7 КОМПЬЮТЕРНЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 621.395.74

МОДЕЛИРОВАНИЕ ЯДРА МУЛЬТИСЕРВИСНОЙ СЕТИ С ОТНОСИТЕЛЬНОЙ ПРИОРИТЕЗАЦИЕЙ НЕОДНОРОДНОГО ТРАФИКА

Т.И. Алиев, И.Е. Никульский, В.О. Пяттаев

Для ядра городской мультисервисной телекоммуникационной сети, ориентированной на передачу трех видов трафика (голос, видео и данные), разработаны математические модели в терминах теории массового обслуживания. Исследование характеристик функционирования (среднего значения и джиттера задержки кадров разных типов) проводится на основе комбинированного подхода, предполагающего совместное применение аналитических и имитационных методов. На основе многочисленных модельных экспериментов при различных предположениях о характере трафика и процесса обработки кадров в узлах выполнен подробный анализ характеристик функционирования сети при использовании способа управления трафиком на основе относительных приоритетов.

Ключевые слова: мультисервисная сеть, неоднородный трафик, характеристики качества обслуживания (QoS), сетевая задержка, джиттер задержки, относительные приоритеты, аналитическое и имитационное моделирование.

Введение

При переходе от традиционных телекоммуникационных сетей к мультисервисным сетям связи (МСС) следующего поколения (NGN – Next Generation Networks) возникает множество технологических, методологических и других проблем, среди которых центральное место занимает проблема обеспечения требуемого качества обслуживания (QoS – Quality of Service) для различных видов трафика и, особенно, для речи. Технологической основой построения мультисервисных сетей является стек протоколов IP/MPLS/Ethernet [1], реализующий пакетную передачу и коммутацию всех видов трафика. Для прогнозирования качества обслуживания на этапе сетевого планирования широкое применение находят методы математического моделирования проектируемой сети.

Основные подходы к оценке качества обслуживания сетей IP специфицированы в Рекомендации МСЭ-Т Y.1541 [2], в соответствии с которой основными показателями качества обслуживания мультимедийного трафика служат средняя сетевая задержка передачи пакета (кадра) u_N и ее вариация (джиттер) σ_N. В то же время большинство исследований подобного рода сетей ограничиваются в основном подробным анализом средней задержки, а анализ джиттера проводится без учета приоритетного механизма управления трафиком в узлах сети. В данной работе с использованием комбинированного подхода к исследованию QoS, предполагающего совместное применение аналитических и имитационных методов, выполнена оценка и детальный анализ показателей качества обслуживания неоднородного трафика с управлением в узлах сети на основе относительных приоритетов.

Объект исследования

Сложность и большая размерность мультисервисных сетей приводят к необходимости применения в качестве основного подхода к исследованию МСС принципа декомпозиции, основанного на делении сети в соответствии с ее иерархической организацией на уровни с последующим их раздельным исследованием. В современных МСС

можно выделить следующие уровни: доступа, агрегирования доступа, ядра МСС и агрегирования услуг.

Ядро (магистраль) МСС является самым мощным и высокопроизводительным фрагментом сети, обеспечивающим перенос и коммутацию больших потоков данных, поступающих от нижних иерархических уровней, а также от уровня агрегирования услуг [3]. Для получения высокой производительности и низких значений задержки на рассматриваемом участке сети применяются высокоскоростные оптические каналы связи, а также специализированные высокопроизводительные маршрутизаторы, использующие технологию MPLS.

В зависимости от передаваемой нагрузки, производительности используемых маршрутизаторов и требуемых значений задержки могут быть выбраны разные варианты топологии ядра сети (рис. 1), различающиеся производительностью, способностью выравнивания нагрузки и стоимостью. Такими топологиями являются: кольцо (рис. 1, a); частичносвязная (рис. 1, b); полносвязная (рис. 1, c).

