

О времени обработки запросов в распределенных информационных системах на основе протокола Z39.50

© Жижимов О.Л., Скибин С.В.

Объединенный институт геологии, геофизики и минералогии им. А.А. Трофимука СО РАН
zhizhim@uiggm.nsc.ru, skibin@uiggm.nsc.ru

Аннотация

В докладе рассматриваются вопросы, связанные с временем выполнения многосерверных поисковых запросов в распределенных информационных системах на основе протокола Z39.50, объединяющих информацию, которая по тем или иным причинам не может существовать как единое хранилище данных. Для тестовой распределенной информационной системы на основе шести серверов проведены измерения времени выполнения запросов с многократной переадресацией. Показано, что основные потери времени происходят в момент открытия вторичных сеансов связи между серверами. В докладе обсуждаются способы минимизации временных потерь в подобных распределенных системах, приводятся рекомендации для администраторов систем.

Как правило, в гетерогенных информационных системах на основе протокола Z39.50 существует одна или несколько точек доступа, реализованных в виде шлюзов Z39.50-http или в виде серверов Z39.50 с функциями перенаправления запросов [3]. Также предполагается, что в информационной системе все серверы доступны по протоколу Z39.50 и обрабатывают стандартные запросы RPN. В качестве базовых серверов РИС рассматриваются серверы ZooPARK [4], широко распространенные в России, т.к. другие доступные серверы Z39.50 не обладают возможностями переадресации поисковых запросов и поддержкой широкого спектра СУБД и схем данных. Также существует возможность прозрачной для пользователя обработки запросов одновременно в нескольких базах данных, в том числе находящихся на других серверах.

При многобазовом поиске информации в распределенных информационных системах происходит обращение к множеству баз данных, в том числе расположенных на различных серверах. При этом

каждый сервер, входящий в информационную систему, может управлять одной или несколькими базами данных, хранящимися на различных физических носителях информации. Если учесть, что все серверы информационной системы имеют различное быстродействие и реальную загрузку, что они соединены каналами связи различной пропускной способностью, то возникает вопрос о наиболее оптимальной конфигурации распределенной информационной системы с точки зрения скорости обработки сложного запроса, т.е. времени, прошедшего с момента прихода запроса в точку входа в РИС до отправки клиенту ответа.

Для проведения измерений времени обработки сложных многобазовых и многосерверных запросов была задействована РИС, состоящая из шести серверов ZooPARK на различных аппаратных и программных платформах, но с примерно одинаковыми дисковыми подсистемами:

	Процессор	RAM, Mb	Операционная система	ДБ
S1	P4-1500	512	Win2000 Srv	
S2	P-200MMX	128	Linux RH6.2	
S3	P3-1000	256	Solaris 8 x86	GG
S4	P4-1700	256	Linux RH8	
S5	2 x P3-1000	1024	Win2000 Srv	GG
S6	2 x P4-2000	2048	Win2000 Srv	GG

Таблица 1. Состав тестируемой РИС

Серверы S3, S5 и S6 управляли одинаковыми базами данных – GG. Эта база данных является групповой и состоит из пяти разрозненных, но однотипных, баз данных под управлением CDS/ISIS с суммарным количеством записей около 800 тыс. Доступ к базам данных осуществлялся или напрямую или через точки доступа S1, S2, S4. Ниже для определенности будет указано имя базы данных, включающее путь доступа. Так, имя базы данных GG_1_3 будет означать доступ к базе данных GG на сервере S3 через сервер S1, а имя GG_1_4_3 – доступ к БД GG на сервере S3 последовательно через серверы S1 и S4.

