

ВЛИЯНИЕ СПОСОБА ДИНАМИЧЕСКОЙ НАСТРОЙКИ РЕСУРСОВ НА ПАРАМЕТРЫ КАЧЕСТВА ПЕРЕДАЧИ ВИДЕО ИНФОРМАЦИИ

Курапов А.С.

Московский физико-технический институт
141700, М.О., г. Долгопрудный, Институтский пер., д. 9, кафедра ИСС

С внедрением новых телекоммуникационных технологий, одной из которых является передача оцифрованной видео информации через сети с коммутацией пакетов, появилась необходимость развития систем управления передачей пакетов данных. Основной функцией этих систем является выбор из всех пакетов данных, находящихся в буферном пространстве порта сетевого оборудования, того пакета, который должен быть отправлен в следующем цикле [1]. Этот выбор должен быть осуществлен так, чтобы при это были обеспечены требования к параметрам качества передачи каждого потока информации.

Одним из наиболее часто используемых систем управления передачи пакетов данных является способ WFQ (Weighted Fair Queuing - Взвешенная Справедливая Очередность) [2]. В данном способе каждому потоку информации выделена определенная полоса пропускания, которая в случае отсутствия в буферном пространстве порта сетевого оборудования пакетов данных какого либо потока информации делится между присутствующими потоками информации пропорционально выделенным им полосам пропускания.

С целью повышения степени использования ресурсов сетевого оборудования автором данного доклада был разработан способ управления передачи пакетов данных DRA (Dynamic Resource Adjustment – Динамическая Настройка Ресурсов), основанный на способе WFQ, отличающийся тем, что выделенные потокам ресурсы изменяются в соответствии с текущими условиями передачи трафика.

Для одного узла сети требования к качеству передачи потока i характеризуются двумя параметрами $(D_i^{total}, L_i^{total})$, где D_i^{total} является границей допустимой задержки, а L_i^{total} – допустимым обобщенным количеством потерянных пакетов, включающим в себя потери, вызванные переполнением буфера и превышением максимальной величины задержки. Таким образом, предполагается, что если передача пакета превысила задержку D_i^{total} , то пакет считается потерянным.

Спецификация ресурсов узла сети определяется параметрами $(B_i^{allocated}, R_i^{allocated})$, где $B_i^{allocated}$ является размером буфера, выделенного потоку i , а $R_i^{allocated}$ – частью общей полосы пропускания сетевого оборудования, выделенной этому потоку. Для того чтобы обеспечить отправку всех пакетов, находящихся в части буфера $B_i^{allocated}$, в рамках задержки $D_i^{allocated}$, было использовано соотношение: $B_i^{allocated} = D_i^{allocated} \cdot R_i^{allocated}$ [3].

Перед началом передачи трафика, устанавливаются максимальное и минимальное значения занимаемой части выделенного потоку буфера – P_{max} и P_{min} .

Работа способа DRA реализована в соответствии со следующей схемой:

- 1) В случае, если после получения пакета данных потока информации i , занимаемое им буферное пространство $B_i^{current} \geq B_i^{allocated}$, а потеря этого пакета приведет к превышению допустимых потерь

L_i^{total} , то $B_i^{allocated}$ увеличивают на величину $\min\left\{\frac{1-P_{max}}{P_{max}} \cdot B_i^{allocated}, B - \sum_{i=1}^N B_i^{allocated}\right\}$, где N –

общее количество передаваемых потоков информации;

- 2) Из буферного пространства сетевого оборудования отправляется пакет, принадлежащий потоку информации i с наименьшим виртуальным завершающим временем F_i , вычисленным в соответствии со спецификацией способа WFQ.

- 3) Если после отправления пакета данных потока информации i , занимаемое им буферное пространство $B_i^{current} \leq P_{max} \cdot B_i^{allocated}$ в течение времени T_c , то $B_i^{allocated}$ уменьшают на величину

$\frac{1-P_{max}}{P_{max}} \cdot B_i^{allocated}$.

