

Ю.И. ЛОСЕВ¹, В.И. БОРОВОЙ¹, В.Е. КУРИЛКО²

¹ *Харьковский военный университет, Украина*

² *Харьковский институт Военно-Воздушных Сил им. Ивана Кожедуба, Украина*

МЕТОДИКА ПРИСВОЕНИЯ ПРИОРИТЕТНОСТИ ОБСЛУЖИВАНИЯ СООБЩЕНИЙ В СЕТИ ОБМЕНА ДАННЫМИ

Рассматриваются вопросы, связанные с обеспечением заданного качества обслуживания (QoS) для различных классов трафика. Показаны механизмы обеспечения заданного QoS при использовании IP - протокола. Вводятся понятия ценности и старения информации и методы их количественной оценки

сеть обмена данными, приоритетность обслуживания сообщений, ценность информации, оптимизация пропускной способности, полоса пропускания

Постановка проблемы

При управлении сетью обмена данными (СОД) сложной динамической системы, к которым несомненно относятся группировки различных видов вооруженных сил, основной является задача не только восстановления ее работоспособности, но и обеспечения необходимого качества обслуживания потребителей. Знание основных концепций и возможностей технологий QoS позволит максимально оптимизировать производительность сетей и обеспечить стабильное функционирование нового поколения мультимедийных и голосовых приложений.

Анализ литературы

Вопросы, связанные с обеспечением заданного качества обслуживания рассматриваются в ряде работ [1-3]. Так, в [1] описаны два подхода к обеспечению качества работы сети, причем, если в первом случае сеть (например, frame relay или ATM) гарантирует пользователю заданный уровень пропускной способности, то второй подход (best effort) основан на том, что сеть старается, по возможности, более качественно обслужить пользователя, но ничего при этом не гарантирует. Работы [2, 3] посвящены вопросам реализации технологий QoS в сетях IP. Однако в перечисленных публикациях не учитывается тот факт, что информация имеет свою ценность,

причем с увеличением времени задержки в передаче того или иного вида трафика эта ценность уменьшается, т.е. происходит старение информации.

Цель статьи

Разработка методики присвоения приоритетности обслуживания сообщений в сети обмена данными с учетом их ценности и старения информации.

Методика присвоения приоритетности обслуживания сообщений в СОД

Качество обслуживания (QoS) – это способность сетевых средств обеспечить требуемый сервис для определенных классов трафика. Основное назначение технологий QoS - обеспечить приоритет, контролируемую задержку и вариацию этой задержки для определенного вида трафика. При этом важно обеспечить приоритетность обслуживания для какого-либо потока с возможностью передачи других потоков данных. Технологии QoS включают широкий спектр средств, которые при правильном их применении могут повысить эффективность использования выделенной полосы пропускания.

Такие технологии, как ATM и Frame Relay, изначально предусматривают использование механизмов QoS. Поэтому органы стандартизации большое внимание уделяют IP, Ethernet и другим технологиям, требующим поддержки качества сервиса.

В этой области IEEE разрабатывает два направления. Одно направление - Diffserv - позволяет внедрить QoS в среде, не ориентированной на соединение. Основной задачей является стандартизация наборов блоков для потребителя QoS, используя которые, провайдеры могут улучшить IP-сервис. Предполагается внедрить QoS-механизмы в сетях с помощью устройств доступа и последующую его транспортировку по магистрали с помощью маршрутизаторов, поддерживающих Diffserv. Второе направление - MPLS-технология - поддерживает QoS-механизмы, но влечет за собой существенное изменение в архитектуре сети.

Базовая архитектура QoS включает в себя механизмы идентификации и маркировки, предназначенные для координирования QoS между окончными сетевыми элементами, и QoS в пределах одного элемента сети (например, для организации очередей). Идентификация и маркировка трафика предназначена для обеспечения предпочтительного обслуживания определенного типа трафика. Трафик может быть идентифицирован по значениям полей заголовка (MAC-адрес, IP-адрес), по TOS-байту в IP-заголовке, по типу приложений и т. п.

Способы присвоения приоритетов могут быть явными и неявными. При неявном присвоении приоритетов обслуживающий прибор автоматически присваивает пакетам приоритет исходя из некоторых критериев, например, адреса. При явной приоритизации запрашивается определенный уровень службы и маршрутизатор пытается удовлетворить запрос. Самым популярным механизмом в этом случае является протокол IPTOS. Он резервирует ранее неиспользуемое поле TOS в заголовке пакета IP, где указываются принципы качества обслуживания (время задержки, скорость передачи). Три первых бита этого поля (0-2) позволяют установить восемь уровней приоритетов. Например, 100-сверхсрочный, 011-срочный, 010-неотложный, 001-приоритетный, 000-обычный.

