался, правда, с относительным, а не абсолютным, как у нас, приоритетом.

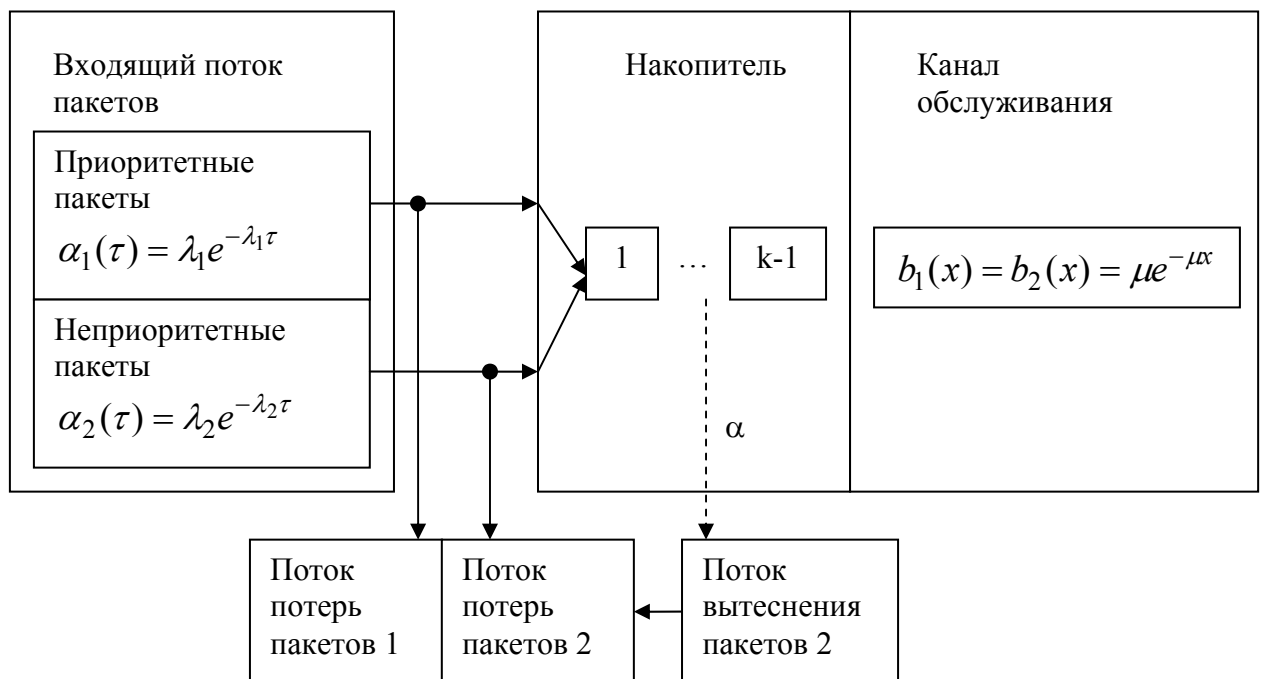


Рис 2. Схема приоритетной СМО класса $\bar{M}_2 / M / 1 / k / f_2^1$

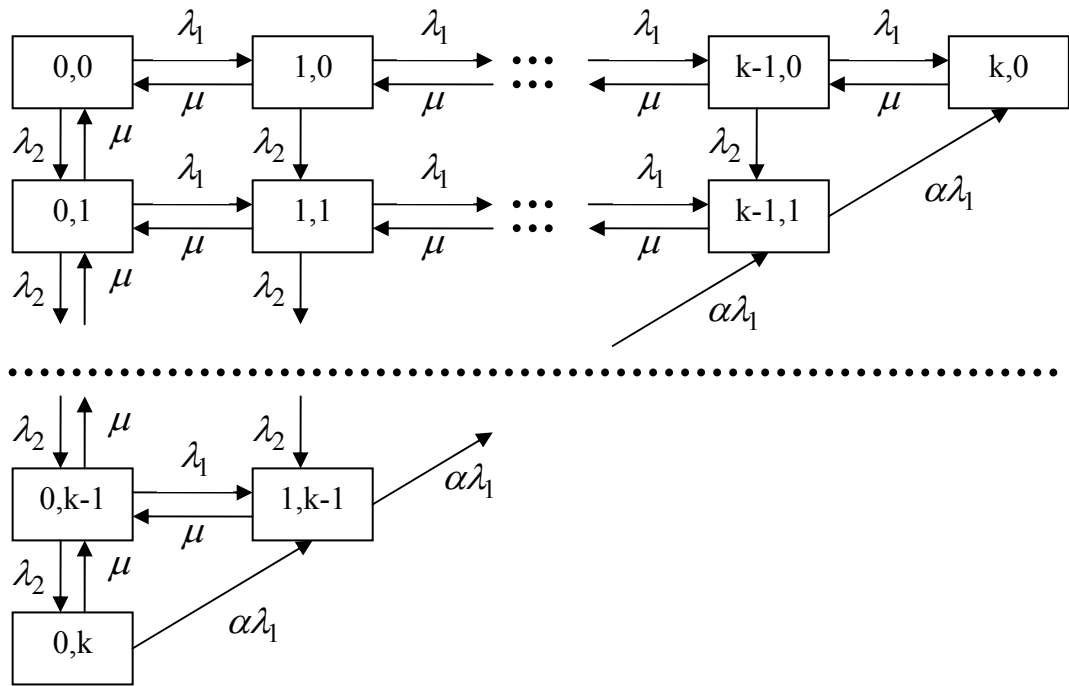


Рис.3. Размеченный граф состояний системы $\bar{M}_2/M/1/k/f_2^1$

Суммарный входящий поток СМО, изображенный на рис. 2, будет простейшим с интенсивностью $\lambda = \lambda_1 + \lambda_2$. Если отвлечься от различий между типами требований и отслеживать только общее число пакетов в системе, то такая упрощенная однопоточковая модель имела бы сокращенное обозначение по Д.Кендаллу $M/M/1/k$ [5]. Специальная модификация системы обозначений Кендалла, предназначенная для приоритетных СМО, принадлежит Г.П.Башарину [3]. В модифицированной системе общая структура обозначения и смысл отдельных его позиций сохраняется, однако в каждой позиции используется векторная символика. Кроме того предусмотрен дополнительный символ f_i^j , где i указывает на тип приоритета (0 – без приоритета, 1 – относительный, 2 – абсолютный), а j – на тип выталкивающего механизма (0 – без выталкивания, 2 – детерминированное выталкивание).

В оригинальной системе [3] значение $j=1$ не было задействовано. На наш взгляд, его целесообразно использовать для обозначения вероятностного выталкивающего механизма, как промежуточного между вариантами $j=0$ и $j=2$. С этим дополнением по правилам [3] СМО, изображенная на рис. 2, должна быть отнесена к классу $\bar{M}_2/M/1/k/f_2^1$.

История исследования одноканальных двухпоточковых приоритетных систем насчитывает уже более полувека, однако, насколько нам известно, имеется лишь одна работа [4], где изучался бы вероятностный выталкивающий механизм (в сочетании с относительным приоритетом для СМО класса $\bar{M}_2/M/1/k/f_1^1$). Вместе с тем, для типовых моделей выталкивающего механизма ($j=0$ и $j=2$) задача, в основном, решена.

Значительный вклад в развитие этой теории внес Г.П.Башарин и его ученики. Типовые модели $\bar{M}_2/\bar{M}_2/1/k/f_2^0$, а также $\bar{M}_2/\bar{M}_2/1/k/f_2^2$ изучены им в [6]. Вторая часть указанной работы посвящена анализу систем $\bar{M}_2/\bar{M}_2/1/\bar{k}/f_2^0$ и $\bar{M}_2/\bar{M}_2/1/\bar{k}/f_2^2$ с отдельными очередями. Такая же структура системы, но с конечным числом источников нагрузки l , обозначаемая как $\bar{M}_2/M/1/k/l/f_2^2$, разобрана в работе [7]. Там же

рассмотрена еще одна более сложная модель $\bar{M}_2/M/1/\bar{k}/\bar{l}/f_2^2$ с отдельными очередями и разделением источников нагрузки по типам требований.

Задачи по исследованию приоритетных СМО, подобных рассмотренным в работах [3,6,7] первоначально возникли в телекоммуникационных и информационных системах, связанных, в частности, с анализом реальных дисциплин диспетчеризации в управляющих ЭВМ. В последние годы подобного рода модели СМО, а также различные их обобщения широко используются при теоретическом анализе реальных интернет-систем.

Интересная задача решена А.Бонди [8]. В этой работе типовые модели с общей очередью $\bar{M}_2/M/1/k/f_2^0$ и отдельными очередями $\bar{M}_2/M/1/\bar{k}/f_2^0$ сравниваются с частично разделенными очередями при сохранении некоторой общей части накопителя. Решена задача оптимального распределения емкости накопителя между его полностью и частично доступными частями (по критерию максимума средней суммарной очереди). В работе [9] рассмотрен пороговый выталкивающий механизм, при котором выталкивание разрешено лишь при достаточно большой длине очереди неприоритетных пакетов.