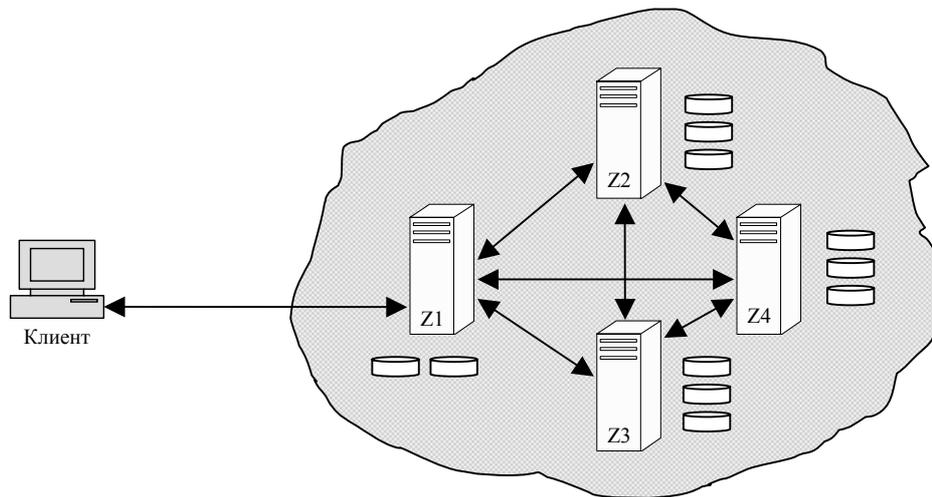


Рис. 1. Модель распределенной информационной системы

Уровень	Имя БД	Время 1, мс	Время 2, мс
0	GG_5	59 ± 9	
0	GG_3	60 ± 50	
0	GG_6	19 ± 7	
1	GG_1_3	941 ± 57	45 ± 13
1	GG_1_5	151 ± 9	68 ± 8
1	GG_1_6	4545 ± 8	23 ± 9
1	GG_2_3	919 ± 12	
1	GG_4_3	912 ± 10	

Таблица 2. Среднее время выполнения запросов

Для измерения времени выполнения запросов использовалась серия из 100 запросов, половина из которых были составными. Усредненные значения времени в миллисекундах приведены в таблице 2, при этом в колонке «Уровень» проставлено число переадресаций запросов.

Здесь следует сделать отступление и пояснить значения «Время 1» и «Время 2» в приведенной

таблице. Работа сервера ZooPARK в составе распределенной информационной системы возможна в двух режимах функционирования провайдера данных удаленного сервера (Z-REMOTE). Эти два режима проиллюстрированы на рис. 2 и рис. 3 в виде временных диаграмм взаимодействия клиента и серверов в распределенной системе.

Режим 1 соответствует диаграмме (см. рис. 2) сеанса связи клиента и сервера 1, перенаправляющего запросы серверу 2, который взаимодействует с СУБД. Сеанс состоит из процедуры открытия сеанса (1), двух запросов на поиск (2, 7) и процедуры закрытия сеанса. При этом обработка запросов на поиск сервером 1 сводится к открытию им сеанса связи с сервером 2 (3, 8), послышки ему запроса на поиск (4, 9) и закрытия сеанса (8, 11). Доступ к СУБД имеет сервер 2 (5, 10) при обработке им запроса на поиск от сервера 1. Заметим, что в этом режиме любое обращение сервера 1 к серверу 2 сопровождается установлением сеанса связи, выполнением операции и закрытием сеанса. С одной стороны, это позволяет не заботиться о сохранении контекста вторичных сеансов, что экономит сеансовую память сервера 1, но, с другой стороны, приводит к увеличению

