

ПРИОРИТИЗАЦИЯ ТРАФИКА РЕАЛЬНОГО ВРЕМЕНИ В МУЛЬТИСЕРВИСНЫХ СЕТЯХ НА ОСНОВЕ РАССТОЯНИЯ

PRIORITIZATION OF REAL-TIME TRAFFIC IN MULTI-SERVICE NETWORKS ON THE BASIS OF DISTANCE

D. Khomenchuk

Summary. The article proposes mechanisms for equalizing delays for real-time traffic transmitted to different distances. It takes into account the propagation time of the signal from the source to the receiver, as well as the current delays when the packets pass through the network. It is proposed to divide all real-time traffic into three priorities depending on the transmission distance. A comparison bases on the FIFO model. A conclusion is made about the average waiting time for packets in case of a priority transmission and without priorities. The use of the priorities under consideration does not exclude the use of other mechanisms for providing QoS.

Keywords: Multiservice Network, Quality of Service (QoS), Real Time Service, Propagation delay, Serialization delay, Delay variation.

Хоменчук Дмитрий Валерьевич

Аспирант, Санкт-Петербургский Государственный
Университет Телекоммуникаций имени профессора
М. А. Бонч-Бруевича, Россия dhom@mail.ru

Аннотация. В статье предлагаются механизмы выравнивания задержек для трафика реального времени, передаваемого на разные расстояния. Учитывается время распространения сигнала от источника до получателя, а также текущие задержки при прохождении пакетов по сети. Предлагается разделить весь трафик реального времени по трем приоритетам в зависимости от расстояния передачи. Проводится сравнение с моделью FIFO, делается вывод о среднем времени ожидания для пакетов в случае беспriorитетной передачи и с приоритетами. Использование рассматриваемых приоритетов не исключает использование других механизмов обеспечения QoS.

Ключевые слова: Мультисервисная сеть, Качество обслуживания, Услуги реального времени, задержка распространения, время передачи, вариация задержки.

На сегодняшний день наблюдается значительное увеличение доли трафика реального времени в мультисервисных сетях. Широкое распространение получили приложения, использующие технологию VoIP (Voice over IP). Причиной этого является экономически более выгодное их использование по сравнению с обычной телефонией. К примеру, звонить за границу на много дешевле через приложения, использующие VoIP, а не обычные средства связи, так как взимается плата только за использование интернет соединения.

По прогнозам Nokia Bell Labs [1], в течение предстоящих пяти лет объем IP-трафика более чем удвоится и в 2022 году достигнет уровня в 330 эксабайт в месяц. Ежегодные темпы роста IP-трафика составят 25%.

В тоже время, согласно прогнозам Cisco (VNI 2017) [2], по доле в общем потоке IP-трафика и по общему приросту интернет-трафика по-прежнему будет доминировать видео (80% всего интернет-трафика к 2021 г., в 2016 г. этот показатель составлял 67%). К 2021 г. в мире будет около 1,9 млрд. пользователей интернет-видео (не считая тех, кто пользуется исключительно мобильной связью), тогда как в 2016 г. таких было 1,4 млрд. К 2021 г. через мировой Интернет в месяц будет передаваться 3 трлн. минут видео (эквивалентно 5 млн. видео-лет в месяц или 1 млн. видео-минут в секунду).

Так же изменится соотношение долей трафика в зависимости от расстояния передачи.

- ◆ К 2021 г. 35% интернет-трафика в мире будет замыкается внутри городских сетей, в 2016 г. доля такого трафика составляла 22%.
- ◆ К 2021 г. 23% интернет-трафика будет передаваться по региональным магистралям без захода в федеральные магистральные сети (в 2016 г. доля такого трафика составляла 20%).
- ◆ К 2021 г. 41% интернет-трафика в мире будет проходить по федеральным магистралям (в 2016 г. — 58%).

Службы реального времени, к которым относится голосовой трафик и видеоконференции, чувствительны к задержкам и её вариациям. В пакетных сетях задержка при передаче информации состоит из задержки на узле-отправителе, на узле-получателе и в транспортной сети. В дальнейшем будем рассматривать только задержку в транспортной сети. Она состоит из задержки распространения сигнала, времени обработки пакета каждым узлом, ожидания в очередях и передачи пакетов в канал на каждом узле сети. Задержки, связанные с распространением сигнала и передачей между интерфейсами можно считать постоянными задержками. Сумма задержек в очередях на всех сетевых узлах является переменной задержкой [3].