Как показано в [4], вероятностный выталкивающий механизм является более удобным и эффективным по сравнению с другими рассмотренными в литературе математическими моделями выталкивания. Он адекватно описывает реальные процессы сетевого трафика и при этом достаточно прост с математической точки зрения. Все это делает данное направление исследований весьма актуальным и практически значимым.

3. Вычисление производящей функции

Будем характеризовать состояние системы двумерным процессом $\vec{N}(t) = \{N_1(t), N_2(t)\}$, где $N_i(t)$ есть число требований i -го типа в системе в момент t . При сделанных допущениях, процесс $\vec{N}(t)$ будет марковским и развиваться в фазовом пространстве

$$\Omega = \{(n_1, n_2) : n_1 \geq 0, n_2 \geq 0, n_1 + n_2 \leq k\}. \quad (3.1)$$

Размеченный граф состояний процесса $\vec{N}(t)$ представлен на рис. 3.

По теореме Маркова процесс \vec{N} эргодичен. Далее ограничимся анализом установившегося режима. Финальные вероятности состояний обозначим через p_{ij} ($i = \overline{0, k}; j = \overline{0, k-i}$).

Составляя с помощью рис. 3 по обычным правилам [5] систему уравнений Колмогорова, получим

$$\begin{aligned} & -[\lambda_1(1 - \delta_{j, k-i}) + \alpha\lambda_1(1 - \delta_{j, k})\delta_{j, k-i} + \lambda_2(1 - \delta_{j, k-i}) + \mu(1 - \delta_{i, 0}\delta_{j, 0})]p_{ij} + \\ & + \mu p_{i+1, j} + \mu\delta_{i, 0}p_{i, j+1} + \lambda_2 p_{i, j-1} + \lambda_1 p_{i-1, j} + \alpha\lambda_1 \delta_{j, k-i} p_{i-1, j+1} = 0, \end{aligned} \quad (3.2)$$

$(i = \overline{0, k}; j = \overline{0, k-i}),$

где $\delta_{i, j}$ обозначает дельта-символ Кронекера. Система (3.2) должна решаться совместно с условием нормировки

$$\sum_{i=0}^k \sum_{j=0}^{k-i} p_{ij} = 1. \quad (3.3)$$

При реальных (достаточно больших) k система (3.2) – (3.3) плохо обусловлена, и ее непосредственное численное решение приводит к большим вычислительным погрешностям. Многие авторы, в частности, Г.П.Башарин обходят эту трудность, применяя метод рекуррентных соотношений [6,7]. В этом методе задача сводится к решению некоторой вспомогательной системы линейных алгебраических уравнений порядка $\frac{1}{2}k(k+1)$ с треугольной матрицей.

В настоящей статье используется альтернативный метод производящих функций [4] в его классическом варианте, предложенном еще полвека назад Х.Уайтом, Л.Кристи и Ф.Стефаном применительно к системе с ожиданием $\bar{M}_2/M/1/f_2$ [10,11]. В ряде задач, как, например, для системы $\bar{M}_2/M/1/f_2$ или системы $\bar{M}_2/M/1/k/f_1^2$ [4] данный метод позволяет получить явное аналитическое решение. Далее в п.4 будет показано, что аналитическое решение получается и для СМО класса $\bar{M}_2/M/1/k/f_2$, в которую переходит наша система при $\alpha = 1$. Отметим, что выражение для вероятности потери для системы $\bar{M}_2/M/1/k/f_1^2$, найденное в [4] методом производящих функций, оказывается более простым, чем аналогичное выражение, полученное в [12,13] методом рекуррентных соотношений.

В случае $\alpha < 1$ метод [4], вообще говоря, не приводит к замкнутому аналитическому решению, однако и в этом случае он дает ощутимый выигрыш по сравнению с методом рекуррентных соотношений. Так же, как и в последнем, задача сводится к решению системы линейных уравнений с треугольной матрицей, но порядок системы уменьшается с $\frac{1}{2}k(k+1)$ до $(k+1)$. Это снижает сложность вычислений и при тех же требованиях к их точности позволяет охватить более широкий диапазон изменения емкости буфера k .

Введем производящую функцию финальных вероятностей в виде

$$G(u, v) = \sum_{i=0}^k \sum_{j=0}^{k-i} p_{ij} u^i v^j. \quad (3.4)$$

Условие нормировки (3.3) переписывается в виде

$$G(1, 1) = 1. \quad (3.5)$$

Умножая обе части равенства (3.2) на $u^i v^j$ и суммируя по всем допустимым значениям (i, j) , получаем

$$\begin{aligned} G(u, v) = & \{ \mu(u-v)G(0, v) + \mu u(v-1)G(0, 0) + [\alpha \lambda_1(u-v) + \lambda_1(1-u)v + \\ & + \lambda_2(1-v)v] u \sum_{i=0}^k p_{i, k-i} u^i v^{k-i} + \alpha \lambda_1 u^{k+1} (v-u) p_{k, 0} \} \{ [\lambda_1 u(1-u) + \lambda_2 u(1-v) + \\ & + \mu(u-1)] v \}^{-1} \end{aligned} \quad (3.6)$$

Полученное представление позволяет выразить функцию G через вероятности граничных состояний $p_{0,j}$ и $p_{i, k-i}$ размеченного графа рис.3. С помощью особого приема, разработанного в [4], удастся исключить и вероятности $p_{0,j}$. Если затем воспользоваться стандартными рассуждениями [10,11] основанными на соображениях аналитичности функции G , то задача сведется к решению некоторой системы линейных уравнений относительно неизвестных $p_{i, k-i}$.

Вначале рассмотрим более простую аналитически разрешимую задачу с детерминированным выталкивающим механизмом.

4. Система $\bar{M}_2/M/1/k/f_2^2$

Аналитическая разрешимость задачи вычисления вероятности потери при $\alpha = 1$ в случае относительного приоритета была показана в [4]. Этот результат сохраняет силу и при абсолютном приоритете.

Полагая $\alpha = 1$, после деления числителя и знаменателя (3.6) на μ находим

$$\begin{aligned} G(u, v) = & [(u-v)G(0, v) + u(v-1)G(0, 0) + u(1-v)(\rho_2 v + \rho_1 u) \cdot \\ & \cdot \sum_{i=0}^k p_{i, k-i} u^i v^{k-i} + \rho_1 (v-u) u^{k+1} \rho_{k, 0}] \cdot \{ [\rho_1 u(1-u) + \rho_2 u(1-v) + (u-1)] v \}^{-1}, \end{aligned} \quad (4.1)$$

где $\rho_i = \lambda_i / \mu$ означает коэффициент использования системы по i -му типу требований.

Получим распределение числа приоритетных требований N_1 . Положим $q_n = P\{N_1 = n\}, (n = \overline{0, k})$ и выразим производящую функцию q_n через G в виде

$$G_1(u) = \sum_{n=0}^k q_n u^n = G(u, 1). \quad (4.2)$$

Подставляя в (4.1) аргумент $v = 1$, сокращая на $(u - 1)$ и используя тождества

$$G(0, 1) = \sum_{j=0}^k p_{0,j} = q_0, \quad p_{k,0} = q_k, \quad (4.3)$$

находим

$$G_1(u) = \frac{q_0 - \rho_1 q_k u^{k+1}}{1 - \rho_1 u} \quad (4.4)$$

Функция (4.4) имеет полюс первого порядка при $u = \rho_1^{-1}$, между тем согласно (4.2) она должна быть полиномом степени k . Чтобы устранить эту особенность, приравняем к нулю вычет G_1 в указанном полюсе

$$\operatorname{Res}_{u=\rho_1^{-1}} G_1(u) = 0, \quad (4.5)$$

откуда выводим

$$q_k = q_0 \rho_1^k. \quad (4.6)$$

Затем из условия нормировки $G_1(1) = 1$ вычисляем q_0

$$q_0 = \frac{1 - \rho_1}{1 - \rho_1^{k+1}} \quad (4.7)$$

и окончательно

$$G_1(u) = \frac{1 - \rho_1}{1 - \rho_1^{k+1}} \cdot \frac{1 - (\rho_1 u)^{k+1}}{1 - \rho_1 u}. \quad (4.8)$$

Производящая функция (4.8) отвечает усеченному геометрическому распределению

$$q_n = \frac{(1 - \rho_1) \rho_1^n}{(1 - \rho_1^{k+1})}, \quad (n = \overline{0, k}). \quad (4.9)$$

Зная его, нетрудно найти все технические характеристики СМО, касающиеся приоритетных заявок. Основной из них в сетевых задачах является вероятность потери приоритетного требования $P_{nom}^{(1)}$. Для более детального анализа приоритетного трафика представляют интерес также среднее число приоритетных требований в системе $\bar{n}^{(1)}$, средняя длина очереди приоритетных требований $\bar{n}_{оч}^{(1)}$, вероятность ожидания $P_{ож}^{(1)}$ (вероятность того, что вновь подошедшие приоритетное требование встанет в очередь), среднее число каналов, занятых этими требованиями $\bar{n}_{зк}^{(1)}$.

