

Использование распределенной иерархии кеш-прокси серверов для маршрутизации Web-трафика

А. С. Петросян

Институт проблем информатики и автоматизации НАН РА и ЕрГУ
E-mail: arthur@sci.am

В настоящее время трафик World Wide Web (или Web-трафик) составляет большую часть общего трафика передаваемого в Internet и его процент постоянно растет. Под Web-трафиком понимается информация, передаваемая от различных серверов Internet Web-браузеру по протоколам, поддерживаемым данным Web-браузером (http, ftp, gopher и т.д.). По оценкам, иногда Web-трафик достигает 80% от общего трафика в магистральных каналах некоторых глобальных сетей США.

Поэтому то, как будет направляться этот трафик в сетях (т.е. его маршрутизация) играет существенную роль. Доступ к системе World Wide Web зачастую организуется не прямым доступом Web-браузера к Web-серверу (или другому серверу Internet), а через кеш-прокси сервер, который выполняет роль посредника и располагается вблизи Web-браузера (часто в локальной или региональной сети). Этот промежуточный сервер предоставляет Web-браузерам доступ в Internet и одновременно осуществляет кеширование информации которая передается через него (так называемое "Web-кеширование").

Термин "Web-кеширование" означает промежуточное хранение Web-трафика, которое имеет целью удовлетворить последующие запросы на те же документы предоставлением уже сохраненной (кешированной) информации, вместо того, чтобы вновь запрашивать эти документы из Internet.

Хотя, как уже было указано, Web-кеширование можно эффективно реализовать на одном кеш-прокси сервере, однако более эффективно использование распределенной иерархии нескольких кеш-прокси серверов. Данная статья обсуждает использование подобной распределенной иерархической системы кеширования для маршрутизации Web-трафика. Для начала необходимо вкратце вспомнить как осуществляется маршрутизация в Internet.

TCP/IP маршрутизация

Работа в сети Internet основана на сетевых протоколах семейства TCP/IP. Для связи между компьютерами каждому из них на сетевом уровне присваивается уникальный адрес из четырех байтов - IP-адрес. Этот IP-адрес состоит из 2-х частей: сетевой и машинной. Сетевая часть обозначает логическую подсеть, к которой относится данный адрес (на основании этой принадлежности и принимаются решения о маршрутизации на сетевом уровне). Машинная часть определяет номер самого компьютера в данной подсети. IP-адреса обычно записываются как четыре десятичных числа разделяемых точками (например 193.232.229.65). С помощью дополнительного аналогичного набора из четырех байтов, называемого "маской подсети", логическим сложением (AND) его с самим IP-адресом определяется, какое количество IP-адресов входит в данную подсеть (т.е. какая часть IP-адреса относится к сетевой, а какая к машинной части).

В некоторой локальной сети компьютерам могут быть присвоены определенные IP-адреса и обеспечена работа между ними по протоколам TCP/IP. А

для связи их с другими компьютерами в других сетях необходимо наличие связующего звена (называемого *маршрутизатором*), которое бы осуществляло пересылку данных между сетями, т.е. маршрутизацию. Для выполнения этой работы *маршрутизатор* должен знать, где находятся какие сети (т.е. иметь таблицу маршрутов к различным IP-адресам). Поскольку неэффективно и невозможно держать таблицу о маршрутах ко всем сетям, обычно определяются маршруты к ближайшим сетям и один дополнительный маршрут "по умолчанию" (*default route*), который определяет направление пересылки всех пакетов, для которых маршруты не определены в таблице маршрутизации.

Любые услуги Internet (в том числе система WWW) работают на основе этих сетевых протоколов TCP/IP. Поэтому когда Web-браузер на каком-либо компьютере должен послать запрос серверу, этот запрос формируется на прикладном уровне с указанием IP-адреса сервера, а затем на сетевом уровне определяется маршрут (с помощью протоколов TCP/IP) к серверу для установки соединения с ним.