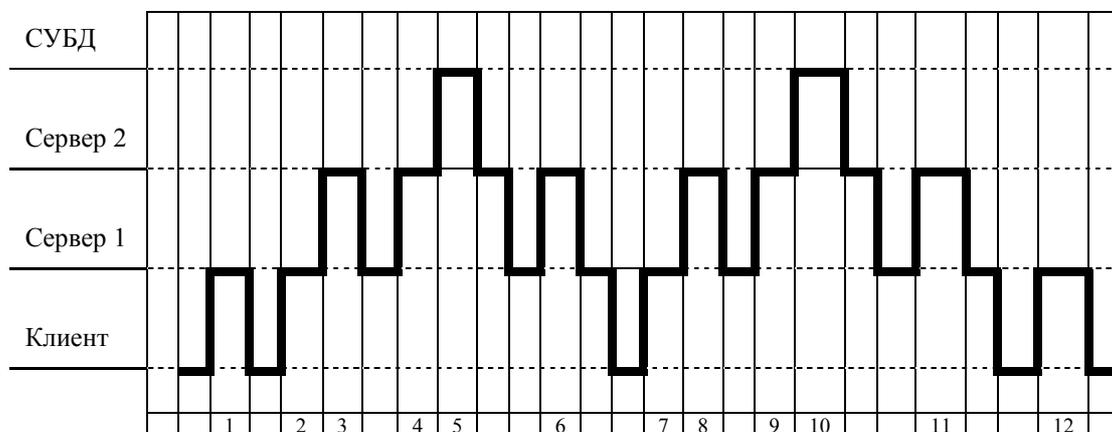


Рис. 2. Временная диаграмма режима 1 взаимодействия клиента и серверов в РИС

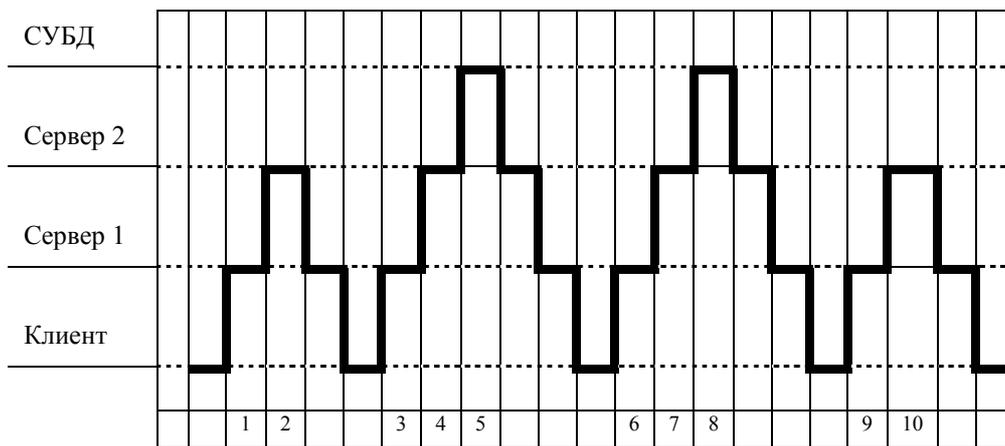


Рис. 3. Временная диаграмма режима 3 взаимодействия клиента и серверов в РИС

сетевого трафика и времени ожидания клиентом окончания операций. Дело в том, что время, необходимое на открытие сеанса, может быть намного больше времени выполнения любой операции, например, поиска в базе данных. При этом непроизводительные потери клиентского времени растут пропорционально количеству операций, выполняемых клиентом в течение одного сеанса. Тем не менее, именно этот режим работы серверов ZooPARK и провайдеров Z-REMOTE свободен от ловушек, о которых речь пойдет ниже, и прост для администрирования распределенной системы в целом. Также следует заметить, что именно режим 1 принят основным в распределенных системах CIP (CIP - Catalogue Interoperability Protocol – прикладной профиль Z39.50, созданный для космического мониторинга земной поверхности) [5].

Режим 2 соответствует диаграмме, изображенной на рис. 3. В этом режиме при открытии клиентом сеанса связи с сервером 1 (1) последний сразу открывает сеанс связи с сервером 2 (2) и сохраняет контекст этого сеанса в своих сеансовых переменных. Поэтому последующие операции клиента, например, операции поиска (3, 6), для сервера 1 сводятся лишь к передаче соответствующих запросов серверу 2 (4, 7) в контексте открытого ранее сеанса. При закрытии сеанса связи клиентом (9) закрывается сеанс связи сервера 1 с сервером 2 (10).

Как видно из логики работы сервера, режим 2 свободен от непроизводительных потерь клиентского времени в режиме 1. Однако следует иметь в виду, что в распределенной системе сервер 1 может общаться не только с сервером 2, но и с другими серверами, например, серверами 2, 3, 4. При этом для каждого из них в контексте сеанса «клиент-сервер1» должен быть открыт свой сеанс связи, например, «сервер1-сервер2», «сервер1-сервер3», «сервер1-сервер4» и т.д. Процесс открытия этих вторичных сеансов требует времени, а сохранение их контекста – ресурсов сервера 1. И если последнее не очень существенно при достаточной оперативной памяти, то время ожидания клиентом открытия сеанса связи с сервером 1 в режиме 2 будет пропор-

ционально количеству серверов в распределенной системе и может оказаться очень большим.

Впрочем, проблема большого времени ожидания открытия сеанса решается на уровне архитектуры провайдера данных Z-REMOTE. Дело в том, что поскольку операции открытия сеансов связи с разными серверами независимы, они могут происходить в отдельных потоках и тем самым выполняться практически одновременно.

Однако вернемся к таблице 2. Для экономии места в ней не приведены значения времени выполнения запросов с числом переадресаций, большим 1. Ясно, что если в распределенной системе выполняются запросы с числом переадресаций большим 1, то такая система просто не оптимизирована и требуется вмешательство администратора. Тем не менее, мы приведем усредненную диаграмму зависимости времени выполнения запросов от числа переадресаций (рис. 4), отбросив крайние значения. Как и следовало ожидать, эта зависимость с учетом отклонений практически линейная. Характерно, что угол наклона этой аппроксимирующей прямой может характеризовать степень эффективности распределенной информационной системы в части обработки многосерверных запросов – чем меньше угол, тем меньше времени тратится на переадресацию запросов.