Опуская несложные вычисления, будем иметь

$$\begin{aligned} P_{nom}^{(1)} = q_k &= \frac{(1 - \rho_1) \rho_1^k}{(1 - \rho_1^{k+1})}, \quad P_{ож}^{(1)} = \sum_{n=1}^{k-1} q_n = \frac{(1 - \rho_1^{k-1}) \rho_1}{(1 - \rho_1^{k+1})}, \\ \bar{n}_{оч}^{(1)} &= \sum_{n=1}^k (n-1) q_n = \frac{\rho_1^2}{1 - \rho_1} - \frac{(1 - \rho_1) \rho_1^{k+1}}{(1 - \rho_1^{k+1})}, \quad n_{зк}^{(1)} = 1 - q_0 = \frac{(1 - \rho_1^k) \rho_1}{(1 - \rho_1^{k+1})}, \\ \bar{n}^{(1)} &= \bar{n}_{оч}^{(1)} + \bar{n}_{зк}^{(1)} = \frac{\rho_1}{1 - \rho_1} - \frac{(k+1) \rho_1^{k+1}}{(1 - \rho_1^{k+1})}. \end{aligned} \quad (4.10)$$

Приведенные характеристики полностью идентичны аналогичным показателям для однопоточковой системы $M/M/1/k$ с коэффициентом использования ρ_1 [5]. Это вполне

естественно: при такой организации СМО приоритетные пакеты как бы «не замечают» остальные требования.

Теперь исследуем общее число пакетов в системе $N = N_1 + N_2$. Определим вероятности $r_n = P\{N = n\}$ и введем их производящую функцию

$$G_{\Sigma}(u) = \sum_{n=0}^k r_n u^n = G(u, u). \quad (4.11)$$

Полагая $v = u$ в (4.1), сокращая на $(u-1)$ и используя вместо (4.3) тождества, касающиеся вероятностей r_n ,

$$G(0,0) = p_{0,0} = r_0, \quad \sum_{i=0}^k p_{i,k-i} = r_k, \quad (4.12)$$

получим для G_{Σ} представление, аналогичное (4.4), в котором числовые параметры q_1, ρ_1, q_k следует заменить, соответственно, на $r_0, \rho = \rho_1 + \rho_2, r_k$.

Функция G_{Σ} , подобно функции (4.4) будет иметь особенность в полюсе $u = \rho^{-1}$. Устраняя ее так, как это было показано выше, и нормируя результат на единицу при $u = 1$, находим

$$G_{\Sigma}(u) = \frac{1-\rho}{1-\rho^{k+1}} \cdot \frac{1-(\rho u)^{k+1}}{1-\rho u}. \quad (4.13)$$

Следовательно, общее число требований также распределено по усеченному геометрическому закону, но с параметром $\rho = \rho_1 + \rho_2$, отвечающим полной загрузке

$$r_n = \frac{(1-\rho)\rho^n}{(1-\rho^{k+1})}, \quad (n = \overline{0, k}). \quad (4.14)$$

Легко видеть, что среднее число всех требований в системе \bar{n} , средняя длина общей очереди $\bar{n}_{оч}$, среднее число каналов, занятых всеми требованиями $\bar{n}_{зк}$ и вероятность ожидания неприоритетного требования $P_{ож}^{(2)}$ находятся по формулам, аналогичным (4.10), где ρ_1 должно быть заменено на ρ . Для требований низшего приоритетного класса отсюда получаем

$$\bar{n}^{(2)} = \bar{n} - \bar{n}^{(1)}, \bar{n}_{оч}^{(2)} = \bar{n}_{оч} - \bar{n}_{оч}^{(1)}, \bar{n}_{зк}^{(2)} = \bar{n}_{зк} - \bar{n}_{зк}^{(1)}. \quad (4.15)$$

Как видим, неприоритетные требования подстраиваются под приоритетные, занимая оставшиеся свободными места на обслуживании и в накопителе таким образом, чтобы суммарные показатели соответствовали однопоточковой СМО $M/M/1/k$ с полной нагрузкой ρ .

Несколько сложнее находится вероятность потери неприоритетного требования $P_{пот}^{(2)}$. Представим ее в виде суммы $P_{пот}^{(2,0)} + P_{пот}^{(2,1)}$, где $P_{пот}^{(2,0)} = r_k$ означает вероятность потери на входе системы, а $P_{пот}^{(2,1)}$ есть вероятность потери из накопителя путем выталкивания. Введем также вероятность вытеснения приоритетным требованием какого-либо неприоритетного

$$P_{выт} = r_k - q_k = \frac{(1-\rho)\rho^k}{(1-\rho^{k+1})} - \frac{(1-\rho_1)\rho_1^k}{(1-\rho_1^{k+1})}. \quad (4.16)$$

Фиксируем некоторый достаточно большой интервал времени $[0, T]$. В среднем в этом интервале за счет вытеснения в накопитель дополнительно попадут $\lambda_1 TP_{выт}$ требований. С другой стороны, за тот же период в среднем $\lambda_2 TP_{пот}^{(2,1)}$ неприоритетны требований будут вытеснены из накопителя. В силу установившегося режима работы системы и законов сохранения [14] указанные величины должны совпадать. Вычисляя отсюда $P_{пот}^{(2,1)}$, находим

$$P_{nom}^{(2,1)} = \frac{(1-\rho)\rho^k}{(1-\rho^{k+1})} \left(1 + \frac{\rho_1}{\rho_2}\right) - \left(\frac{\rho_1}{\rho_2}\right) \frac{(1-\rho_1)\rho_1^k}{(1-\rho_1^{k+1})}. \quad (4.17)$$

5. Система $\vec{M}_2 / M / 1 / k / f_2^1$

В случае $\alpha \neq 1$ уже не удается вывести простые аналитические формулы для характеристик СМО, однако все эти характеристики можно выразить через вероятности

$$p_i = p_{k-i}, \quad (i = \overline{0, k}), \quad (5.1)$$

описывающие различные состояния СМО с переполненным накопителем. Сами вероятности p_i определяются решением некоторой системы линейных алгебраических уравнений.

Покажем вначале, что через p_i выражается распределение числа приоритетных требований. Подставляя $v = 1$ в (3.6) и учитывая (4.3), получаем вместо (4.4) выражение

$$G_1(u) = \frac{q_0 - (1-\alpha)\rho_1 \sum_{i=0}^{k-1} p_{k-i} u^{i+1} - \rho_1 q_k u^{k+1}}{1 - \rho_1 u}. \quad (5.2)$$

Условие аналитичности (4.5) сохраняет силу, но вместо (4.6) приводит к уравнению

$$q_0 - (1-\alpha) \sum_{i=0}^{k-1} p_{k-i} \rho_1^{-i} - q_k \rho_1^{-k} = 0. \quad (5.3)$$

Вычтем левую часть (5.3) из числителя (5.2), сократим на $(1 - \rho_1 u)$ и приравняем коэффициенты при одинаковых степенях u в левой и правой части равенства. В результате приходим к уравнениям

$$q_{k-j} = (1-\alpha) \sum_{i=1}^j p_i \rho_1^{i-j} + q_k \rho_1^{-j}, \quad (j = \overline{1, k}). \quad (5.4)$$

При $j = 0$ уравнение (5.4) также выполняется, но превращается в тривиальное тождество.

Далее просуммируем, используя (5.4), все вероятности q_j по j от 0 до $k-1$, заменим эту сумму на $1 - q_k$ и решим полученное линейное уравнение относительно q_k . Это дает следующее выражение для q_k

$$q_k = \frac{(1-\rho_1)\rho_1^k}{(1-\rho_1^{k+1})} \left[1 - (1-\alpha) \sum_{i=1}^k p_i \frac{(1-\rho_1^{k-i+1})}{(1-\rho_1)\rho_1^{k-i}}\right]. \quad (5.5)$$

Естественно при $\alpha = 1$, оно переходит в (4.9). Остальные вероятности q_j выражаются по формуле (5.4), которая также сводится к (4.9) при $\alpha = 1$.

Что касается распределения общего числа требований N , то подставляя в (3.6) значение $v = u$, приводя подобные члены и повторяя рассуждения п.4, приходим к прежнему распределению общего числа требований (4.14), которое, таким образом, оказывается инвариантным относительно вероятности вытеснения α . Это вполне естественно, так как вытеснение не меняет заполненность накопителя, а лишь увеличивает представительство в нем приоритетных требований.

Теперь перейдем к получению выражений через p_i произвольных вероятностей p_{ij} . Для этого перепишем (3.6) в виде

$$G(u, v) = \frac{R_{10}(u, v)}{R_{20}(u, v)}, \quad (5.6)$$

где R_{10} и R_{20} – полиномы, степень которых относительно u равняется, соответственно, $k + 2$ и 2. Их выражения таковы:

$$R_{10} = (u-v) \cdot G(0, v) + u(v-1) \cdot G(0, 0) + \alpha \rho_1 (v-u) u^{k+1} p_{k,0} +$$

$$+ [\alpha \rho_1 (u-v) + \rho_1 (1-u)v + \rho_2 (1-v)v] \cdot u \sum_{i=0}^k p_{i, k-i} u^i v^{k-i}, \quad (5.7)$$

$$R_{20} = [\rho_1 u(1-u) + \rho_2 u(1-v) + (u-1)]v. \quad (5.8)$$

Знаменатель (5.6) имеет два корня

$$u_{1,2} = \frac{[\rho_1 + \rho_2(1-v) + 1] \mp \sqrt{[\rho_1 + \rho_2(1-v) + 1]^2 - 4\rho_1}}{2\rho_1}. \quad (5.9)$$

По формулам Виета выполняются тождества

$$u_1 + u_2 = [\rho_1 + \rho_2(1-v) + 1]\rho_1^{-1}, u_1 u_2 = \rho_1^{-1}. \quad (5.10)$$

Корни (5.9) являются простыми полюсами функции (5.6).