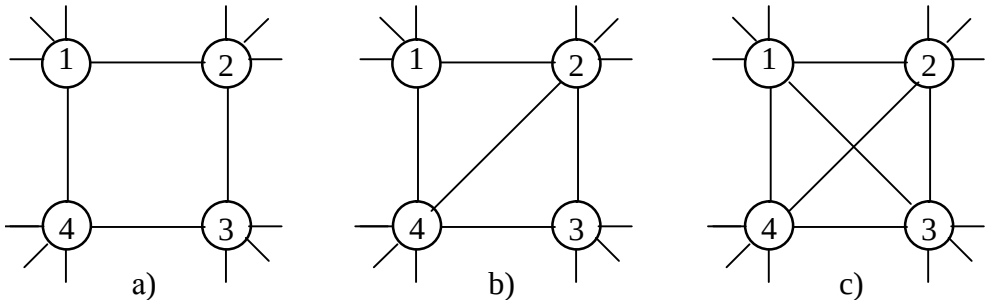

Рис. 1. Топологии ядра МСС

В качестве протокольной единицы в исследуемой сети рассматривается кадр Ethernet. Процесс передачи кадров в рассматриваемом фрагменте сети сводится к следующему. Потоки кадров с присоединенными метками поступают от граничных маршрутизаторов на входы портов маршрутизаторов ядра сети. В соответствии с применяемыми процедурами дифференцированного обслуживания различных классов трафика, реализуемыми в узлах MPLS, поступающие потоки на основе анализа аппаратно-программными средствами маршрутизаторов меток, присоединенных к кадрам, разделяются на три очереди, соответствующие трем рассматриваемым классам обслуживания. Эти очереди различаются длинами поступающих в них кадров, а также приоритетами их обслуживания. Выбираемые из очередей кадры передаются в соответствующие выходные порты маршрутизаторов и далее по каналам связи – на другие маршрутизаторы ядра сети.

Модели ядра и узла МСС

Модель ядра МСС (рис. 2), представленная в виде графа разомкнутой сети массового обслуживания (СеМО), соответствует наиболее общей полносвязной топологии. Узел 0 СеМО отображает внешнюю среду, в качестве которой выступают уровни агрегирования доступа и агрегирования услуг, а узлы 1–4 – обработку кадров в маршрутизаторах ядра МСС и их передачу по каналам связи. Маршруты передаваемых кадров описываются в модели вероятностями передач p_{ij} $(i, j = \overline{0,4})$, причем $\sum_{j=0}^{4} p_{ij} = 1$ для всех $i = \overline{0,4}$.

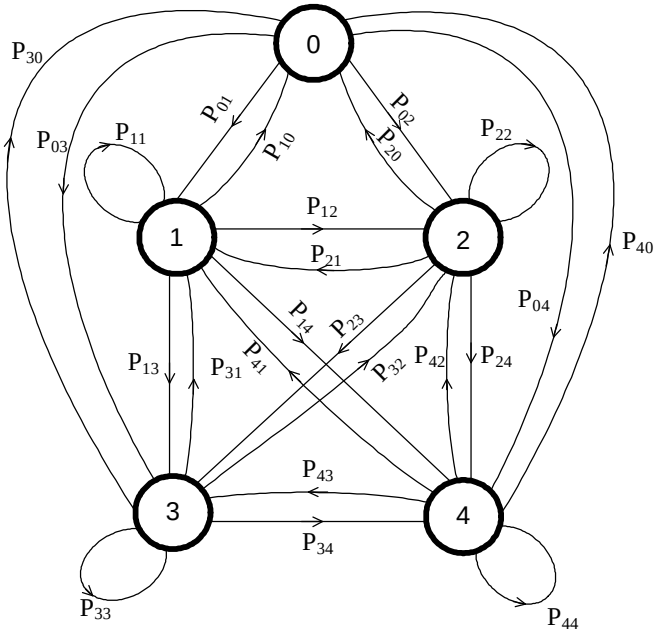

Рис. 2. Граф модели ядра МСС

Таким образом, рассмотренный фрагмент сети представляет собой систему, включающую большое число распределенных взаимосвязанных ресурсов, выполняющих действия с заявками (кадрами). При этом заявки могут передаваться и перераспределяться между этими ресурсами.

Модель узла (маршрутизатора) может быть представлена в виде разомкнутой трехфазной СеМО (рис. 3). В фазе 1 выполняется прием поступающих с интенсивностью Λ кадров Ethernet в накопитель, анализ меток и распределение кадров по трем очередям: 1) речевые (VoIP – Voice over IP); 2) видео (IPTV); 3) данные (DoIP – Data over IP). В фазе 2 реализуется передача кадров в выходные порты маршрутизатора в соответствии с маршрутной таблицей и присвоенными очередям приоритетами. В фазе 3 осуществляется передача кадров к другим маршрутизаторам ядра или к пограничным маршрутизаторам, расположенным в сети агрегирования доступа или в сети агрегирования услуг.