Что же касается запросов с однократной переадресацией, то из таблицы 2 видно, что в режиме 2, обсуждавшемся выше, работа распределенной информационной системы намного эффективней. Время выполнения запросов в режиме 1 может в десятки и сотни раз превышать время выполнения в режиме 2. Последнее свидетельствует о том, что в тестируемой распределенной системе в режиме 1 основное время тратится не на собственно обработку поисковых запросов к базам данных, а на процедуры установления сеансов связи между серверами. Поскольку в режиме 2 все внутренние сеансы открываются в момент открытия первичного сеанса, суммарное время обработки запроса резко уменьшается и содержит в основном время обращения к базам данных.

Таким образом, эффективность распределенной информационной системы зависит от времени открытия сеанса Z39.50 между серверами системы. Можно утверждать, что основная часть этого времени уходит на процедуры взаимной аутентификации серверов. Сюда следует отнести вторичные запросы к серверам имен (DNS, WINS) и, при наличии централизованных служб аутентификации, к серверам аутентификации, например, контроллерам доменов Windows. Кстати, именно последняя процедура сильно увеличивает время ответа для БД GG_1_6 (см. табл. 2), т.к. в этом случае два сервера Windows относились к различным доменам аутентификации.

Анализируя время выполнения многосерверных запросов с многократной переадресацией, можно построить очень эффективную распределенную информационную систему на основе нескольких серверов ZooPARK. При продуманном распределении баз данных по разным серверам можно получить высокую степень распараллеливания поисковых запросов и тем самым существенно уменьшить время их выполнения. Однако существует несколько моментов, на которые следует обратить внимание.

При работе нескольких серверов в распределенной системе в режиме 2 возможна ситуация возникновения инициализационных петель. Такая петля появляется уже в системе из двух серверов, работающих в режиме 2. Суть ее состоит в том, что если сервер 1 настроен на открытие вторичного сеанса с сервером 2, а сервер 2 – на открытие вторичного сеанса с сервером 1, т.е. создана симметричная конфигурация серверов, то клиент *никогда* не сможет открыть сеанс связи ни с сервером 1, ни с сервером 2. Действительно, в этом случае событийная цепочка будет выглядеть следующим образом:

- клиент запрашивает сервер 1 для открытия сеанса «клиент-сервер1»
- открывая этот, сеанс сервер 1 запрашивает сервер 2 для открытия вторичного сеанса «сервер1-сервер2»
- открывая этот, сеанс сервер 2 запрашивает сервер 1 для открытия вторичного сеанса

«сервер2-сервер1»

- открывая этот, сеанс сервер 1 запрашивает сервер 2 для открытия вторичного сеанса «сервер1-сервер2»
- и т.д. до бесконечности

Нетрудно догадаться, что в системе из большого количества серверов ситуация может принять еще более тяжелый характер, оставаясь по своей сути той же самой. Поэтому для избежания инициализационных петель не следует создавать симметричные конфигурации серверов в распределенных системах. Идеальный случай сводится к разделению функций серверов на серверы, перенаправляющие запросы (точки входа в распределенную систему), и серверы, которые взаимодействуют с СУБД.

Другой момент, который ухудшает временные характеристики распределенной системы и присущ как режиму 2, так и режиму 1, состоит в возможности появления вторичной переадресации. Эту ситуацию можно проиллюстрировать на примере. Пусть сервер 1 переадресует запрос серверу 2, а сервер 2 его не обрабатывает, а переадресует запрос серверу 3. В этом случае конфигурация, когда сервер 1 сразу бы обращался к серверу 3, была бы более экономичной. Переадресации уровня выше первого в распределенной системе не должно быть вообще, т.к. ее наличие ничего кроме проблем не дает. Более того, вторичная переадресация может привести к появлению петли переадресации, в том числе и в режиме 1.

Для защиты от избыточной переадресации, инициализационных петель и петель переадресации в сервере ZooPARK предусмотрены ограничения, задаваемые в конфигурационном файле провайдера данных Z-REMOTE как на максимально допустимое количество переадресаций запроса, так и на количество петель.

В заключение ниже приведен список рекомендаций, которым следует руководствоваться при построении распределенных информационных систем на основе серверов ZooPARK:

- Распределенная система должна содержать несколько серверов ZooPARK.

