

МОДЕЛЬ ОБЕСПЕЧЕНИЯ КАЧЕСТВА ОБСЛУЖИВАНИЯ В СЕТЯХ ПЕРЕДАЧИ ДАННЫХ

Г.А. Абрамян;

А.С. Крутолапов, кандидат технических наук, доцент.

Санкт-Петербургский университет ГПС МЧС России.

С.М. Одоевский, доктор технических наук, профессор.

Военная академия связи имени С.М. Буденного, Санкт-Петербург

Предложена модель управления процессом обработки заявок с поддержкой качества обслуживания в сети передачи данных автоматизированной системы диспетчерского управления средствами пожарной сигнализации и пожаротушения.

Ключевые слова: сеть, передача данных, автоматизированная система, диспетчерское управление, заявка, приоритет, информация, качество обслуживания

QUALITY OF SERVICE ASSURANCE MODEL IN DATA NETWORKS

G.A. Abramian; A.S. Krutolapov.

Saint-Petersburg university of State fire service of EMERCOM of Russia

S.M. Odoevsky. Saint-Petersburg Military academy of telecommunications of S.M. Budenny.

A control model for editing applications process with support quality of service in data networks of automated dispatch control of fire alarm and extinguishing systems is offered.

Key words: network, data transmission, an automated system, dispatch control, application, priority, information, quality of service

Активные исследования в области обеспечения качества обслуживания в сетях передачи данных (СПД) начались с середины 80-х гг., потребность в подобных средствах проявилась с началом роста СПД автоматизированных систем диспетчерского управления (АСДУ). Функционирование таких сетей базируется на сквозном режиме обслуживания пакетов данных без ориентации на установку соединения, что делает АСДУ более гибкой и устойчивой к сбоям. Динамика передаваемых потоков данных приводит сети к перегрузкам, которые чаще всего возникают в местах стыка двух сетей со значительно различающимися пропускными способностями.

Начнем с классификации заявок на информационное обслуживание в СПД. Целью классификации является создание системы приоритетов по степени важности оперативности обработки заявки. Степень важности оперативности обработки заявки определяется двумя факторами.

Во-первых, какая информация должна быть передана (насколько эта операция важна для принятия решения).

Во-вторых, кто из пользователей сделал запрос (насколько должностные обязанности пользователя важны для принятия решения).

Существующие классификации в различных информационных системах ставят своей целью выделение классов пользователей для назначения прав доступа к сервисам и данным. При реализации системы обеспечения качества обслуживания в каждом конкретном случае необходим анализ предметной области и выделение соответствующих классов

пользователей. Можно предложить достаточно общий подход, который может быть адаптирован в частных случаях.

Одной из перспективных технологий разграничения доступа в сложных СПД является технология ролевой авторизации [1]. Модель (политика) ролевого разграничения доступа основывается на функциях, которые каждый пользователь должен выполнять в соответствующей системе. Реализация такой политики требует разработки набора необходимых ролей (библиотеки ролей), соответствующей СПД. Роль определяется как набор прав доступа к объектам системы и прав на выполнение над ними определенного набора элементарных действий, соответствующих функций, которые должен выполнять в системе определенный пользователь.

С точки зрения важности для функционирования любой СПД можно выделить три класса пользователей:

- потребители информационных услуг, использующие предоставляемые ресурсы и сервисы для исполнения своих должностных обязанностей;
- производители информационных услуг, обеспечивающие создание и поставку информационных ресурсов, необходимых для принятия решения;
- администрация системы, обеспечивающая сбор и анализ информации об информационной среде, ее сопровождении и развитии.

Функции администрации СПД превалируют над остальными по критерию срочности. Однако необходимо отметить ряд особенностей административных функций:

- служебная информация имеет малый вес в нормальных условиях функционирования;
- подавляющая часть административных функций реализуется с помощью специализированного программного обеспечения, не входящего в состав сервисов, управление которыми осуществляется в подсистеме;
- работы, требующие значительного трафика (например, репликация каталогов распределенных баз данных) всегда проводятся в ночное время.

Графическое представление классификации представлено на рис. 1.