Большинство современных маршрутизаторов используют параллельную обработку кадров сверхбыстродействующими аппаратными средствами, поэтому время обработки кадров незначительно (несколько микросекунд), а основной вклад в задержку вносят время передачи и время ожидания в очередях.

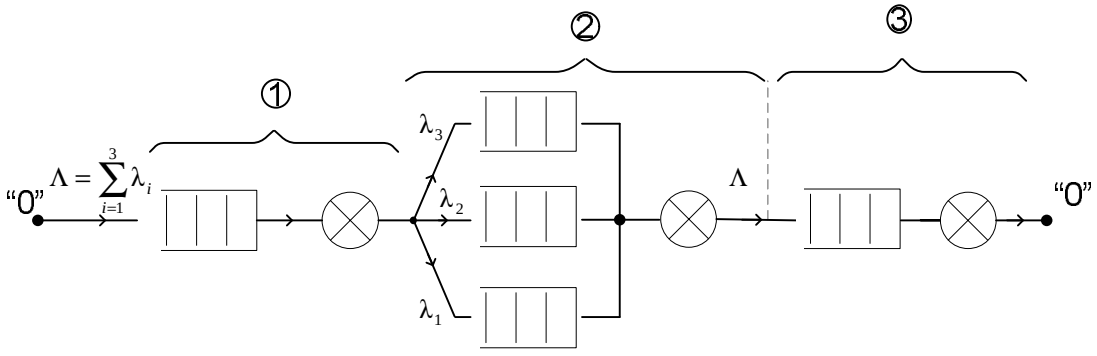

Рис. 3. Модель узла МСС

Исследование модели ядра сети выполнялось с применением аналитических и статистических методов при следующих предположениях и допущениях, принятых на основе анализа характеристик современного сетевого оборудования, используемого для построения МСС, и протоколов передачи данных в сетях NGN. Из внешней среды поступают кадры трех типов: 1) VoIP (речь) с интенсивностью $\lambda_1 = 0{,}12\Lambda$; 2) IPTV (видео) с интенсивностью $\lambda_2 = 0{,}2\Lambda$; 3) DoIP (данные) с интенсивностью $\lambda_3 = 0{,}68\Lambda$, где Λ – суммарная интенсивность потока кадров, варьируемая величина при аналитических расчетах и проведении имитационных экспериментов. Кадрам разных типов назначены приоритеты, причем наивысший приоритет имеют кадры VoIP, а самый низкий – кадры DoIP.

Обслуживание в первой фазе – бесприоритетное со средней длительностью $b_1 \approx 0{,}5$ мкс, включающей время приема кадра в буфер приемника. Обслуживание во второй фазе – с относительными приоритетами, причем длительности обслуживания кадров разных типов равны $b_{21} = 0{,}5$ мкс, $b_{22} = 1$ мкс, $b_{23} = 3$ мкс соответственно. Длительность обслуживания в третьей (бесприоритетной) фазе $b_3 \approx 0{,}5$ мкс.

Аналитическое моделирование МСС

Аналитическое моделирование базируется на математическом аппарате теории массового обслуживания и толерантном преобразовании разомкнутых СеМО, позволяющем свести задачу расчета неоднородной приоритетной СеМО к расчету независимых одноканальных систем массового обслуживания (СМО) M/G/1, что в терминах символики Кендалла [4] означает: поступающие заявки (кадры) разных типов образуют простейшие (марковские) потоки (M – Markovian), а длительности обслуживания заявок распределены по произвольному закону общего вида (G – General). Таким образом, в качестве базовой математической модели, отображающей обработку кадра в одной фазе узла, будем рассматривать СМО типа M/G/1 с неоднородным потоком заявок и, в общем случае, приоритетным обслуживанием. Для расчета базовой модели воспользуемся результатами, полученными в [5].