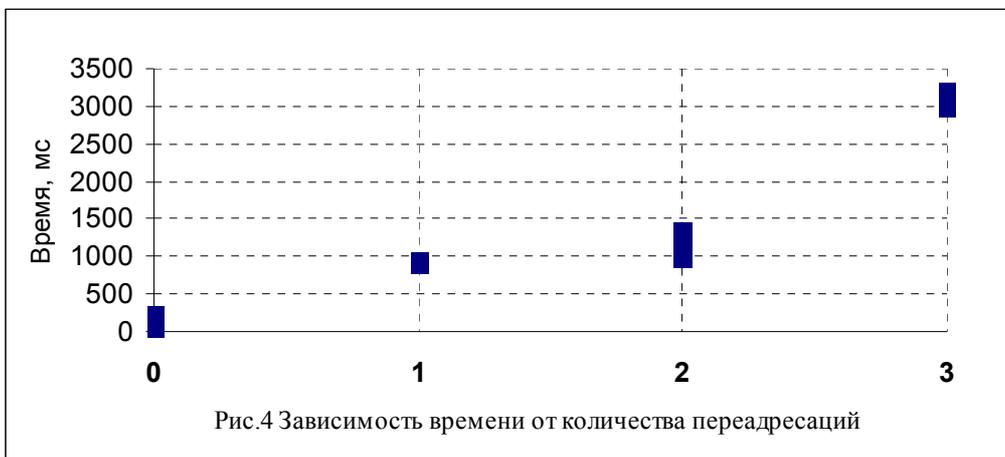


Рис.4 Зависимость времени от количества переадресаций

- Все серверы должны быть объединены надежными линиями связи. Требуемая ширина полосы пропускания этих линий зависит от величины сторонней загрузки, но сама по себе слабо сказывается на эффективности информационной системы.
- Выбор аппаратной платформы для функционирования серверов ZooPARK не принципиален, но предпочтение следует отдавать многопроцессорным системам с быстрыми дисковыми устройствами.
- В качестве программной платформы можно использовать Windows NT/2000 или UNIX (Linux, Solaris, FreeBSD). Временные характеристики систем на Windows обычно лучше, но надежность UNIX выше. Из всех перечисленных UNIX наиболее эффективной является Solaris, наименее – Linux.
- Все серверы РИС должны или относиться к одному домену аутентификации или не использовать централизованных служб.
- Базы данных в распределенной системе следует группировать по серверам в соответствии с их аппаратной конфигурацией – на более мощном сервере должно находиться большее количество баз данных.
- В распределенной системе должны существовать несколько серверов, способных переадресовывать запросы. На остальных серверах провайдер данных Z-REMOTE лучше всего заблокировать.
- Выделенные для переадресации серверы должны функционировать в режиме 2 (см. выше). Список удаленных серверов для этого режима должен включать все серверы распределенной системы, кроме тех, которые работают в режиме 2 (точки входа).
- Конфигурация серверов, которые являются точками входа в систему, должна содержать в качестве имен баз данных групповые имена баз данных. Группировать базы данных следует так, чтобы в одну группу попадали базы данных с разных серверов.
- Для пользователей следует открывать доступ только к групповым именам баз данных.
- На каждом сервере должна существовать база данных IR-Explain-1, в которой должны быть описаны все базы данных этого сервера, в том числе и фиктивные групповые.

Отделения РАН на базе протокола Z39.50. Электронные библиотеки, 1999, т. 2, вып. 2, ISSN 1234-5678.

- [4] ZooPARK. Модульный сервер Z39.50. Версия 4.01.
(<http://z3950.uiggm.nsc.ru:210/doc/index.htm>)
- [5] Catalogue Interoperability Protocol (CIP) Specification - Release B -2.4 - CEOS/WGISS/PTT
http://www.dfd.dlr.de/ftp/pub/CIP_documents/cip2.4/S_cover.pdf

RCDL 2003 Submission Style Guide (MS Word document)

© Oleg Zhizhimov, Sergey Skibin

Trofimuk United Institute of Geology, Geophysics and
Mineralogy,

Siberian Branch of the Russian Academy of Sciences,
Novosibirsk, Russia

zhizhim@uiggm.nsc.ru, skibin@uiggm.nsc.ru

The questions concerned with multiserver retrieval requests run-time in distributed informational systems based on Z39.50 protocol joining information that cannot exist as a single data warehouse because of some reasons, are examined in the paper. Measurements of run-time for requests with multiple redirections were carried out for test distributed informational system based on six servers. The most of time leakage is shown to occur when opening repeated communication sessions between servers. Means of time leakage reduction in similar distributed informational systems are discussed in the paper; recommendations for system administrators are given.

Литература

- [1] ANSI/NISO Z39.50-1995. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. Z39.50 Maintenance Agency Official Text for Z39.50-1995, July 1995.
- [2] *Жижимов О.Л.* Введение в Z39.50. Издание 4-е. Новосибирск, Изд-во НГОНБ, 2003.
- [3] *Жижимов О.Л., Мазов Н.А.* Модель распределенной информационной системы Сибирского