Согласно рекомендации G.114 [4], суммарная задержка в телефонной сети общего пользования не должна превышать 150 мс. Однако, при передаче трафика на большие расстояния, задержка распространения может достигать величин порядка 100 мс. В этом случае оставшегося резерва задержки может не хватить.

Для решения этой проблемы, весь трафик реального времени можно разделить на 3 типа по дальности передачи:

- ◆ передаваемый на малые расстояния (внутри городской сети);
- ◆ передаваемый на средние расстояния (уходящий в региональную магистраль);
- ◆ передаваемый на дальние расстояния (уходящий в федеральную магистраль).

Современные механизмы не учитывают это разделение при передаче трафика реального времени, в результате чего весь трафик, как на малые, так и на очень большие расстояния передается с одинаковым приоритетом. Это приводит к тому, что трафик внутри городских сетей передается с избыточно низкими задержками, в то время как трафик, проходящий по федеральным магистралям, может прийти с существенной задержкой.

Для решения данной проблемы на каждом сетевом узле можно одни потоки задерживать дольше, т.е. передавать с меньшим приоритетом, чем другие. Приоритетность потоков может определяться расстоянием, на которое передается поток. Т.е. трафик, передающийся на малые расстояния, будет иметь наименьший приоритет, в то время как трафик, передающийся на большие расстояния, будет иметь наибольший приоритет. При установлении соединения необходимо определить количество промежуточных узлов между отправителем и получателем, а также задержку распространения. Число промежуточных узлов определяется простым подсчетом узлов на пути следования. Чтобы определить задержку распространения, необходимо отметить время отправления пакета, а на каждом промежуточном узле к этому времени прибавлять задержку, имеющее место на данном узле. Величина смещения записывается в поле управления пакета. Начальное смещение на узле-отправителе равно 0. После прохождения пакета от отправителя к получателю, он передается в обратном направлении. В этом пакете отмечается количество промежуточных узлов N и задержки, полученных при прохождении пакета от отправителя до получателя. На обратном пути, в нем также отмечается количество промежуточных узлов и задержки. Если маршрут прохождения пакета в прямом и обратном направлении не меняется, то количество сетевых узлов будет одинаковым, а задержки могут быть разными. Зная время получения пакета, можем опреде-

лить задержку распространения. Для этого из общей задержки необходимо вычесть суммарную задержку на всех узлах, а полученное значение поделить пополам. Основываясь на полученном результате, можно оценить расстояние, на которое передается трафик и выставить соответствующий приоритет. В расчетах так же можно использовать количество сетевых узлов на пути следования.

Другим вариантом определения расстояния может быть использование иерархической адресации, по аналогии с телефонной сетью общего пользования, где адрес получателя (номер телефона) состоит из кода страны, кода города и номера абонента. Это может быть реализовано в IPv6.

Сравним эффективность предлагаемого решения со стандартной моделью без приоритетов. Пусть имеется система с одним обслуживающим устройством и бесконечной очередью. Пусть в систему поступают заявки M типов с интенсивностями $\lambda_1, \dots, \lambda_M$. Предположим, что каждый из входящих потоков заявок – пуассоновский и его суммарная интенсивность: $\Lambda = \sum_{i=1}^M \lambda_i$. Для вычислений в дальнейшем предполагается, что все чувствительные к задержке пакеты имеют фиксированный размер. Что позволяет использовать известную из теории массового обслуживания модель $M/D/1$. В таком случае время обслуживания ϑ_j можно принять за 1 [5].

Допустим, что в произвольный момент времени в систему поступает заявка с приоритетом k ($k=1, \dots, M$). Время ожидания такой заявки в очереди будет вычисляться как [6],[7]:

$$W_k = T_0 + \sum_{j=1}^k T_j + \sum_{j=1}^{k-1} T_j', \quad (1)$$

где время T_0 - время необходимое для завершения обслуживания ранее поступившей заявки, $\sum T_j$ - время всех ранее поступивших заявок с более высоким приоритетом, $\sum T_j'$ - время обслуживания заявок с более высоким приоритетом, поступившие в систему за время W_k и обслуживаемые раньше рассматриваемой заявки. Переход к математическим ожиданиям осуществляется следующим образом:

$$E[W_k] = \omega_k; \quad E[T_j] = \lambda_j \vartheta_j \omega_j = \rho_j \omega_j;$$

$$E[T_j'] = \lambda_j \vartheta_j \omega_k = \rho_j \omega_k$$

$$\text{Отсюда: } \omega_k = E[T_0] + \sum_{j=1}^k \rho_j \omega_j + \sum_{j=1}^{k-1} \rho_j \omega_k$$

$$\omega_k = \frac{E[T_0] + \sum_{j=1}^k \rho_j \omega_j}{(1 - R_{k-1})} \quad (2)$$

где $R_{k-1} = \rho_1 + \dots + \rho_{k-1}$ — суммарная загрузка системы потоками заявок с приоритетом не ниже $(k-1)$.