Средняя задержка кадров типа k $(k = 1, 2, 3)$ в приоритетной фазе i определяется как среднее время пребывания заявок в СМО типа M/G/1 с относительными приоритетами [5]:

$$u_{ik} = w_{ik} + b_{ik}, \tag{1}$$

где b_{ik} – среднее время обработки кадров; w_{ik} – среднее время ожидания в очереди:

$$w_{ik} = \frac{\sum_{j=1}^{3} \lambda_{ij} b_{ij}^{(2)}}{2(1-R_{ik-1})(1-R_{ik})}. \tag{2}$$

Джиттер задержки кадров типа k $(k = 1, 2, 3)$ представляет собой среднеквадратическое отклонение времени пребывания заявок и определяется как

$$\sigma_{ik} = \sqrt{w_{ik}^{(2)} + 2w_{ik}b_{ik} + b_{ik}^{(2)} - u_{ik}^2}, \tag{3}$$

где $b_{ik}^{(2)}$ и $w_{ik}^{(2)}$ – вторые начальные моменты времени обслуживания и ожидания в очередях соответственно:

$$w_{ik}^{(2)} = \frac{\sum_{j=1}^{3} \lambda_{ij} b_{ij}^{(3)}}{3(1-R_{ik-1})^2(1-R_{ik})} + \frac{\sum_{j=1}^{k} \lambda_{ij} b_{ij}^{(2)} \sum_{j=1}^{3} \lambda_{ij} b_{ij}^{(2)}}{2(1-R_{ik-1})^2(1-R_{ik})^2} + \frac{\sum_{j=1}^{k-1} \lambda_{ij} b_{ij}^{(2)} \sum_{j=1}^{3} \lambda_{ij} b_{ij}^{(2)}}{2(1-R_{ik-1})^3(1-R_{ik})}. \tag{4}$$

В выражениях (2)–(4) используются следующие обозначения: $R_{ik} = \sum_{j=1}^{k} \rho_{ij} = \sum_{j=1}^{k} \lambda_{ij} b_{ij}$ – частичная суммарная загрузка, создаваемая первыми k классами $(k = 1, 2, 3)$, причем $R_0 = 0$; λ_{ij} – интенсивность потока заявок типа j; $b_{ij}^{(2)}$ и $b_{ij}^{(3)}$ – соответственно второй и третий начальные моменты времени обслуживания заявок типа $j = \overline{1,3}$ в фазе $i = \overline{1,3}$.

Значения $b_{ij}^{(2)}$ и $b_{ij}^{(3)}$ могут быть получены экспериментальным путем в процессе измерения реальных параметров обработки данных в маршрутизаторах или рассчитаны для заданных априори конкретных законов распределений.

Для фаз, в которых реализуется бесприоритетная обработка кадров, после преобразования выражений (1)–(4) получим следующие формулы для расчета среднего значения и джиттера задержки кадров типа k:

$$u_{ik} = \frac{\sum_{j=1}^{3} \lambda_{ij} b_{ij}^{(2)}}{2(1-R_i)} + b_{ik}; \qquad \sigma_{ik} = \sqrt{\frac{\sum_{j=1}^{3} \lambda_{ij} b_{ij}^{(3)}}{3(1-R_i)} + \frac{\left(\sum_{j=1}^{3} \lambda_{ij} b_{ij}^{(2)}\right)^2}{4(1-R_i)^2} + b_{ik}^{(2)} - b_{ik}^2},$$

где $R_i = R_{i3}$ – суммарная загрузка системы. Среднее значение и джиттер задержки кадров типа k в узле определяются по формулам

$$u_y = \sum_{i=1}^{3} u_{ik}, \qquad \sigma_y = \sqrt{\sum_{i=1}^{3} \sigma_{ik}^2}.$$

Средняя сетевая задержка и ее джиттер вычисляются через соответствующие значения узловых характеристик:

$$u_N = \sum_{j=1}^{N} \alpha_j u_j, \quad \sigma_N = \sqrt{\sum_{j=1}^{N} \alpha_j \sigma_j^2}, \tag{5}$$

где N – число узлов в сети; σ_j – джиттер задержки в узле $j = \overline{1, N}$; u_j – средняя задержка заявки в узле j; α_j – коэффициент передачи для узла j СеМО. Коэффициенты передач α_j, используемые в выражениях (5), вычисляются на основе матрицы вероятностей передач путем решения системы линейных алгебраических уравнений

$$\alpha_j = \sum_{i=0}^{N} p_{ij} \alpha_i, \quad (j = 0,1,\mathbf{K}, N)$$

с учетом того, что $\alpha_0 = 1$. Выражения (1)–(5) получены для стационарного режима функционирования сети, в которой узлы и каналы связи абсолютно надежны.