Производящая функция (3.4), будучи полиномом, должна быть аналитична по u во всей плоскости. Поэтому необходимо приравнять к нулю ее вычеты в обоих полюсах

$$\operatorname{Res}_{u=u_1} G(u, v) = 0, \quad \operatorname{Res}_{u=u_2} G(u, v) = 0. \quad (5.11)$$

Первое уравнение (5.11) сводится к равенству

$$R_{10}(u_1, v) = 0. \quad (5.12)$$

Вычитая из числителя (5.6) левую часть (5.12), после сокращения на $(u-u_1)$ находим

$$G(u, v) = \frac{R_{11}(u, v)}{R_{21}(u, v)}, \quad (5.13)$$

где полиномы

$$R_{11}(u, v) = \frac{R_{10}(u, v) - R_{10}(u_1, v)}{u - u_1}, R_{21}(u, v) = -\rho_1(u - u_2) \quad (5.14)$$

имеют по u степень на единицу меньшую чем, соответственно, R_{10} и R_{20} .

В представлении (5.13) сохраняется только полюс $u = u_2$. Из второго условия (5.11) заключаем, что

$$R_{11}(u_2, v) = 0. \quad (5.15)$$

Разность $R_{11}(u, v) - R_{11}(u_2, v)$ без остатка делится на $(u - u_2)$, после чего получаем

$$G(u, v) = R_{12}(u, v),$$

где

$$R_{12}(u, v) = -\frac{R_{11}(u, v) - R_{11}(u_2, v)}{\rho v(u - u_2)}. \quad (5.16)$$

Таким образом, производящая функция (3.4) может быть получена из полинома (5.7) применением следующего оператора разделенной разности второго порядка

$$G(u, v) = -\frac{1}{\rho v(u - u_2)} \left[\frac{R_{10}(u, v) - R_{10}(u_1, v)}{(u - u_1)} - \frac{R_{10}(u_2, v) - R_{10}(u_1, v)}{(u_2 - u_1)} \right]. \quad (5.17)$$

Поскольку R_{10} есть полином относительно u , то дело сводится к применению оператора вида (5.17) к целым степеням u . При этом удобно использовать тождество

$$\frac{1}{(u - u_2)} \left(\frac{u^{i+2} - u_1^{i+2}}{u - u_1} - \frac{u_2^{i+2} - u_1^{i+2}}{u_2 - u_1} \right) = \sum_{j=0}^i u^j Q_{i-j+1}, \quad (5.18)$$

где по определению

$$Q_i = \frac{u_2^i - u_1^i}{u_2 - u_1}, (i \geq 0). \quad (5.19)$$

В первоначальном выражении (5.7) присутствовали вероятности состояний вида $(0, j)$ и $(i, k-i)$. После вычисления по u разделенной разности (5.17) в правой части

(5.17) сохраняются лишь вероятности вида (5.1). Если представить производящую функцию (3.4) в виде полинома относительно u , расположенного по возрастающим степеням

$$G(u, v) = \sum_{i=0}^k g_i(v) u^i, \quad (5.20)$$

то коэффициенты g_i будут такими

$$\begin{aligned} g_k &= p_0, g_{k-1} = \rho_1^{-1} p_0 + (v - \alpha) p_1 \\ g_i &= [Q_{k-i+1} - (1 + \varepsilon(1 - v)) Q_{k-i}] p_0 + (1 - \alpha v^{-1}) p_{k-i} v^{k-i} + \\ &+ \sum_{j=i+1}^{k-1} [(1 - \alpha v^{-1}) Q_{j+1-i} + (\alpha - (1 + \varepsilon(1 - v)) Q_{j-i})] p_{k-j} v^{k-j}, (i = \overline{0, k-2}), \end{aligned} \quad (5.21)$$

где $\varepsilon = \rho_2 / \rho_1$ характеризует соотношение интенсивностей.

Используя формулы Виета (5.10), нетрудно показать, что

$$Q_{j-i+1} - (1 + \varepsilon(1 - v)) Q_{j-i} = \rho_1^{-1} (Q_{j-i} - Q_{j-i-1}). \quad (5.22)$$

Это позволяет еще более упростить выражение (5.21)

$$\begin{aligned} g_k &= p_0, g_{k-1} = \rho_1^{-1} p_0 + (v - \alpha) p_1, \\ g_i &= \rho_1^{-1} (Q_i - Q_{i-1}) p_0 + \sum_{j=1}^{i-2} [(\rho_1^{-1} + \alpha) Q_{i-j} - \rho_1^{-1} Q_{i-j-1}] p_j v^j + \\ &+ (\rho_1^{-1} + \alpha) p_{i-1} v^{i-1} - \alpha \sum_{j=0}^{i-2} Q_{i-j} p_{j+1} v^j + (1 - \alpha v^{-1}) p_i v^i, (i = \overline{2, k}). \end{aligned} \quad (5.23)$$

Из определения (3.4) ясно, что функция $g_i(v)$ должны быть полиномами относительно v степени не выше, чем $(k - i)$. Следовательно, задача свелась к получению полиномиального представления для коэффициентов Q_i . Для ее решения применим общий прием, разработанный в статье [4].

Представим корни u_j в комплексной форме

$$u_j = a_j e^{i\varphi_j}, \quad (j = \overline{1, 2}), \quad (5.24)$$

где a_j и φ_j обозначают, соответственно, модуль и аргумент числа. В соответствии с (5.9)-(5.10)

$$a_1 = a_2 = \rho_1^{-1/2}, \varphi_1 = -\varphi_2 = \varphi, \quad (5.25)$$

где φ определяется уравнением

$$\cos(\varphi) = \frac{1}{2} (\rho_1 + \rho_2(1 - v) + 1) \rho_1^{-1/2} = t(v). \quad (5.26)$$

Важно заметить, что функция $t(v)$ линейна по v .

Из определения коэффициентов Q_j следует, что они выражаются через v следующим образом

$$Q_j = \rho_1^{-\frac{(j-1)}{2}} U_{j-1}(t(v)), (j = \overline{1, k+1}), \quad (5.27)$$

где $U_n(x)$ обозначает полином Чебышева второго рода [15]

$$U_n(\cos \varphi) = \frac{\sin[(n+1)\varphi]}{\sin \varphi}. \quad (5.28)$$

Раскладывая правую часть (5.27) по степеням v , получаем

$$Q_j = \sum_{i=0}^{j-1} \frac{(-1)^i \rho_2^i}{i! 2^i \rho_1^{i+j-1/2}} \cdot U_{j-1}^{(i)}(t(0)) v^i. \quad (5.29)$$

Производные полиномов Чебышева второго рода можно выразить в виде [15]

$$U_n^{(m)}(x) = 2^m m! C_{n-m}^{m+1}(x), \quad (5.30)$$

где $C_n^v(x)$ – полином Гегенбауэра порядка n с индексом v . В результате, окончательно находим

$$Q_j = \sum_{i=0}^{j-1} C_{j-i-1}^{i+1}(t_0) \beta^i v^j, \quad (5.31)$$

где

$$\beta = -\rho_2 \rho_1^{-1/2}, t_0 = t(0) = \frac{1}{2}(1 + \rho) \rho_1^{-1/2}. \quad (5.32)$$

Таким образом, в правой части уравнений (5.23) действительно стоят полиномы относительно v . С другой стороны, согласно определению (3.4)

$$g_{k-i}(v) = \sum_{j=0}^i p_{k-i,j} v^j, \quad (i = \overline{0, k}). \quad (5.33)$$

Согласно определению (5.1) вероятностей p_i коэффициент при v^i в выражении (5.33) должен равняться p_i . Если выразить все такие коэффициенты из (5.23) для $1 \leq i \leq k$, то получим систему k линейных уравнений, связывающих $(k+1)$ вероятность (5.1). Коэффициенты при v^0 совпадают тождественно, и недостающее уравнение придется выводить особо.