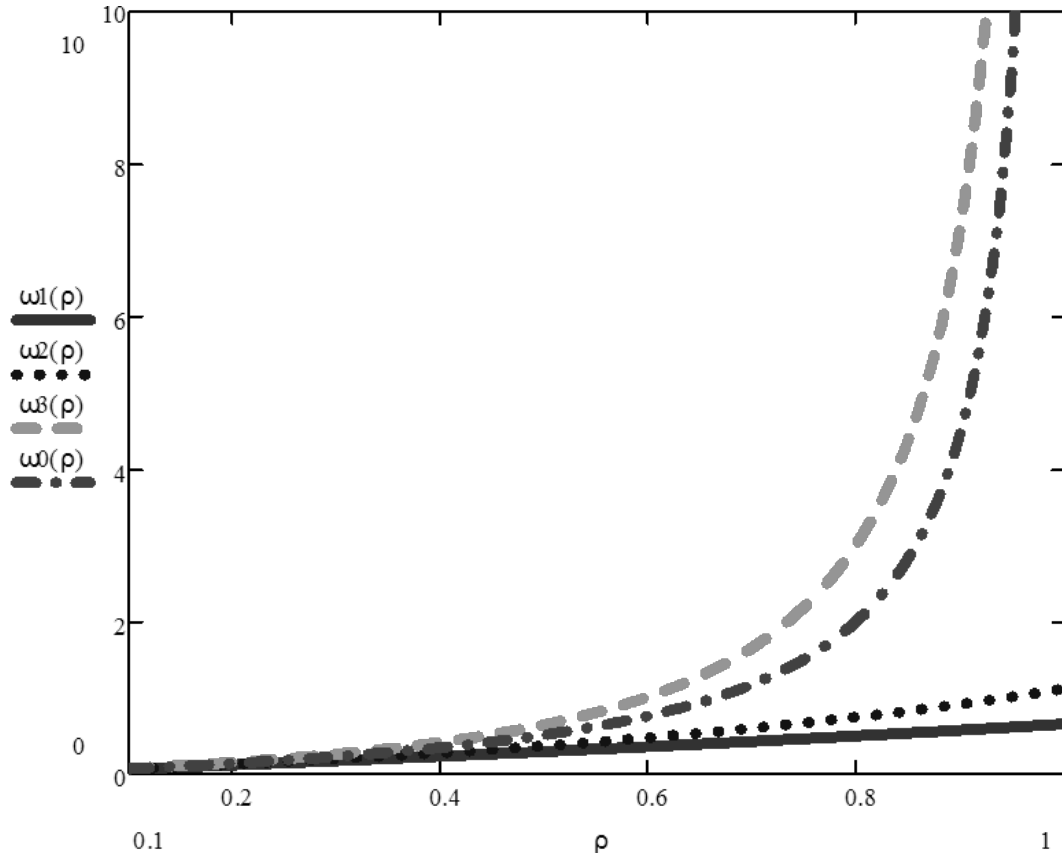


Рис.1. Зависимость среднего времени ожидания от загрузки для разных приоритетов (ω0 – без приоритетов, ω1-ω3 – с приоритетами)

Из (2) следует:

$$\omega_k = \frac{E'[T_0]}{(1 - R_{k-1})(1 - R_k)} \quad (3)$$

Подставив в (3) $E[T_0] = 0.5 \sum_{i=1}^M \lambda_i \vartheta_i$ получается среднее время ожидания заявок с приоритетами $k=1, \dots, M$:

$$\omega_k = \frac{0.5 \sum_{i=1}^M \lambda_i \vartheta_i^{(2)}}{(1 - R_{k-1})(1 - R_k)} \quad (4)$$

Выразив второй начальный момент через коэффициент вариации:

$$\omega_k = \frac{\sum_{i=1}^M \lambda_i \vartheta_i (1 + v_i^2)}{2(1 - R_{k-1})(1 - R_k)} \quad (5)$$