Биты 3, 4 и 5 выделены для указания трех классов обслуживания:

бит 3 - задержка: 0 - нормальная; 1 - низкая;

бит 4 - пропускная способность: 0 - нормальная; 1 - высокая;

бит 5 - надежность: 0 - обычная; 1 - высокая.

Биты 6 и 7 зарезервированы для будущего использования. После принятия документа RFC 1349 биты 3, 4, 5 и 6 стали рассматриваться как единое целое и называться полем TOSS. Содержание этого поля означает: 1000 - с низкой задержкой; 0100 - с высокой пропускной способностью. Разница указанных вариантов заключается в следующем: первый предназначен для указания приоритета конкретной датаграммы и учитывается при обслуживании очередей, а второй определяет соотношение между пропускной способностью, задержкой и надежностью.

При обслуживании пакета с различными приоритетами организуются очереди, представляющие собой области буферного запоминающего устройства, в которых группируются заявки с одинаковыми приоритетами. Необходимо обеспечить наилучшее обслуживание трафика с более высокими приоритетами при условии, что пакету с низким приоритетом гарантируется обслуживание. В этих условиях более эффективной является дисциплина пропорционального обслуживания. При этом каждому потоку присваивается некоторый вес. Этот вес, а, следовательно, и приоритет можно оценить численно.

Пусть обслуживается N потоков информации. Сообщения каждого потока могут обслуживаться сразу при поступлении или с некоторой задержкой. Если i -ый поток не обслуживается, то происходит потеря I_{nom} количества ценной информации. При задержке в обслуживании на $t_{ож.і}$ ценная информация также теряется за счет ее старения.

Следовательно, среднее количество потерянной ценной информации будет вычисляться по формуле:

$$\overline{I_{nom}} = \sum_{i=1}^N I_i X_i + \sum_{i=1}^N I(t_{ож_i}) (1 - X_i) \quad (1)$$

где I_i – количество информации в i -м сообщении;

X_i – параметр обслуживания ($X_i = 1$, если сообщение не обслуживается; $X_i = 0$, если сообщение обслуживается);

$I(t_{ож_i})$ – количество ценной информации в i -м сообщении в зависимости от времени ожидания его обслуживания.

Под ценностью информации будем понимать пользу, которую данное количество информации может принести для решения поставленной задачи, т.е. ценность информации должна определяться численным значением увеличения вероятности выполнения поставленной задачи.

Обозначим через P_{0i} вероятность решения задачи при отсутствии сообщений i -го типа. Тогда неопределенность при решении этой задачи равна:

$$I_{0i} = -\log P_{0i}. \quad (2)$$

При получении сообщений i -го типа вероятность решения задачи будет равна P_{1i} , а неопределенность этого решения равна:

$$I_{1i} = -\log P_{1i}. \quad (3)$$

В результате величина снятой неопределенности, т.е. количество полученной информации равно:

$$I_{iy} = I_{0i} - I_{1i} = \log \frac{P_{1i}}{P_{0i}}. \quad (4)$$

Т.к. количество информации связано с поставленной целью, т.е с решением задачи, то величина, вычисленная по выражению (4), и будет определять количество ценной информации.

Понятие старение информации неразрывно связано с понятием ее ценности. Под старением информации понимают уменьшение со временем ее ценности. Поэтому старение информации $I(t_{ож_i})$ можно определить, если известна зависимость от времени вероятности решения поставленной задачи $P_{1i}(t)$. Эта зависимость $P_{1i}(t)$ определяется содер-

жением конкретной задачи. Однако можно указать и некоторые общие требования, которым должна удовлетворять функция $P_{1i}(t)$. Это следующие требования:

1) $P_{1i}(t) = P_{1i}$ при $0 \leq t \leq t_k$, где t_k – допустимое время, в течение которого ценность передаваемой информации не уменьшается;

2) $P_{1i}(t) = P_{0i}$ при $t > \Delta T_k$, где ΔT_k – интервал времени, по истечении которого информация, содержащаяся в сообщении считается полностью устаревшей;

3) при $t_k < t \leq \Delta T_k$ вероятность $P_{1i}(t)$ должна уменьшаться по некоторой функции от P_{1i} до P_{0i} . С некоторым приближением для многих информационных процессов эта функция может быть линейной.