Опуская громоздкие, но несложные выкладки, приходим к следующим k уравнениям

$$p_i = \frac{\rho_1^{-1}(\rho_1^{-i} - \zeta_{i+1})p_0 + \sum_{j=1}^{i-1} [(1-\alpha)\rho_1^{j-i-1} - \zeta_{i+1,j}]p_j}{\zeta_{i+1,i} - (1-\alpha)\rho_1^{-1}}, \quad (i = \overline{1, k-1}) \quad (5.34)$$

$$p_k = \frac{\rho_2(1-r_k) - \alpha\rho_1(r_k - p_0) + r_0 - \rho_1^{-1}\zeta_k p_0 - \sum_{j=1}^{k-1} \zeta_{k,j} p_j}{(1-\alpha)},$$

где r_n обозначает вероятности (4.14), а числовые коэффициенты ζ_i и $\zeta_{i,j}$ выражаются через полиномы Гегенбауэра по формулам:

$$\zeta_i = \sum_{j=0}^{i-1} [C_{i-j-1}^{j+1}(t_0) - C_{i-j-2}^{j+1}(t_0)] \beta^j, \quad (i = \overline{1, k}) \quad (5.35)$$

$$\zeta_{i,j} = \sum_{s=j}^i \{ \rho_1^{-1} [C_{i-s-1}^{s-j+1}(t_0) - C_{i-s-2}^{s-j+1}(t_0)] - \alpha [C_{i-s}^{s-j+1}(t_0) - C_{i-s-1}^{s-j+1}(t_0)] \} \beta^{s-j}, \quad (5.36)$$

$$(i = \overline{2, k}; j = \overline{0, i-1})$$

Последнее уравнение системы (5.35) получается из очевидного равенства

$$q_k = p_0, \quad (5.37)$$

с учетом которого на основании (5.5) получим

$$\frac{(1-\rho_1^{k+1})}{(1-\rho_1)\rho_1^k} p_0 + (1-\alpha) \sum_{i=1}^k \frac{(1-\rho_1^{k-i+1})}{(1-\rho_1)\rho_1^{k-i}} p_i = 1. \quad (5.38)$$

Система (5.34), (5.38) решается численно. Если p_i получены, то остальные вероятности $p_{k-i,j}$ в (5.33) для $j < i$ определяются как уравнением коэффициентов при v^j в (5.33) и (5.23). Соответствующие выражения $p_{k-i,j}$ для $j < i$ через p_i ради экономии места здесь не приводятся.

6. Алгоритм вычислений

Перепишем систему уравнений (5.34), (5.38) в виде

$$p_i = \sum_{j=0}^{i-1} \mathcal{G}_{ij} p_j + \mathcal{G}_{k*} \delta_{i,k}, \quad (i = \overline{1, k}),$$

$$\sum_{j=0}^k \mathcal{G}_{0j} p_j = 1,$$
(6.1)

где \mathcal{G}_{0j} обозначает коэффициент при p_j в уравнении (5.38), \mathcal{G}_{ij} при $i > 0$ - аналогичный коэффициент в i -м уравнении (5.34), \mathcal{G}_{k*} - свободный член в k -м уравнении (5.34), а δ_{ik} - дельта-символ Кронекера.

Матрица системы (6.1) квазитреугольна. Это позволяет рекуррентно выразить все p_i при $i > 0$ через p_j

$$p_i = \gamma_i p_0 + \mathcal{G}_{k*} \delta_{i,k},$$
(6.2)

где

$$\gamma_0 = 1; \quad \gamma_i = \sum_{j=1}^{i-1} \mathcal{G}_{ij} \gamma_j, \quad (i = \overline{1, k}),$$
(6.3)

а затем с помощью первого уравнения (6.1) получить:

$$p_i = \frac{(1 - \mathcal{G}_{0k} \mathcal{G}_{k*})}{\sum_{j=0}^k \mathcal{G}_{0j} \gamma_j} + \mathcal{G}_{k*} \delta_{i,k}.$$
(6.4)

При вычислении коэффициентов \mathcal{G}_{ij} , как это следует из результатов п.5, все полиномы Гегенбауэра $C_n^k(t)$ берутся при одном и том же значении аргумента $t = t_0$, определяемом (5.32). Это позволяет применить рекуррентное соотношение [15]

$$(n+1)C_{n+1}^k(t) = 2(n+k)C_n^k(t) - (n+2k-1)C_{n-1}^k(t),$$
(6.5)

отправляясь от известных значений $C_0^k(t) = 1$ и $C_1^k(t) = 2kt$.

Технические характеристики СМО легко выражаются через вероятности p_i , вероятности q_i , определяемые по формулам (5.4) – (5.5) и вероятности r_i , инвариантные относительно α и задаваемые (4.14). Наиболее важна в сетевых задачах вероятность потери. Для приоритетных требований потеря происходит безусловно при $N_1 = k$, а также в случае несрабатывания выталкивающего механизма при $1 \leq N_2 \leq k-1$, так что

$$P_{nom}^{(1)} = q_k + (1 - \alpha) \sum_{i=1}^{k-1} p_i,$$
(6.6)

Требования второго типа либо теряются на входе, либо вытесняются с ожидания или обслуживания:

$$P_{nom}^{(2)} = r_k + \alpha \frac{\rho_1}{\rho_2} \sum_{i=1}^k p_i + \frac{\rho_1}{\rho_2} p_k,$$
(6.6)

Вероятность потери определяет пропускную способность системы. Различают абсолютную пропускную способность по i -му типу требований $\lambda_i(1 - P_{nom}^{(i)})$, относительную пропускную способность для этого типа:

$$\mathfrak{a}_i = 1 - P_{nom}^{(i)}, \quad (i = \overline{1, 2}),$$
(6.8)

а также взаимную пропускную способность по i -му типу требований, относительно j -го их типа

$$\mathfrak{a}_{ij} = \mathfrak{a}_i / \mathfrak{a}_j = (1 - P_{пот}^{(i)}) / (1 - P_{пот}^{(j)}).$$
(6.9)

Пропускной способностью можно управлять, изменяя параметр α .

Поскольку рассматриваемая СМО является эргодической, величина $\bar{\alpha}_{ij}$ допускает следующее толкование. Пропустим через систему достаточно большое число пакетов n вначале i -го, а затем j -го типа. Зарегистрируем число нормально принятых пакетов $n_*^{(i)}$ и $n_*^{(j)}$ каждого типа. Тогда в пределе при $n \rightarrow \infty$ отношение этих чисел будет стремиться к $\bar{\alpha}_{ij}$.

Получим сетевые задержки. Обозначим через \bar{s}_i среднее время пребывания в сети пакетов i -го типа. Введем нормированную задержку $\theta_i = \frac{\bar{s}_i}{\bar{\tau}_i}$, выразив \bar{s}_i в долях среднего интервала между требованиями i -го типа. По формулам Литтла для многопоточковых систем [14]

$$\theta_i = \frac{\bar{s}_i}{\bar{\tau}_i} = \frac{\bar{n}_{оч}^{(i)}}{(1 - \bar{P}_{ном}^{(i)})} + \rho_i, \quad \bar{\tau}_i = \frac{1}{\lambda_i}, \quad (i = \overline{1,2}). \quad (6.10)$$

Средняя длина приоритетной очереди $\bar{n}_{оч}^{(1)}$ определяется уравнением (4.10), в которое следует подставить q_n из (5.4)-(5.5), для $\bar{n}_{оч}^{(2)}$ сохраняется прежнее значение (4.15). Вероятность ожидания (что пакет попадет в буфер) дается равенствами

$$P_{ож}^{(1)} = 1 - q_0 - q_k, \quad P_{ож}^{(2)} = 1 - r_0 - r_k, \quad (6.11)$$

а вероятность немедленного обслуживания такова:

$$P_{но}^{(i)} = 1 - P_{ном}^{(i)} - P_{ож}^{(i)}, \quad (i = \overline{1,2}). \quad (6.12)$$

7. Числовые результаты

Для удобства сравнения с данными работы [4] зададим ту же самую емкость буфера (30 пакетов), что соответствует $k = 31$.

Вначале изучим зависимость $P_{ном}^{(i)}$ от параметра вытеснения α в режиме умеренной загрузки при преобладании неприоритетных требований ($\rho_1 = 0.2, \rho_2 = 0.9$). Результаты для нашей системы и системы [4] представлены на рис 4.

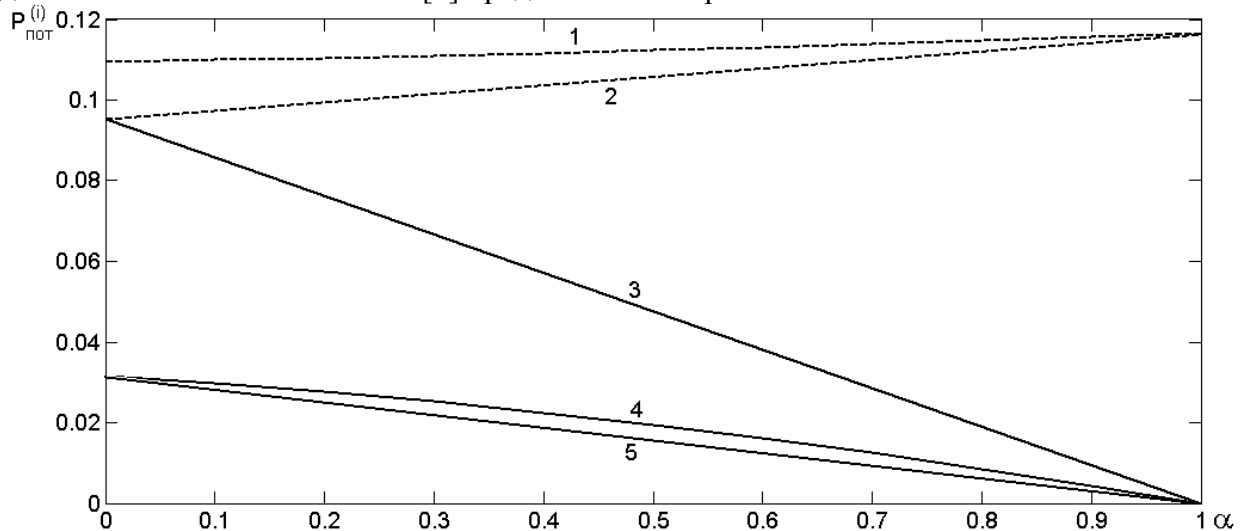


Рис. 4. Зависимость $P_{ном}^{(i)}$ от α при $\rho_1 = 0.2, \rho_2 = 0.9$: 1 – $i = 2$, абсолютный приоритет; 2 – $i = 2$, относительный приоритет; 3 – $i = 1$, относительный приоритет; 4 – $i = 1$, абсолютный приоритет; 5 – $i = 1$, абсолютный приоритет, линейная аппроксимация.