Web маршрутизация

Как уже было отмечено, современные Web-браузеры имеют возможность работы через кеш-прокси сервер, что существенно изменяет способ соединения Web-браузера с серверами. Web-браузер, настроенный на работу через кеш-прокси сервер, направляет все запросы не соответствующим серверам, а кеш-прокси серверу и от него же получает ответы. Таким образом при работе с кеш-прокси сервером в сети осуществляется некая дополнительная маршрутизация Web трафика на прикладном уровне (называемая *Web-маршрутизацией*), так как весь трафик направлен через кеш-прокси сервер и затем распределяется по Web-браузерам, расположенным в локальной или региональной сети (рис.1). Можно сказать, что кеш-прокси сервер играет как-бы роль маршрута "по умолчанию" (*default route*) (в примерах в качестве Web-браузера принят Netscape Navigator for Windows95 версии 3.x или выше, а в качестве кеш-прокси сервера - Squid Internet Object Cache for UNIX, рис.1).

Web-маршрутизация представляющая собой перенаправление потоков данных позволяет создавать "реальные" маршруты для Web-трафика, отличные от TCP/IP маршрутов, действующих в данной сети. Например, если компьютер на котором работает Web-браузер согласно действующей сетевой TCP/IP маршрутизации не имеет прямого доступа к Internet, то он может получить его благодаря перенаправлению Web-трафика через кеш-прокси сервер.



(рис.1)

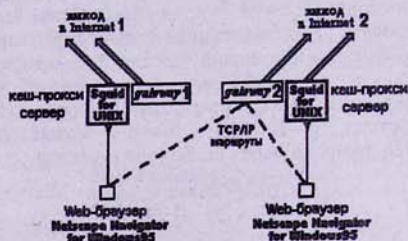
Web-браузер, пользуясь *своей* таблицей TCP/IP маршрутов, достигает кеш-прокси сервера, расположенного в той же сети и передает ему свой запрос. В свою

очередь кеш-прокси сервер (имеющий прямой доступ к Internet) используя TCP/IP маршруты, уже установленные в его маршрутной таблице, осуществляет доступ к Internet и доставляет Web-браузеру запрошенную им информацию. Фактически компьютер на котором работает Web-браузер как бы использует TCP/IP маршруты кеш-прокси сервера, хотя на нем самом могут быть установлены совершенно иные TCP/IP маршруты, которые будут неизменно действовать для других TCP/IP приложений запускаемых с данного компьютера (кроме Web-браузера, настроенного на работу через кеш-прокси сервер).

При использовании одного кеш-прокси сервера основные преимущества состоят в том, что Web-браузеры не имеющие прямого доступа к Internet, могут его получить. На кеш-прокси сервере также организуется Web-кеширование, которое ускоряет ответы на повторные запросы документов, и способствует более эффективному использованию канала выхода в Internet.

Еще больше возможностей Web-маршрутизации открывается в распределенной иерархии кеш-прокси серверов и при наличии более одного выхода в Internet. В этом случае кеш-прокси серверы можно установить в местах наиболее близких к точкам выхода в Internet, и определить TCP/IP маршруты на них таким образом, что каждый кеш-прокси сервер мог бы осуществлять независимый прямой доступ к Internet по своему каналу выхода в Internet (рис.2). В то же время TCP/IP маршруты локальной или региональной сети должны быть установлены таким образом, чтобы все кеш-прокси серверы в распределенной иерархии были достижимы любым Web-браузером в этих сетях.

После подобной, установленной на сетевом уровне TCP/IP маршрутизации, на прикладном уровне каждому Web-браузеру задаются маршруты связи с кеш-прокси серверами. Существует два способа настройки Web-браузера на работу через кеш-прокси сервер. Первый — ручной, когда в конфигурации Web-браузера непосредственно прописывается адрес и порт одного конкретного кеш-прокси сервера для работы через него.