- 4) После любых изменений $B_i^{allocated}$ осуществляют преобразование выделенной потоку i полосы пропускания $R_i^{allocated}$ в соответствии с выражением $R_i^{allocated} = B_i^{allocated} / D_i^{total}$.

Параметры P_{max} и P_{min} необходимы для контроля степени использования ресурсов, выделенных потокам информации, а T_c - для избежания настройки ресурсов, вызванных лишь кратковременных изменением плотности потока информации.

Таким образом, способ DRA позволяет, во-первых, контролировать и значение допустимых потерь, и величину требуемой задержки, а во-вторых, динамически настраивать распределение ресурсов в соответствии с текущими величинами параметров качества передачи каждого потока информации.

Для изучения характеристик алгоритма динамической настройки ресурсов DRA были проведены исследования, в которых в качестве приложения реального времени использовалась передача видео потоков с кодировкой MPEG-1. Данные потоки имели частоту образования видео кадров, равную 30 кадров/сек. Трафики этих потоков характеризовались двумя различными функциями распределения I, P и B кадров. Структура потока первого типа имела почти правильную периодическую форму, при этом видео сигналу второго типа соответствовали резкие изменения амплитуды, имеющие различную временную длительность.

Во всех исследованиях, описанных в данном докладе, в качестве сетевого моделирующего устройства использовалась программа NS-2 (Network Simulator-Version 2) [4].

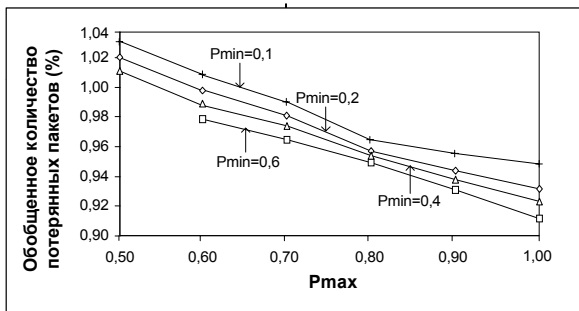


Рис. 1.

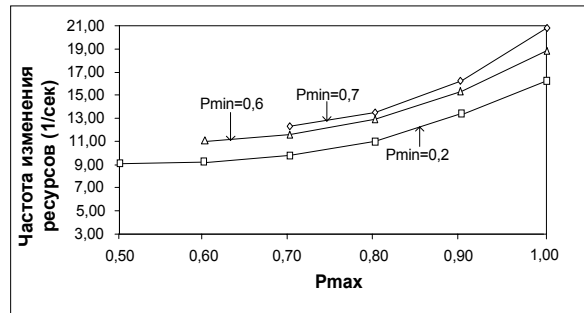


Рис. 2.

В исследованиях, направленных на нахождение возможного диапазона пороговых параметров P_{max} и P_{min} , было установлено, что при увеличении P_{max} и P_{min} уменьшается обобщенное количество потерянных пакетов (Рис. 1) и увеличивается частота изменения ресурсов (Рис. 2), причем наиболее значительным является влияние параметра P_{max} . В соответствии с полученными зависимостями оптимальными значениями параметров P_{max} и P_{min} являются диапазоны от 0,7 до 0,9 и от 0,2 до 0,6 соответственно.

Аналогичны исследования, проведенные для временного параметра T_c (Рис. 3 и 4), показали, что оптимальным значением этого параметра является диапазон от 0,030 до 0,035 с, что приблизительно равно периоду формирования видео кадров – 0,033 с.

Теоретическим обоснованием этого, является то, что значение параметра T_c , равное среднему периоду, в течение которого скорость потока i приблизительно постоянна, позволяет удовлетворить следующим требованиям:

- 1) Данное значение T_c является достаточно большим для избежания настройки ресурсов, вызванной лишь кратковременным изменением скорости потока i .
- 2) Данное значение T_c является достаточно малым, для того, чтобы уменьшить выделенное потоку i буферное пространство $B_i^{allocated}$, если оно в течение длительного времени больше, чем $B_i^{current}$.