Граничные значения кривых 1 и 4 при $\alpha = 0$ и $\alpha = 1$ полученные численным путем, хорошо согласуются с аналитическими значениями, приведенными в статье. Параметр

вытеснения α позволяет эффективно управлять вероятностью потери приоритетных пакетов в очень широких пределах. Так при увеличении α от 0 до 1 значение $P_{nom}^{(i)}$ уменьшается более чем в $1,836 \cdot 10^{20}$ раз. При этом вероятность потери неприоритетного пакета возрастает всего лишь на 6,4%. Следовательно, при умеренной загрузке сети и небольшой доле приоритетных пакетов вероятностный выталкивающий механизм чрезвычайно эффективно регулирует сетевые взаимодействия в пользу приоритетного трафика, не ущемляя общий трафик.

Отметим, что введение абсолютного приоритета гораздо слабее влияет на пропускную способность сети, чем изменение α . Так, при усилении приоритета от относительного до абсолютного значение $P_{nom}^{(1)}$ уменьшается всего лишь в 2-3 раза, причем примерно в одинаковой пропорции для всех α .

Важно заметить, что зависимость $P_{nom}^{(i)}(\alpha)$, полученная численным путем, очень близка к линейной для обоих типов пакетов. В практике инженерных расчетов при слабом приоритетном трафике приемлема линейная аппроксимация

$$\bar{P}_{nom}^{(i)}(\alpha) = P_{nom}^{(i)}(0) + \alpha(P_{nom}^{(i)}(1) - P_{nom}^{(i)}(0)), \quad (7.1)$$

где $P_{nom}^{(i)}(0)$ и $P_{nom}^{(i)}(1)$ вычисляются аналитически. Относительная погрешность линейной аппроксимации

$$\Delta_i(\alpha) = \frac{|\tilde{P}_{nom}^{(i)}(\alpha) - P_{nom}^{(i)}(\alpha)|}{P_{nom}^{(i)}(\alpha)} \quad (7.2)$$

приводится в таблице 1. Для приоритетного трафика она не превосходит 30%, а для обычного – 0,7%.

Таблица 1. Относительная погрешность линейной аппроксимации

α	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
$\Delta_1(\alpha)$	0	0,0465	0,0881	0,1253	0,1590	0,1894	0,2166	0,2421	0,2651	0,2863	0
$\Delta_2(\alpha)$	0	0,0027	0,0045	0,0063	0,0072	0,0071	0,0061	0,0052	0,0043	0,0017	0

Указанный линейный закон потерь наблюдается только для случая небольшого значения отношения ρ_1/ρ_2 . В случае преобладания приоритетных пакетов картина кардинально меняется. На рис. 5 показана зависимость $P_{nom}^{(i)}$ от α для $\rho_1 = 1.2$, $\rho_2 = 0.2$.

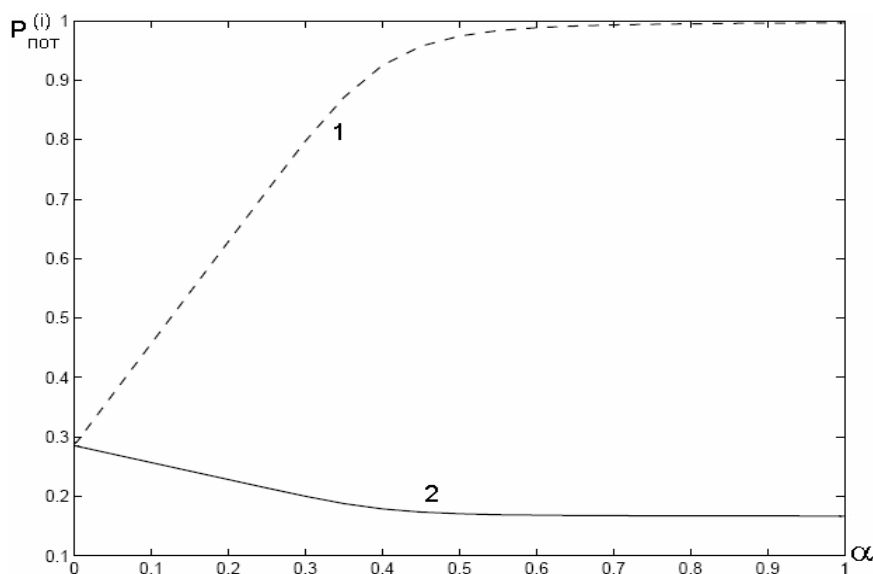


Рис. 5. Зависимость $P_{nom}^{(i)}$ от α при $\rho_1 = 1.2$, $\rho_2 = 0.2$: 1 – $i = 2$, абсолютный приоритет; 2 – $i = 1$, абсолютный приоритет.

При преобладании приоритетного трафика зависимость $P_{nom}^{(i)}(\alpha)$ заметно отличается от линейной для обоих типов пакетов и асимптотически приближается к постоянному уровню при $\alpha \rightarrow 1$. Отметим, что кривые рис.5 практически в точности повторяют аналогичные кривые работы [4], построенные для случая относительного приоритета. Следовательно, для сильно загруженной сети тип приоритета гораздо менее важен, чем наличие выталкивающего механизма и конкретное значение параметра вытеснения α . Выталкивающий механизм позволяет управлять сетевым трафиком и тогда, когда приоритетный механизм перестает действовать.

Разобранные выше два варианта задания сетевых параметров (ρ_1, ρ_2) , взятые из работы [4], соответствуют двум принципиально разным и противоположным по своему физическому смыслу режимам передачи данных. Малые значения ρ_1 означают слабую загрузку сети приоритетными пакетами. Здесь действует простейший линейный закон потерь, аналогичный картине рис 4. Высокий уровень ρ_1 вызывает появление перегрузок сети. В результате кривая потерь отклоняется от прямой и имеет участок насыщения в области больших α . С ростом ρ_1 нижняя граница зоны насыщения сдвигается в сторону малых α .

Представляет интерес изучение промежуточных режимов загрузки, в которых линейность закона потерь уже нарушена, но зона насыщения еще не сформировалась. Для выявления таких режимов был проведен численный эксперимент, в котором ρ_1 варьировалось в широких пределах от 0,1 до 2,5, а ρ_2 поддерживалось постоянным на уровне 1,5. Целью исследования был анализ сетевых взаимодействий при увеличении приоритетного трафика на фоне постоянного, и притом достаточно интенсивного потока неприоритетных пакетов.

Полученные кривые потерь для приоритетного трафика представлены на рис. 6.

Задание $\rho_1 = 0,1$ дает слабо загруженную сеть, для которой имеет силу линейный закон потерь. Увеличение нагрузки ρ_1 до 0,5 приводит к заметному отклонению кривой потерь от прямой, но сохраняется ее монотонность и выпуклость вверх. Пограничным является

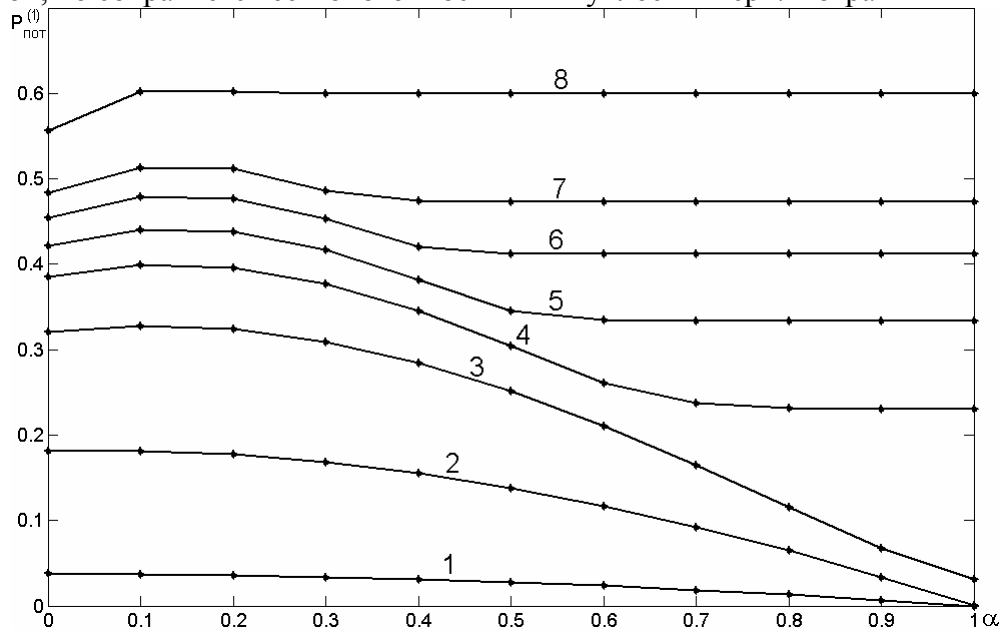


Рис 6. Зависимость $P_{nom}^{(1)}$ при $\rho_2 = 1,5$ для различных значений ρ_1 : 1 – $\rho_1 = 0,1$; 2 – $\rho_1 = 0,5$; 3 – $\rho_1 = 1,0$; 4 – $\rho_1 = 1,3$; 5 – $\rho_1 = 1,5$; 6 – $\rho_1 = 1,7$; 7 – $\rho_1 = 1,9$; 8 – $\rho_1 = 2,5$.