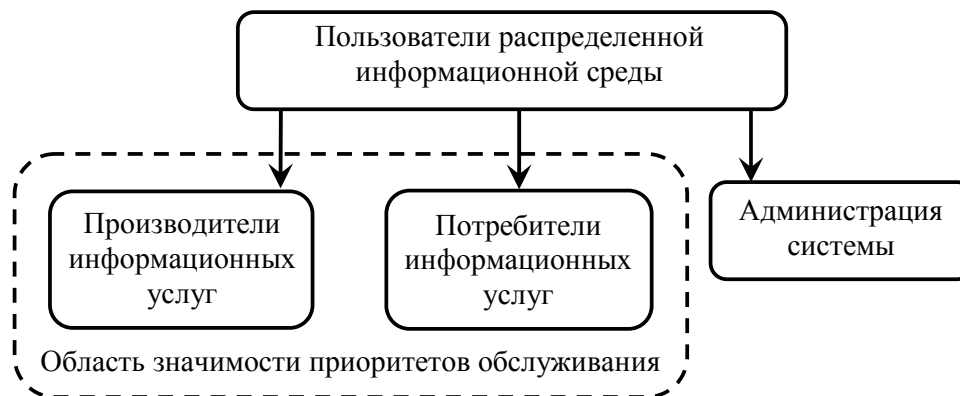


Рис. 1. Классификация пользователей СПД

Один и тот же физический пользователь может выступать и как производитель, и как потребитель информации в зависимости от выполняемых функций.

Отличительной особенностью СПД является неоднородность их среды, поэтому нецелесообразно пытаться перечислить возможные виды активностей пользователей и, тем более, назначить им приоритеты.

Такой подход также негативно отразится на масштабируемости сети в случае добавления новых сервисов. Естественным выглядит подход, основанный на выделении (по

рассматриваемому критерию) классов пользователей и классов видов деятельности пользователей. Общий вид такой классификации представлен на рис. 2.

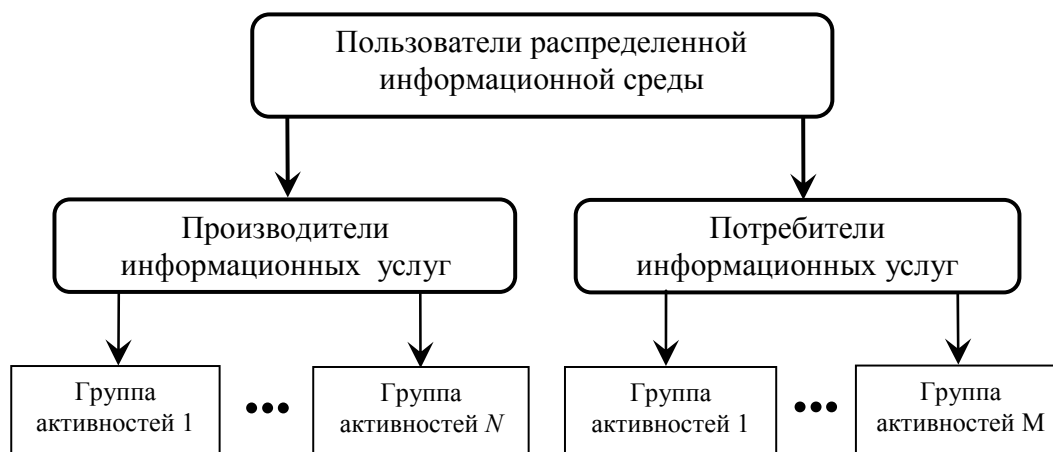


Рис. 2. Принцип классификации активностей пользователей СПД

Цели и критерии эффективности управления процессом обработки заявок с поддержкой качества обслуживания

Исходя из постановки задачи, можно сформулировать требования, предъявляемые к механизму обеспечения качества обслуживания пользователя при работе с удаленными данными:

- элементарным объектом управления модели обеспечения качеством обслуживания является заявка;
- критерием классификации заявок является смысловое значение (семантика) передаваемой информации, важность оперативности ее доставки пользователю с точки зрения принятия решения, функционирования и развития СПД;
- основной характеристикой качества обслуживания является задержка в очереди при ожидании освобождения обслуживающего устройства (под обслуживающим устройством будем понимать комплекс программно-аппаратных средств и каналов передачи данных, необходимый для обслуживания заявки).

Целью работы любого механизма, обеспечивающего качество обслуживания, является справедливое (или равномерное) распределение ресурса обслуживающего устройства. Термин «справедливое» имеет не только разговорное, но и техническое значение (существует несколько определений). Общим во всех формулировках справедливости является распределение между классами ресурса обслуживающего устройства пропорционально назначенным весовым коэффициентам, в том числе той его части, которая не используется пассивными на данный момент классами. Важным отличием этих формулировок является наличие или отсутствие штрафа за использование ресурса в предыдущие периоды, в том числе за счет пассивных классов. Большинство определений справедливого обслуживания предполагают отсутствие подобного штрафа [2–6]. Этой концепции будем придерживаться в дальнейшем.