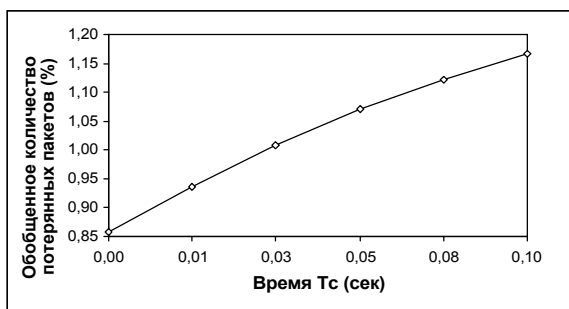


Рис. 3.

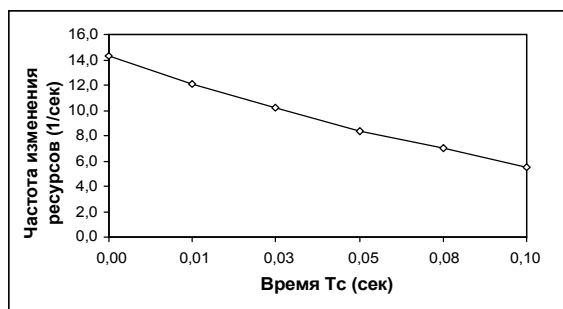


Рис. 4.

Таким образом, оптимальное значение параметра равняется длине временного интервала, в течение которого скорость потока приблизительно постоянна.

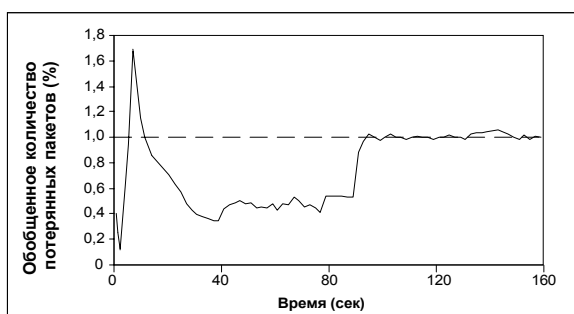


Рис. 5.

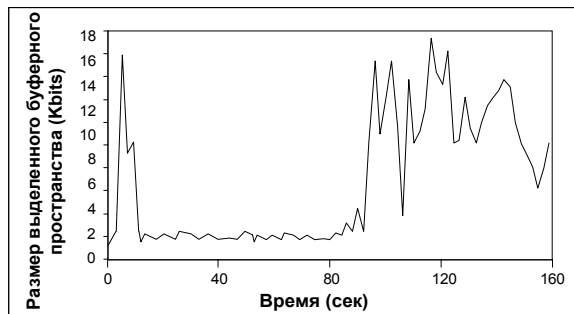


Рис. 6.

Также были проведены моделирования, в которых получено качественное подтверждение уменьшения нагрузки на вычислительные ресурсы сетевого оборудования за счет разделения незанятых ресурсов между всеми потоками, что обеспечивается за счет того, что увеличение выделенного потока буферного пространства (Рис. 6) происходит только, если потеря получаемого пакета приведет к превышению допустимой величины потерянных пакетов (Рис. 5). Данные исследования показали, что возможность занятия остаточных ресурсов уменьшает количество настроек выделенных потокам ресурсов, которые обусловлены всего лишь кратковременным изменением плотности трафика.

Для оценки функциональных возможностей динамического регулирования ресурсов были проведены моделирования, в которых сравнивались характеристики способа DRA и статического варианта WFQ (Static WFQ). Эти исследования состояли из анализа влияния этих способов на общее количество потерянных пакетов, а также на величины потерь, обусловленных переполнением буферного пространства порта сетевого оборудования и превышением допустимой задержки передачи пакета.

Результаты данных исследований показали, что для всех значений размера буфера общее количество потерянных пакетов и величина потерь, обусловленных переполнением буфера, меньше для DRA, чем для способа Static WFQ. Однако в отношении потерь, вызванных превышением максимальной задержки, при достаточно малых размерах буфера алгоритм динамической настройки ресурсов показывает более плохие результаты.