С использованием предложенных аналитических моделей сети и узла выполнен анализ влияния различных типовых топологий (рис. 1) на характеристики качества функционирования ядра МСС. При этом связность сегментов сети представленных топологий задавалась соответствующими матрицами вероятностей передач:

$$P_a = \begin{array}{c|ccccc} & 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 0{,}25 & 0{,}25 & 0{,}25 & 0{,}25 \\ 1 & 0{,}5 & 0 & 0{,}25 & 0 & 0{,}25 \\ 2 & 0{,}5 & 0{,}25 & 0 & 0{,}25 & 0 \\ 3 & 0{,}5 & 0 & 0{,}25 & 0 & 0{,}25 \\ 4 & 0{,}5 & 0{,}25 & 0 & 0{,}25 & 0 \end{array}; \qquad P_b = \begin{array}{c|ccccc} & 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 0{,}25 & 0{,}25 & 0{,}25 & 0{,}25 \\ 1 & 0{,}64 & 0 & 0{,}18 & 0{,}19 & 0 \\ 2 & 0{,}44 & 0{,}19 & 0 & 0{,}25 & 0{,}18 \\ 3 & 0{,}64 & 0 & 0{,}18 & 0 & 0{,}18 \\ 4 & 0{,}44 & 0{,}19 & 0{,}19 & 0{,}18 & 0 \end{array};$$

$$P_c = \begin{array}{c|ccccc} & 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 0{,}25 & 0{,}25 & 0{,}25 & 0{,}25 \\ 1 & 0{,}57 & 0 & 0{,}14 & 0{,}14 & 0{,}15 \\ 2 & 0{,}57 & 0{,}14 & 0 & 0{,}14 & 0{,}15 \\ 3 & 0{,}57 & 0{,}14 & 0{,}14 & 0 & 0{,}15 \\ 4 & 0{,}57 & 0{,}14 & 0{,}14 & 0{,}15 & 0 \end{array}.$$

В результате расчета характеристик передачи кадров с использованием выражений (1)–(5) при различных входных параметрах получены зависимости средней сетевой задержки ядра МСС от интенсивности входного потока заявок для различных топологий (рис. 4, а) и зависимости средней задержки в узлах для различных типов трафика от интенсивности входного потока заявок (рис. 4, б). Анализ полученных результатов показывает, что минимальная сетевая задержка характерна для топологий с большей связностью. С ростом загрузки сети (за счет интенсивности поступления кадров) увеличивается различие между значениями задержек для кольцевой и полносвязной топологий, которое может достигать 100% и более. Кроме того, для кадров VoIP, обладающих самым высоким относительным приоритетом, обеспечивается достаточно хорошее качество обслуживания в узле, т.е. небольшое время задержки, даже при возникновении перегрузок сети, когда суммарная нагрузка становится больше единицы. Это свойство, называемое защитой от перегрузок, обеспечивается за счет существенного увеличения задержки низкоприоритетных кадров DoIP, время ожидания которых в узле резко возрастает.

При достижении суммарной нагрузкой, создаваемой кадрами всех трех типов, значения 1 время ожидания кадров 3-го типа (DoIP) начинает увеличиваться, что, в конечном счете, при ограниченной емкости буферной памяти узла приводит к отказу в обслуживании, при этом приоритетные кадры типа 1 (VoIP) и типа 2 (IPTV) продолжают обслуживаться и имеют конечное время ожидания. Дальнейшее увеличение нагрузки может привести к потере кадров IPTV, когда создаваемая кадрами VoIP и IPTV нагрузка достигнет значения 1. Отметим, что при бесприоритетном обслуживании защита от перегрузок отсутствует для всех типов кадров.

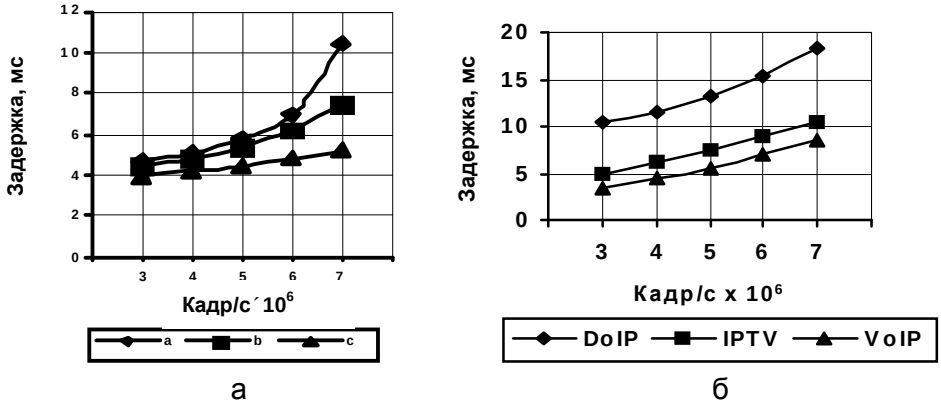

Рис. 4. Задержка для различных топологий МСС (а) и типов трафика (б)