Определим начальные условия и введем некоторые обозначения. Пусть имеется N потоков заявок на обслуживание. Каждый поток включает заявки одинакового приоритета и образует отдельный класс обслуживания. Классу обслуживания выделяется очередь, выбор заявок внутри которой осуществляется по принципу FCFS – «первым пришел – первым обслужен».

Пусть в момент времени, когда заявка поступает в систему, обслуживающее устройство занято, и ни в одной очереди нет ожидающих обслуживания заявок. Тогда заявка помещается в очередь, очередь становится единственной активной. Если до завершения обслуживания заявки, занимающей обслуживающее устройство, количество активных очередей увеличится, то возникнет конфликт классов обслуживания и должен вступить в работу механизм обеспечения качества обслуживания. Обозначим момент времени, когда количество активных очередей стало более единицы t_s . Соответственно, определим условием окончания работы механизма обеспечения качества обслуживания – уменьшение количества активных очередей до одной. Обозначим этот момент времени t_f .

Мерой качества обслуживания ранее была названа задержка заявки в очереди на обслуживание. Если говорить о равномерности обслуживания в целом, то значения времен ожидания заявок в очереди за определенный интервал времени должны быть усреднены. Выбор интервала времени, за который будет осуществляться усреднение, осуществляется в соответствии с требованием, отсутствия влияния предыдущих периодов активности класса на предоставляемое ему в дальнейшем качество обслуживания. Мерой использования ресурса обслуживающего устройства будем считать среднюю задержку в очереди за текущий период активности. Это означает, что как только из очереди извлекается последняя заявка, среднее значение ее задержки приравнивается нулю [7].

Пусть классам качества обслуживания назначены весовые коэффициенты: $0 < m_i < 1$ ($i = 1 \dots N$), причем $\sum_{i=1}^N m_i = 1$. Назначение классу i приоритета m_i означает, что его средняя задержка d_i за время с начала текущей активности должна удовлетворять равенству:

$$\frac{d_i}{d_j} = \frac{m_i}{m_j}, \quad (1)$$

где d_j , m_j – средняя задержка за период текущей активности некоторого активного в данный момент времени класса j и его весовой коэффициент соответственно.

Суммируя по всем активным классам j , получим:

$$\frac{d_i}{\sum_{j \in A} d_j} = \frac{m_i}{\sum_{j \in A} m_j},$$

где A – подмножество очередей, проявляющих активность в рассматриваемый момент времени.

Можно сформулировать условие справедливости, являющееся критерием эффективности процессов информационного обмена в условиях перегрузки.

Распределение ресурса обслуживающего устройства является справедливым, если:

– среднее за период текущей активности время ожидания заявок в очереди i -го класса обслуживания составляет

$$d_i = \frac{m_i}{\sum_{j \in A} m_j} \cdot \sum_{j \in A} d_j ;$$

– класс обслуживания не штрафуются за использование ресурса обслуживающего устройства в предшествующие периоды.

Первое условие автоматически удовлетворяет требованию распределения неиспользованной части ресурса обслуживающего устройства между активными классами в соответствии с назначенными им весовыми коэффициентами. В случае если активной является только очередь i , имеет место тривиальное равенство $d_i = d_i$, что означает полный захват ресурса обслуживающего устройства классом i . Теперь необходимо формализовать приведенные выше принципы управления на основе правил выбора следующей заявки на обслуживание.

Формальные правила управления процессом обслуживания заявок в предлагаемой модели

Перепишем выражение (1) в виде $d_i / m_i = d_j / m_j$ и рассмотрим возможные варианты взаимодействия двух потоков при функционировании СПД.

1. Все очереди пусты и поступающие заявки сразу направляются на обслуживание. Задержки различных классов равны нулю и равенство удовлетворяется.

2. Появляется одна активная очередь i , и равенство нарушается $d_i / m_i > 0$. Для восстановления равновесия необходимо выбрать заявки из очереди i . В результате мы получаем или первый случай, или третий.

3. При функционировании СПД появляются две активные очереди. Возможны ситуации выполнения равенства, и его нарушения. В первом случае выбор класса, заявка которого будет обслужена первой, может быть произвольным (например, розыгрышем случайного числа). Во втором случае для восстановления равновесия выбирается класс, чье отношение d/φ больше.

Обобщив рассмотренный случай на произвольное количество классов обслуживания в модели, получим формальные правила выбора заявок на обслуживание, обеспечивающие удовлетворение условия справедливости.