Первый результат может быть объяснен тем, что общее разделение буферного пространства сетевого оборудования между всеми потоками информации позволяет значительно уменьшить требования к ресурсам, за счет возможности занятия потоком свободных ресурсов. Однако данная характеристика способа DRA является причиной второго результата, который обусловлен тем, что использование свободных ресурсов приводит к нарушению соотношения $B_i = D_i \cdot R_i$, обеспечивающего отправку всех пакетов находящихся в очереди в рамках допустимой задержки.

Данное тестирование также подтвердило рост утилизации ресурсов и увеличение количества потоков, для которых выполнены требования к параметрам качества сервиса (Рис. 7 и 8), причем этот рост особенно значителен для трафиков с большими изменениями плотности потока (Рис. 8). Это доказывает возможность эффективного использования алгоритма DRA для потоков, структура которых предварительно неизвестна или сильно меняется во времени.

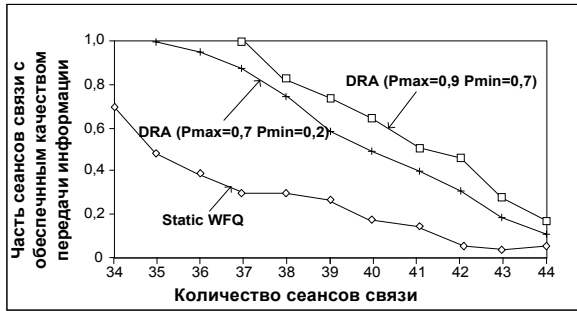


Рис. 7.

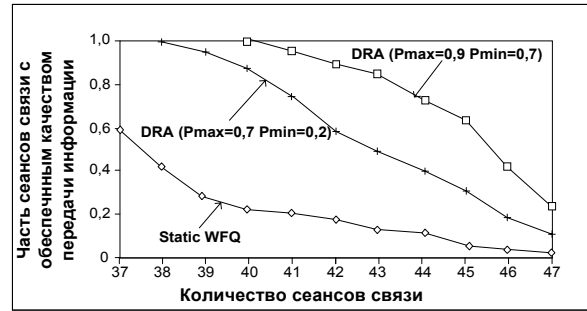


Рис. 8.

Литература

1. Y. Bao - Efficient Resource Management for QoS Guarantees, University of Delaware, Technical Report № 5, 1997
2. O. Abuamsha - Comparison of Fair Queuing Algorithms, In Proc. IEEE INFOCOM'97, volume 2, Paris, 1997
3. D. Stiliadis and A. Varma, - Frame-based Fair Queuing: A New Traffic Scheduling Algorithm for Packet-Switched Networks, Technical Report UCSC-CRL-95-39, 1995
4. K. Fall, K. Varadhan - NS Notes and Documentation, UCB Multicast Network Research Group, April, 1998



THE EFFECT OF DYNAMIC RESOURCE ADJUSTMENT SCHEDULING METHOD ON VIDEO INFORMATION TRANSMISSION PARAMETERS

Kurapov A.S.

Moscow Institute of Physics and Technology
 Institutsky per., 9, Dolgoprudny, Moscow Region, 141700, Russia

Real-time multimedia application communicating over a packet-switched network, e.g. video information transmission, presents new challenges to packet scheduling methods. The function of these methods is to select the packet to be transmitted in the next cycle from the available packets belonging to the communication sessions sharing the output link [1]. This selection must be performed such that Quality of Service (QoS) parameters guaranteed for the individual traffic sessions, e.g. upper bounds on maximum delay, are satisfied.

One of the most popular scheduling methods is WFQ (Weighted Fair Queuing) [2]. This method is a rate-based QoS control algorithm that schedules the flows according to their allocated fixed shares of the bandwidth.

In order to achieve high resource utilization of network equipment the author designed a Dynamic Resource Adjustment (DRA) used in combination with WFQ algorithm. The novel idea of DRA is to adjust the resource shares based on the QoS performance of each data flow.