Приоритетное обслуживание обеспечивает существенное снижение задержки для высокоприоритетных типов трафика VoIP и IPTV как в узлах сети, так и в ядре МСС. Полученные с использованием выражения (3) значения джиттера задержки для трафика данных (DoIP) с наименьшим приоритетом значительно превышают значения для высокоприоритетных типов трафика и меняются нелинейно с ростом нагрузки (рис. 5).

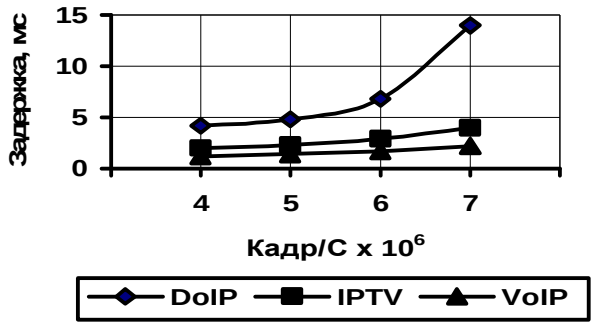

Рис. 5. Джиттер задержки для различных типов трафика

Имитационное моделирование МСС

Для оценки различий между результатами аналитического моделирования при использовании предположения о пуассоновском характере входного потока и при потоках, отличных от пуассоновских, была разработана имитационная модель узла (рис. 3) с неоднородным потоком заявок и приоритетным обслуживанием в среде GPSS World. В модели была заложена возможность выбора следующих законов распределений интервалов между поступлением заявок входного потока: пуассоновского, Эрланга и гиперэкспоненциального. Эти же распределения использовались для интервалов обработки кадров во всех фазах модели узла ядра ММС (рис. 3).

На рис. 6 представлены результаты аналитического и имитационного моделирования, показывающие зависимости средних задержек кадров разных типов в узле ядра МСС от интенсивности входного потока.

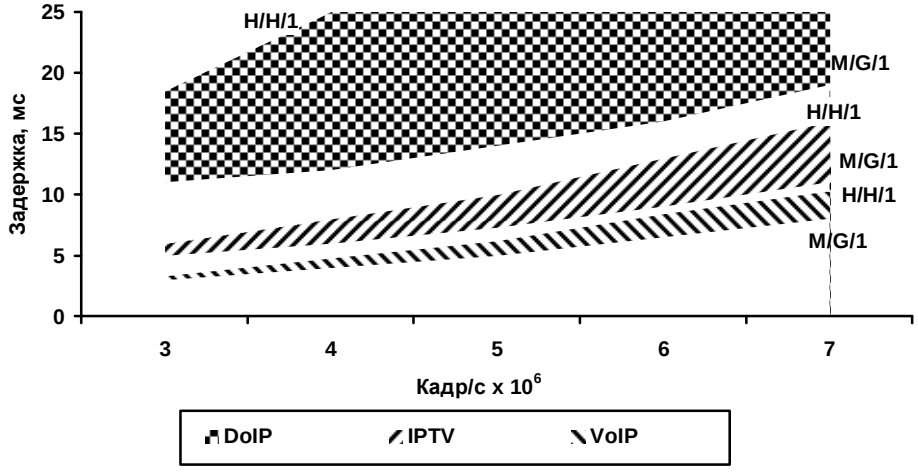

Рис. 6. Сопоставление результатов имитационного и аналитического моделирования

Три заштрихованные области для трех видов трафика с различными приоритетами обслуживания кадров в узле МСС характеризуют области значений задержек кадров при различных законах распределений потоков поступающих кадров и длительностей их обработки в узле МСС.