Так как время обслуживания постоянно, то $v_i=0$, а $\vartheta_i = 1$, получаем:

$$\omega_k = \frac{\sum_{i=1}^M \rho_i}{2(1 - R_{k-1})(1 - R_k)} \quad (6)$$

Известно, что для заявок без приоритетов среднее время ожидания в такой системе будет равно:

$$\omega_k = \frac{\sum_{i=1}^M \rho_i}{2(1 - R_k)} \quad (7)$$

Для сравнения двух решений, возьмем данные 2016 (Cisco VNI 2017) по процентным соотношениям трафика. Получим, что $\rho_1 = 0,22\rho$, $\rho_2 = 0,2\rho$, $\rho_3 = 0,58\rho$. Подставив в формулы 6 и 7 получим:

$$\omega_1 = \frac{\rho}{2(1 - 0,22\rho)}; \omega_2 = \frac{\rho}{2(1 - 0,22\rho)(1 - 0,42\rho)}$$

$$\omega_3 = \frac{\rho}{2(1 - 0,42\rho)(1 - \rho)}$$

$$\omega_0 = \frac{\rho}{2(1 - \rho)} \text{ — для системы без приоритетов.}$$

На рисунке 1 видно, что при высоких приоритетах среднее время ожидания уменьшается, в то время как с уменьшением приоритета время обслуживания увеличивается и для наименьших приоритетов становится больше чем в системе с бесприоритетным обслуживанием. Эта зависимость более наглядно отображена на рисунке 2.

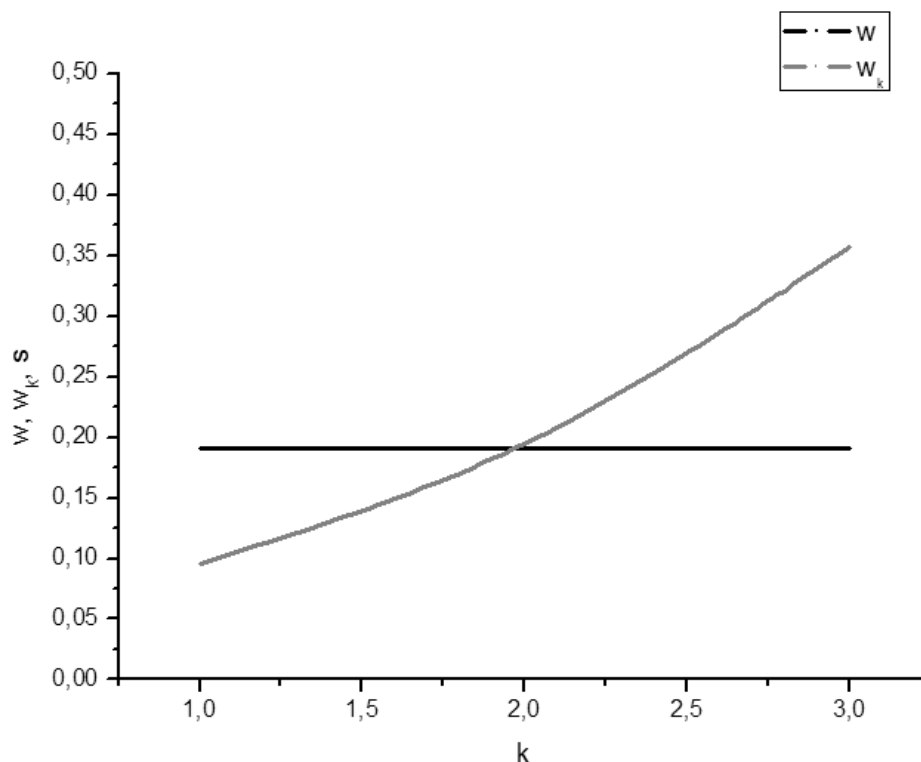


Рис. 2. Влияние относительных приоритетов на время ожидания заявок

Подводя итог, можно сказать, что приоритезация трафика по расстоянию позволяет уменьшить задержку при передаче на большие расстояния и увеличить на малые. В итоге общая задержка выровняется и станет прием-

лемой для трафика реального времени, передаваемого на разные расстояния. Следует отметить, что использование рассматриваемых приоритетов не исключает использование других механизмов обеспечения QoS.

ЛИТЕРАТУРА

1. Nokia Bell Labs 2017
2. Cisco (VNI 2017)
3. Олифер В. Г., Олифер Н. А. Компьютерные сети. Принципы, технологии, протоколы. Учебник для вузов. 5-е изд. — СПб.: Питер, 2016.
4. ITU-T Rec. G.114 (05/2003)
5. T-REC-Y.1541-200602-S!! PDF-R (02/2006)
6. Майоров С. А. Основы теории вычислительных систем (1978)
7. Шварц М. — Сети связи протоколы, моделирование и анализ. Ч. 1, 1992

© Хоменчук Дмитрий Валерьевич (vmdovgal@yandex.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»