(рис.2)

Основной недостаток подобного способа состоит в том, что возможно указание лишь одного кеш-прокси сервера, что каждый Web-браузер настраивается отдельно и в случае изменений в конфигурации (например изменения имени кеш-прокси сервера) необходимо будет вручную перенастраивать все Web-браузеры. Поэтому все больше используется второй способ настройки Web-браузера на работу через кеш-прокси сервер — с помощью файла автоматической конфигурации прокси (Automatic Proxy Configuration). Из наиболее распространенных в настоящее время Web-браузеров возможность автоматической конфигурации прокси имеют: Netscape Communicator, Netscape Navigator v.2.02 и выше, Microsoft Internet Explorer v.3.02 и выше. При таком способе настройки Web-браузера в его конфигурации указывается лишь URL (Uniform Resource Locator) к файлу автоматической конфигурации прокси. Этот файл представляет из себя программу на языке JavaScript и располагается где-либо на

местном Web-сервере. При каждом запуске Web-браузера он пытается обновить этот файл с указанного Web-сервера. Это позволяет централизованно вводить изменения в настройки изменяя этот файл и не беспокоясь о Web-браузерах, так как каждый из них при следующем своем запуске обновит у себя настройки на работу через кеш-прокси сервер, получив новый файл автоматической конфигурации прокси.

Помимо этого, в таком файле предполагается различных конфигураций. Например, можно указать имена доменов, для которых не следует использовать кеш-прокси сервер. Таким образом для этих доменов определяются Web-маршруты, отличные от маршрута "по умолчанию" (*default route*), через кеш-прокси сервер и доступ к серверам в этих доменах будет определяться непосредственно на сетевом уровне на основе TCP/IP маршрутов, установленных на том компьютере, где запущен Web-браузер.

Возможно также указание использовать различные кеш-прокси серверы для различных доменов (например для доступа к сетям ".de" направлять запросы через *proxy.sci.am*, а для доступа к сетям ".com" - через *proxy2.sci.am*). И даже возможно указание не одного кеш-прокси сервера, а последовательности (списка) кеш-прокси серверов, которые будут использоваться по очереди. В случае выхода из строя одного кеш-прокси сервера он на время удаляется из списка, и осуществляется переключение на следующий. Через определенный интервал осуществляется очередная попытка восстановить работу через кеш-прокси сервер ранее удаленный из списка.

Дополнительным преимуществом автоматической конфигурации прокси является то, что решения о Web-маршрутизации принимает не пользователь (вручную настраивая свой Web-браузер), а квалифицированный администратор, который может эффективно определить Web-маршруты определив этим направления потоков данных в сети. Очевидно, что все это открывает большие возможности гибкой конфигурации для эффективного направления трафика в сетях.

Все вышеописанные возможности реализованы в Армянской академической компьютерной сети и позволяют эффективно распределять загрузку каналов выхода в Internet, осуществлением Web-маршрутизации на прикладном уровне.

Литература

1. Ingrid Melve, Why Internet Service Providers should integrate web caches into their networks
<http://w3cache.icm.edu.pl/workshop/talk2/>
2. A. Daviel, Proxy Cache – Van-Pool for the Web
<http://vancouver-webpages.com/proxy>
3. Oskar Pearson, Squid Users Guide <http://cache.is.co.za/squid>
4. Desire Project, Survey of caching requirements and specifications for prototype
<http://www.cc.ruu.nl/~henny/desire/survey.html>
5. Duane Wessels, Configuring Hierarchical Squid Caches
<http://squid.nlanr.net/Squid/Hierarchy-Tutorial>
6. Netscape Corp., Navigator Proxy Auto-Config File Format
<http://home.netscape.com/eng/mozilla/2.0/relnotes/demo/proxy-live.html>
7. Nlanr, Squid Internet Object Cache v1.1 Release Notes
<http://squid.nlanr.net/Squid/1.1/Release-Notes-1.1.txt>
8. A. J. Flavell, WWW Cache Briefing <http://ppewww.ph.gla.ac.uk/~flavell/cache.html>
9. DFN-Cache Service Project <http://www-cache.dfn.de>
10. Artem Belevich, Примерный java-скрипт для Automatic proxy configuration
<http://kulichki.rambler.ru/moshkow/WEBMASTER/proxyauto.txt>