Нижние границы заштрихованных областей соответствуют результатам моделирования с пуассоновской моделью входного потока (M/G/1). В этом случае результаты аналитического и имитационного моделирования практически не отличаются. Верхние границы этих областей соответствуют результатам имитационного моделирования при

использовании наиболее тяжелого режима – гиперэкспоненциального распределения интервалов во входном потоке и гиперэкспоненциального распределения интервалов обслуживания, что в терминах символики Кендалла обозначается как СМО H/H/1. Основанием для выбора гиперэкспоненциального распределения с коэффициентом вариации, равным 3, может служить то, что это распределение позволяет смоделировать наиболее тяжелый режим функционирования ядра МСС. Это объясняется следующими обстоятельствами. С одной стороны, при гиперэкспоненциальном распределении интервалов между поступающими в сеть кадрами существует большая вероятность появления небольших интервалов, что приводит к группированию поступающих кадров и созданию в некоторые периоды времени больших нагрузок. Так, например, при гиперэкспоненциальном распределении вероятность того, что интервалы между поступающими в сеть кадрами будут меньше среднего значения, превышает 0,8, в то время как для пуассоновского потока эта вероятность равна 0,63. С другой стороны, наличие так называемого «тяжелого хвоста» гиперэкспоненциального распределения, означающего, что вероятность появления больших значений случайной величины значительно отличается от нуля, может вызывать большие задержки при обработке кадров в узлах. Кроме того, такое распределение в ряде случаев с высокой степенью адекватности отражает характер входного потока и обслуживания в IP-сетях.

Анализ полученных результатов позволяет сделать вывод о том, что наименьшие расхождения результатов моделирования при различных предположениях о характере трафика и обслуживания кадров в узлах наблюдаются в областях малых нагрузок для трафика с высоким приоритетом VoIP, чувствительного к задержкам. В области больших нагрузок для этого трафика расхождения увеличиваются и составляют 1,6–1,8 раз. Для эластичного (малочувствительного к задержкам) трафика DoIP с наименьшим приоритетом расхождения могут составлять 2–8 раз в зависимости от уровня загрузки.

Заключение

Выполненные исследования с использованием разработанных аналитических и имитационных моделей ядра МСС показали, что использование приоритетной обработки кадров в узлах не только обеспечивает для наиболее чувствительного трафика VoIP наименьшую среднюю задержку, но и позволяет минимизировать джиттер задержки. При этом различие между характеристиками высокоприоритетного трафика при разных законах распределений интервалов между поступающими кадрами и времени их обработки лежит в приемлемых для практики пределах для значений загрузки узлов и каналов связи в интервале от 0,1 до 0,7. Последнее обстоятельство дает основание полагать, что данные модели могут быть использованы для оценочных расчетов сегментов мультисервисных сетей в режимах малых и средних нагрузок при потоках заявок, отличных от простейшего. Дальнейшие исследования представляется целесообразным направить на поиск приближений для моделей СМО типа G/G/1 с неоднородным потоком заявок и приоритетным обслуживанием, что позволит проводить анализ качества обслуживания в современных сетях с произвольным распределением интервалов во входном потоке и обслуживания, в том числе использовать распределения, обладающие свойствами самоподобия.

Литература

1. Гольдштейн А.Б., Гольдштейн Б.С. Технологии и протоколы MPLS. – СПб: БХВ–Санкт-Петербург, 2005. – 304 с.: ил.

2. ITU-T Recommendation Y.1541 (02/2006) – Network performance objectives for IP-based services.
3. Aliev T.I., Nikulsky I.Y., Pyattaev V.O. Modeling of packet switching network with relative prioritization for different traffic types // ICAT. – 2008, Feb. 17–20. – S. 2174–2176.
4. Клейнрок Л. Вычислительные системы с очередями / Пер. с англ. – М.: Мир, 1979. – 600 с.
5. Алиев Т.И. Характеристики дисциплин обслуживания заявок с несколькими классами приоритетов // Известия АН СССР. Техническая кибернетика. – 1987. – № 6. – С.188–191.
6. Руководство пользователя по GPSS World / Пер. с англ. – Казань: «Мастер Лайн», 2002. – 384 с.

Алиев Тауфик Измайлович – Санкт-Петербургский государственный университет информационных технологий, механики и оптики, доктор технических наук, профессор, зав. кафедрой, alive@d1.ifmo.ru

Никульский Игорь Евгеньевич – Ленинградский отраслевой НИИ связи (ФГУП ЛОНИИС), кандидат технических наук, доцент, начальник лаборатории, nikul@loniis.org

Пяттаев Владислав Олегович – Ленинградский отраслевой НИИ связи (ФГУП ЛОНИИС), кандидат технических наук, зам. генерального директора, pvo@loniis.org