Перечисленным требованиям отвечает функция вида:

$$P_{1i}(t) = \begin{cases} P_{1i} & \text{при } 0 \leq t \leq t_k; \\ (P_{1i} - P_{0i}) \left(1 - \frac{t - t_{ki}}{\Delta T_{ki} - t_{ki}} \right) + P_{0i}, & \text{при } t_k < t \leq \Delta T_k; \\ P_{0i} & \text{при } t > \Delta T_k. \end{cases} \quad (5)$$

Тогда ценность информации в сообщении i -го типа в зависимости от времени его задержки при передаче $t_{ож_i}$ будет определяться следующим соотношением:

$$I(t_{ож_i}) = \log \frac{P_{1i}(t)}{P_{0i}} = \log \left[\frac{P_{1i}}{P_{0i}} \left(1 - \frac{t - t_{ki}}{\Delta T_{ki} - t_{ki}} \right) + \frac{t - t_{ki}}{\Delta T_{ki} - t_{ki}} \right] \quad (6)$$

при $t_k < t \leq \Delta T_k$.

Количество потерянной ценной информации за счет ожидания обслуживания i -го сообщения будет равно:

$$I_{nom}(t_{ожз}) = I_i - I_i(t) = \log \frac{P_{li}}{P_{0i}} - \log \frac{P_{li}(t)}{P_{0i}} =$$

$$= \log \left[\frac{P_{li}}{P_{li} \left(1 - \frac{t-t_{ki}}{\Delta T_{ki}-t_{ki}} \right) + \frac{t-t_{ki}}{\Delta T_{ki}-t_{ki}} P_{0i}} \right] \quad (7)$$

Среднее же значение потерянной ценной информации определяется выражением (1).

Необходимо обеспечить такое приоритетное обслуживание, чтобы ее величина была минимальной. При небольшом числе разнообразных потоков такая задача может быть решена методом перебора.

Для уменьшения времени решения оптимизационной задачи нужно определить два пороговых значения отношения $\frac{P_{li}}{P_{0i}}$ (порог А и порог В). Если

для i -го сообщения эта величина больше порога А, то $X_i = 0$ и сообщение обслуживается. Сообщения, для которых эта величина меньше порога В, не обслуживаются и принимается $X_i = 1$.

Для сообщений с отношением $V < \frac{P_{li}}{P_{0i}} < A$ приоритетность обслуживания определяется по критерию минимума количества потерянной ценной информации $\bar{I}_{пот}$.

Таким образом, методика присвоения приоритетности обслуживания в СОД заключается в следующем:

1. Определяются вероятности решения задачи при отсутствии P_{0i} и при поступлении P_{li} сообщений i -того типа в соответствии с выражениями (2) и (3).

2. Определяется количество полученной ценной информации в соответствии с выражением (4).

3. Определяется ценность информации в сообщении i -го типа в зависимости от времени его задержки при передаче $t_{ожз_i}$ в соответствии с выражением (6).

4. Определяется количество потерянной ценной

информации за счет ожидания обслуживания i -го сообщения в соответствии с выражением (7).

5. Определяются пороговые значения отношения $\frac{P_{li}}{P_{0i}}$.

6. Для случая $\frac{P_{li}}{P_{0i}} > A$ сообщения обслуживаются, а для случая $\frac{P_{li}}{P_{0i}} < B$ – не обслуживаются.

7. Для случая $B < \frac{P_{li}}{P_{0i}} < A$ приоритетность обслуживания определяется по критерию минимума количества потерянной ценной информации $\bar{I}_{пот}$.

8. Определяется среднее значение количества потерянной ценной информации в соответствии с выражением (1).

Выводы

Таким образом, в данной работе предложена методика присвоения приоритетности обслуживания сообщений в СОД с учетом ценности и старения информации. При этом обеспечивается заданное качество обслуживания потребителей и оптимизация пропускной способности сети для различных видов трафика.

Литература

1. Компьютерные сети. Принципы, технологии, протоколы / В.Г. Олифер, Н.А. Олифер. – СПб.: Питер, 2002. – 672 с.

2. Вегешна, Шринивас. Качество обслуживания в сетях IP.: Пер. с англ. – М.: Издательский дом “Вильямс”, 2003. – 368 с.

3. Олифер В.Г., Олифер Н.А. Новые технологии и оборудование IP-сетей. – СПб.: БХВ-Петербург, 2001. – 512 с.

Поступила в редакцию 02.10.03

Рецензент: д-р техн. наук, профессор Поповский В.В., Харьковский национальный университет радиоэлектроники, г. Харьков