The request QoS parameters of flow i is modeled by QoS tuple $(D_i^{total}, L_i^{total})$. D_i^{total} is delay bound and L_i^{total} is the requested loss ratio tolerance. It is assumed that these parameters are pertinent to a single network node. There are two situations that could cause the loss of a packet:

- 1) A newly-arrived packet observing a full buffer space;
- 2) A packet experiencing a delay that is longer than its delay bound D_i^{total} ;

The considered resources include buffer space bandwidth pertaining to an outgoing link of a network node. Tuple $(B_i^{allocated}, R_i^{allocated})$ is used to represent the resources allocated to flow i to satisfy its specific QoS parameters, in which $B_i^{allocated}$ refers to the allocated buffer size and $R_i^{allocated}$ refers to the allocated bandwidth (service rate).

The deadlines of the packets are enforced by maintaining $(B_i^{allocated}, R_i^{allocated})$ under the relationship $B_i^{allocated} = D_i^{allocated} \cdot R_i^{allocated}$ [2]. Because of this relationship, if flow i is guaranteed a bandwidth $R_i^{allocated}$, then any packets that is admitted into the allocated buffer will be guaranteed to be serviced within its delay bound D_i^{total} .

Before the start of transmission, two buffer occupancy thresholds are chosen: P_{max} (the maximum threshold) и P_{min} (the minimum threshold).

The following is DRA algorithm for resource adjustment:

- 1) If after packet arrival of flow i the current occupied buffer of flow i $B_i^{current}$ is larger than $B_i^{allocated}$ and losing this arrival will cause the loss ratio performance to be worse than L_i^{total} , then increase $B_i^{allocated}$ by
$$\min \left\{ \frac{1 - P_{max}}{P_{max}} \cdot B_i^{allocated}, B - \sum_{i=1}^N B_i^{allocated} \right\}$$
, where N is the total number of data flows;
- 2) Serve the packet that belongs to the data flow i with minimal finishing time F_i that is computed according to WFQ algorithm.
- 3) If after packet transmission of flow i the current occupied buffer of flow i $B_i^{current}$ is consistently lower than $P_{max} \cdot B_i^{allocated}$ for a period T_c , then decrease $B_i^{allocated}$ by $\frac{1 - P_{max}}{P_{max}} \cdot B_i^{allocated}$;
- 4) After each changes of allocated buffer space $B_i^{allocated}$ perform $R_i^{allocated} = B_i^{allocated} / D_i^{allocated}$;

The parameter P_{max} is used to increase the resources when a buffer of flow i is full and an increase of resources is needed. The parameter P_{min} is used to detect low buffer usage. The parameter T_c allows to reduce the frequency of resource adjustment.

Simulations for multiple real-time video MPEG-1 sources were conducted to study the significance of different values of P_{max} , P_{min} , T_c parameters. In this simulation NSs-2 (Network Simulator – 2) was used.

This simulation shows that the larger P_{max} and P_{min} values results in larger frequency of resource adjustment and smaller loss ratio performance. The impact of different P_{max} values is more pronounced. With reference to T_c the result of this simulation shows that the larger the value of T_c parameter, the lower the frequency of resource adjustment and the worse QoS performance. Generally the value of T_c parameter should be appropriately equal to the time period in which the source has a relatively stable rate, e.g. MPEG source that has frame rate is 30 frames/sec this time period is 0.033 sec.

In order to evaluate the effect of dynamically adjusting the resource shares among the flows, the simulation was conducted to compare the achieved QoS under DRA and Static WFQ algorithms. The QoS satisfaction ratio decreases under both algorithms when the total number of flows sharing the resources increases. DRA presents better than Static WFQ by presenting a higher fraction of flows with achieved QoS under the same number of participating flows. This increase is more significant for video flows that have less predictable (non-repetitive) traffic pattern.

References

1. Y. Bao - Efficient Resource Management for QoS Guarantees, University of Delaware, Technical Report № 5, 1997
2. O. Abuamsha - Comparison of Fair Queuing Algorithms, In Proc. IEEE INFOCOM'97, volume 2, Paris, 1997
3. D. Stiliadis and A. Varma, - Frame-based Fair Queuing: A New Traffic Scheduling Algorithm for Packet-Switched Networks, Technical Report UCSC-CRL-95-39, 1995
4. K. Fall, K. Varadhan - NS Notes and Documentation, UCB Multicast Network Research Group, April, 1998