Выбор следующей заявки на обслуживание осуществляется в соответствии со следующими правилами:

– право занять обслуживающее устройство предоставляется заявке из очереди i -го класса обслуживания, где:

$$i = \max_{d_j/\varphi_j} \{j \mid j \in A\};$$

- если таких классов окажется более одного, выбор определяется случайным образом;
- порядок следования заявок внутри каждой очереди определяется принципом FCFS;

С целью формализации понятия средней задержки за текущий период активности рассмотрим его более подробно. Пусть имеется активная очередь i . Обозначим текущий момент времени t , а время начала текущего периода активности очереди t_{ia} . Число, покинувших очередь заявок за интервал времени $(t_{ia}; t]$ обозначим n_{if} . Определим их суммарную задержку:

$$\sum d_{if} = \sum_{j \in F_i} (t_{if}^j - t_{is}^j),$$

где F_i – множество заявок, покинувших очередь за время $(t_{ia}; t]$; t_{is}^j , t_{if}^j – соответственно время постановки и время изъятия из i -ой очереди заявки с номером j .

Использовать это значение, отнесенное к количеству выбранных из очереди заявок n_{if} , в качестве меры использования ресурса обслуживающего устройства d_i не представляется возможным. Такой подход делает невозможным обновление информации о состоянии

очереди без выбора заявки из нее, а такой выбор может не произойти по причине отсутствия достоверной информации о состоянии очереди.

Состояние очереди характеризуется набором заявок, находящихся в ней на момент времени t . Обозначим их количество n_{ia} и определим суммарную задержку ожидающих в очереди i заявок:

$$\sum d_{ia} = \sum_{j \in A_i} (t - t_{is}^j),$$

где A_i – множество ожидающих в очереди i заявок.

Отсюда получим среднюю задержку ожидания в очереди i за текущий период активности:

$$d_i = \frac{\sum d_{if} + \sum d_{ia}}{n_{if} + n_{ia}}. \quad (2)$$

Сформулирована идеальная модель качества обслуживания по величине задержки. Применение ее на практике обуславливается следующими ограничениями:

Во-первых, усреднение нестационарных величин (величины задержек в общем случае за период активности изменяются произвольно) не дает объективной информации о поведении системы.

Во-вторых, при длительной перегрузке (в реальных условиях применение модели эффективно именно при длительных перегрузках) слагаемые $\sum d_{ia}$ и n_{ia} числителя и знаменателя формулы (2) становятся пренебрежимо малыми по сравнению с $\sum d_{if}$ и n_{if} соответственно. То есть поведение системы определяется предысторией и практически не зависит от текущего состояния, что приводит к принятию неправильных решений при управлении.

В-третьих, одной из основополагающих концепций предложенного подхода является независимость приоритета заявки на обслуживание от трудоемкости обработки заявки, в то время как длительность ее обработки может не быть пренебрежимо малой по сравнению с временем ожидания в очереди (например, при необходимости обращения к удаленным данным). Это может привести к необоснованным задержкам высокоприоритетных заявок, пришедших во время обслуживания низкоприоритетных заявок большой сложности.

Литература

1. Бартунов О.С., Прохоров М.Е., Родичев Е.Б. Управление доступом в сложных информационных системах на основе ролевой авторизации // Единая образовательная информационная среда. 2004. № 2. С. 103–108.
2. Demers A. Keshav S. Shenker S. Analysis and simulation of a fair queueing algorithm // Internetworking: Research and Experience. 1990. V 1. P. 3–26.
3. Bennett J.C. Zhang H.R. Hierarchical packet fair queueing algorithms // In Proceedings of the ACM-SIGCOMM96. 1996. P. 143–156.
4. Floyd S. Jacobson V. Link-sharing and resource management models for packet networks // IEEE/ACM Transactions on Networking. 1995. № 4. V3. P. 365–386.
5. Stoica I. Zhang H. A Hierarchical Fair Service Curve Algorithm for Link-Sharing, Real-Time and Priority Services // In Proceedings of the ACM-SIGCOMM97. 1997. P. 249–262.
6. Zhang L. Virtual clock: A new traffic control algorithm for packet switched networks // IEEE/ACM Transactions on Computer Systems. 1990. № 2. V 9. P. 10–24.
7. Константинов И.С., Фролов А.И. Модель обеспечения качества обслуживания по величине задержки, основанная на семантических приоритетах // Вестник компьютерных и информационных технологий. 2005. № 3. С. 30–34.