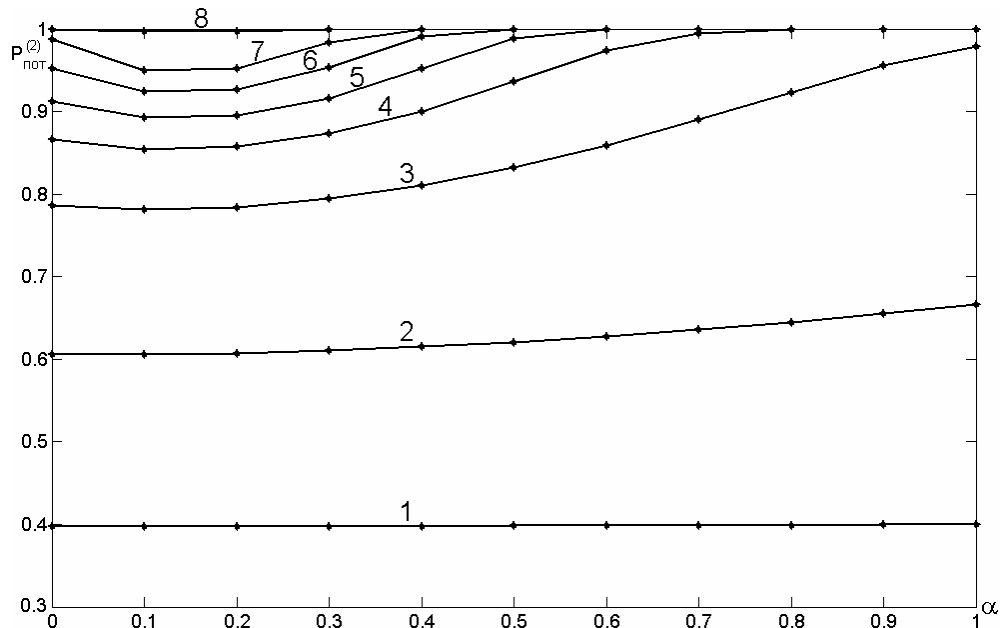


Рис. 7 Зависимость $P_{nom}^{(2)}$ от α для режимов загрузки, указанных на рис.6.

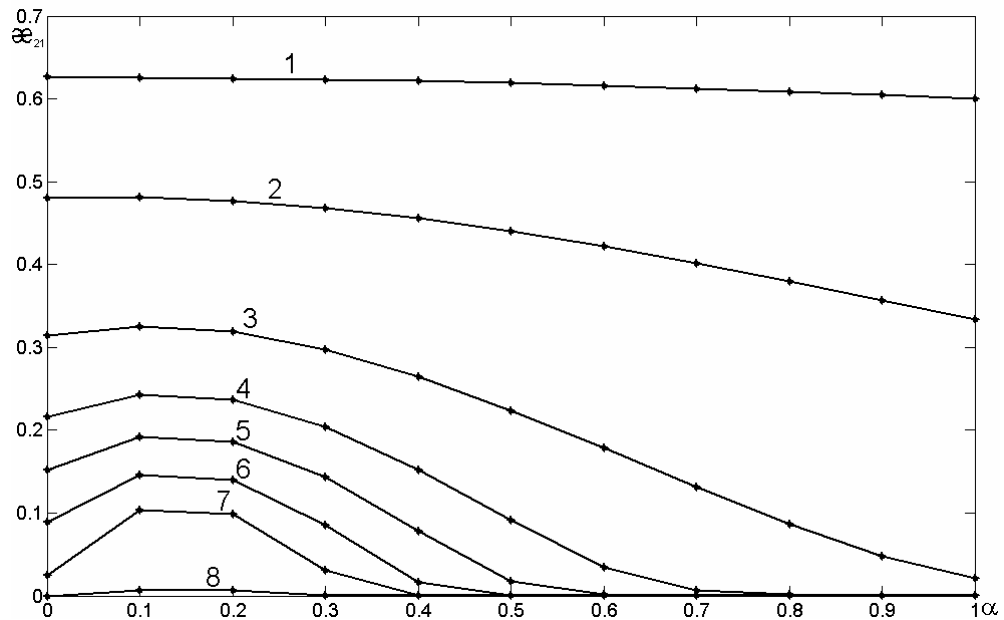


Рис. 8. Взаимная пропускная способность ζ_{21} для режимов загрузки, указанных на рис.6

режим $\rho_1 = 1$. Монотонность функции $P_{nom}^{(1)}(\alpha)$ сохраняется, но кривая теряет выпуклость, и вблизи точки $\alpha = 1$ появляется точка перегиба. При дальнейшем увеличении ρ_1 вогнутый участок кривой в области больших α сглаживается и на его месте при $1 < \rho_1 < \rho_2$ формируется зона насыщения, где $P_{nom}^{(1)}(\alpha)$ близка к постоянной. Ордината горизонтального участка кривой увеличивается с ростом ρ_1 , а в области малых α появляется локальный максимум. При дальнейшем увеличении ρ_1 область максимума сужается, а сам максимум разглаживается. Для значений $\rho_1 > 2$ функция $P_{nom}^{(1)}(\alpha)$ имеет характерный для перегруженной сети вид монотонной кривой с насыщением.

Графики $P_{nom}^{(2)}(\alpha)$ для тех же режимов загрузки приведены на рис. 7. Эти графики проясняют смысл «пограничного» режима $\rho_1 = 1$. При такой загрузке в случае детерминированного вытеснения ($\alpha = 1$) имеем $P_{nom}^{(2)}(\alpha)$, близкое к единице, то есть сеть

можно практически «закрыть» для неперіоритетных пакетов. При больших значениях ρ_1 эффект «закрытия» сети для обычного трафика происходит при значениях $\alpha < 1$, причем для $\rho_1 > 2$ при очень малых α . Вся сеть начинает работать только на передачу приоритетных пакетов, причем даже для них пропускная способность стабилизируется где-то на уровне всего лишь 40% для $\rho_1 = 2,5$.

При $1 < \rho_1 < 2$ наблюдается неожиданный эффект в области малых значений α . Функция $P_{nom}^{(1)}(\alpha)$ имеет локальный максимум, а функция $P_{nom}^{(2)}(\alpha)$ в той же области имеет локальный минимум. В результате взаимная пропускная способность α_{21} имеет вид, представленный на рис. 8. Взаимная пропускная способность обычного трафика по отношению к приоритетному имеет ярко выраженный максимум в области α порядка 10-20%. Для $\rho_1 > 2$ значение α_{21} пренебрежимо мало во всем диапазоне изменения α .

Полученные результаты хорошо иллюстрируют графики нормированного времени задержки приоритетных пакетов θ_1 , определяемого согласно (6.10), представленные на рис. 9. При $\rho_1 = 0,1$ приоритетные пакеты практически не задерживаются в буфере, и поэтому $\theta_1 \approx \rho_1 = 0,1$ (второе слагаемое в (6.10) пренебрежимо мало). При $\rho_1 = 0,5$ задержка θ_1 увеличивается примерно до 1, причем слабо зависит от α . «Пограничный» режим $\rho_1 = 1$ характеризуется тем, что во-первых задержка начинает ощутимо зависеть от α и, во-вторых, достигает значительной величины порядка 15-20 единиц $\bar{\tau}_1$. При дальнейшем увеличении ρ_1 имеем монотонно возрастающую зависимость $\theta_1(\alpha)$ с зоной насыщения в области больших α .

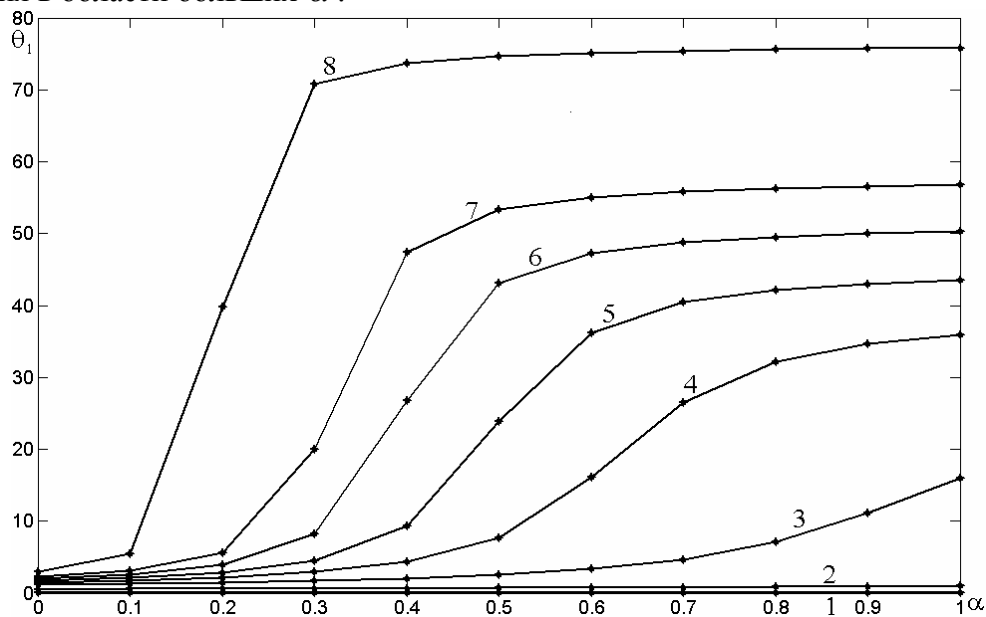


Рис. 9. Нормированное время задержки приоритетных пакетов θ_1 для режимов загрузки, указанных на рис. 6.

Отметим, что приведенные сложные немонотонные зависимости сетевых параметров от параметра вытеснения α имеют место только при достаточно большом значении интенсивности фонового потока неперіоритетных пакетов ρ_2 . В работе, при тех же значениях ρ_1 , был детально изучен случай $\rho_2 = 0,5$, отвечающий умеренной загрузке. При этом на графиках, аналогичных графикам рис. 6-8, какие-либо экстремумы отсутствуют, а все кривые меняются монотонным образом.

Приведенные результаты показывают, что описанные сложные нелинейные эффекты сетевого взаимодействия проявляются только при достаточно большой загрузке сети, что

согласуется с данными натурных наблюдений. Модели приоритетных многопоточковых СМО с вероятностным выталкивающим механизмом позволяют успешно моделировать эти эффекты и объяснять их.

8. Пример управление мультимедиа потоком данных.

Рассмотрим простейший пример управления двухпоточковой сетью, описанной в работе [4] и использованный при построении графиков рис. 4 ($\rho_1 = 0.2, \rho_2 = 0.9$). В данном случае, как мы видели, выполняется линейный закон потерь, и справедлива линейная аппроксимация (7.1). По графикам рис. 4. получаем в случае диспетчеризации сети с абсолютным приоритетом пропускные способности

$$\alpha_1 = 0,9688 + 0,0312\alpha, \alpha_2 = 0,8904 - 0,007\alpha \quad (8.1)$$

и взаимную пропускную способность второго потока относительно первого

$$\alpha_{21} = \frac{0,8904 - 0,007\alpha}{0,9688 + 0,0312\alpha}. \quad (8.2)$$

Потребуем, чтобы первый поток передавался с пропускной способностью не ниже 99%, а пропускная способность второго по отношению к нему была не меньше 89%. Тогда решая неравенства

$$\alpha_1 > 0,99, \alpha_{21} > 0,89 \quad (8.3)$$

относительно параметра вытеснения α находим

$$0,679 < \alpha < 0,887. \quad (8.4)$$

Если ту же самую сеть диспетчеризовать с относительным приоритетом, как это описано в работе [4], то получим вместо (8.1), (8.2) и (8.4), соответственно,

$$\alpha_1 = 0,904 + 0,096\alpha, \alpha_2 = 0,904 - 0,0152\alpha \quad (8.5)$$

$$\alpha_{21} = \frac{0,904 - 0,0152\alpha}{0,904 + 0,096\alpha}, \quad (8.6)$$

$$0,896 < \alpha < 0,900, \quad (8.7)$$

то есть в нашем случае требуется менее интенсивное вытеснение неприоритетных пакетов из накопителя, чем в модели [4], что повышает гибкость управления.

В случае сильно загруженной сети вместо линейных выражений (8.1) и (8.5) приходится использовать нелинейные зависимости, подобные представленным на графиках п.7 и получаемые численным путем. При этом приходится существенно снизить ограничения на пропускные способности по сравнению с (8.3) и учитывать все нюансы сложной немонотонной кривой потерь.

Заключение.

В статье изложен метод расчета характеристик двухпоточковой компьютерной сети. Абсолютный приоритет по обслуживанию, а также приоритет по постановке в очередь предоставляется TCP пакетам, как управляющему виртуальному соединению, а видеопоток UDP играет роль фонового трафика. Сеть моделируется с помощью СМО конечной емкости, снабженной вероятностным выталкивающим механизмом в комбинации с абсолютным приоритетом. Предложен эффективный вычислительный алгоритм, позволяющий сетевым инженерам и проектировщикам без больших усилий оценивать возможные варианты эксплуатации сетей.

Показано, что при умеренной нагрузке, наблюдающейся, в частности, в космическом эксперименте, описанном в п.1, вероятностный выталкивающий механизм является чрезвычайно эффективным средством управления высокоприоритетным трафиком, не ущемляющим при этом низкоприоритетный трафик. Чувствительность сети к выбору параметра вытеснения α на несколько порядков выше, чем к выбору типа приоритета, причем сеть можно плавно настроить на нужную пропускную способность. Существенно,

что выталкивающий механизм управления сетью продолжает работать даже в сильно загруженной сети, когда приоритетный механизм управления резко снижает свою эффективность.

В сильно загруженных сетях пропускная способность весьма сложным образом зависит от параметра вытеснения α . При определенном соотношении параметров загрузки возможен своего рода «резонансный» режим. Взаимная пропускная способность низкоприоритетного трафика по отношению к высокоприоритетному в этом режиме имеет максимум по α .

Сеть с абсолютным приоритетом дает максимум преимуществ наиболее важным типам требований. При этом по сравнению с относительным приоритетом можно уменьшить вероятность вытеснения α , что делает алгоритм управления более гибким.

Настоящая работа продолжает цикл статей по исследованию вероятностного выталкивающего механизма в приоритетных СМО, задуманный профессором Санкт-Петербургского государственного политехнического университета Никитой Олеговичем Вильчевским и начатый работой [4]. Авторы статьи считают своим долгом вспомнить этого замечательного ученого и прекрасного человека, чутко уловившего перспективность и важность исследований в этом направлении. Данная работа посвящается светлой памяти Никиты Олеговича.

Литература

1. Гнеденко Б.В., Даниелян Э.А., Димитров Б.Н., Климов Г.П., Матвеев В.Ф. Приоритетные системы обслуживания. М.: МГУ, 1973.
2. Джейсоул Н. Очереди с приоритетами. М.: Мир, 1973.
3. Башарин Г.П. Некоторые результаты для систем с приоритетом // Массовое обслуживание в системах передачи информации. М.: Наука, 1969, с. 39-53.
4. Avrachenkov K.E., Vilchevsky N.O., Shevljakov G.L. Priority queueing with finite buffer size and randomized push-out mechanism // Proceedings of the ACM international conference on measurement and modeling of computer (SIGMETRIC 2003). San Diego: 2003, p. 324-335.
5. Клейнрок Л. Теория массового обслуживания М.: Машиностроение, 1979.
6. Башарин Г.П. Обслуживание двух потоков на однолинейной системе с ограниченным числом мест для ожидания и абсолютным приоритетом // Известия ДН СССР. Техническая кибернетика, 1967, №5, с.106-116.
7. Башарин Г.П. О пуассоновских обслуживающих системах с абсолютным приоритетом // Массовое обслуживание в системах передачи информации. М.: Наука, 1969, с. 3-20.
8. Bondi A. An analysis of finite capacity queues with priority scheduling and common or reserved waiting areas // Computers and operations research, 1989, vol. 16, no. 3, p. 217-233.
9. Hegde N., Avrachenkov K.E. Service differentiation and guarantees for TCP elastic traffic // Lecture notes in computer science, 2002, vol. 2511, p. 159-168.
10. White H., Christie L.S. Queueing with preemptive priorities or with breakdown // Operations research, 1958, vol. 6, no. 1, p. 79-95.
11. Stephan F.F. Two queues under preemptive priority with Poisson arrival and service rates // Operations research, 1958, vol. 6, no.3, p. 399-418.
12. Kapadia A.S., Kazmi, M.F., Mitchell A.C. Analysis of a finite capacity non-preemptive priority queue // Computers and operations research, 1984, vol. 11, no. 3, p. 337-343.
13. Kapadia A.S., Chiang Y.K., Kazmi M.F. Finite capacity priority queues with potential health applications // Computers and operations research, 1985, vol. 12, no. 4, p. 411-420.
14. Бронштейн О.И., Духовный И.М. Модели приоритетного обслуживания в информационно-вычислительных системах. М.: Наука, 1976.
15. Бейтмен Г., Эрдейн А. Высшие трансцендентные функции. Функции Бесселя, функции параболического цилиндра, ортогональные многочлены. М.: Наука, 1974.