

**Репликация данных на основе технологии peer-to-peer  
в информационных системах библиотек**

**Peer-to-Peer Data Replication in Libraries'  
Information Systems**

**Реплікація даних на основі технології Peer-to-Peer  
в інформаційних системах бібліотек**

*Р. Ф. Халабия, А. В. Романов, А. В. Рыбаков*

*Московский государственный университет приборостроения и информатики  
(МГУПИ), Москва, Россия*

*R. F. Khalabiya, A. V. Romanov, and A. V. Rybakov*

*Moscow State University of Instrumentation Technology and Informatics, Moscow, Russia*

*Р. Ф. Халабія, А. В. Романов, А. В. Рыбаков*

*Московський державний університет приладобудування та інформатики  
(МДУПІ), Москва, Росія*

Одной из задач автоматизации крупных библиотек, имеющих территориально-распределенную структуру, является оперативный обмен данными между библиотеками-филиалами. Обеспечить такой обмен, призвана система репликации, которая объединяет разрозненные данные филиалов-библиотек в единую репликационную сеть. В докладе изложены основные проблемы единой репликационной сети в информационных системах библиотеки и предложено решение их с помощью технологии peer-to-peer. В статье предложен инновационный подход репликации данных на основе данной технологии, а также приведены примеры объединения библиотек, различных филиалов и их пользователей в единую сеть.

One of the automation tasks in large libraries with a distributed territorial structure is on-line data exchange between their affiliate. Replication system are designed to ensure such exchange: They unite dispersed data of affiliate libraries into an integrated replication network. The main problems of the integrated replication system in libraries' information systems are discussed in the paper. The authors offer an innovative approach towards data replication based on the discussed technology, and provide the examples of integrating libraries, affiliates and their users into a single network.

Одним із завдань автоматизації великих бібліотек, які мають територіально розподілену систему, є оперативний обмін даними між бібліотеками-філіалами. Забезпечити такий обмін покликана система реплікації, яка об'єднує розрізнені дані філіалів бібліотек в єдину реплікаційну мережу. В доповіді викладені основні проблеми єдиної реплікаційної мережі в інформаційних системах бібліотеки і запропоновано їх вирішення за допомогою технології peer-to-peer. В статті запропоновано інноваційний підхід реплікації даних на основі даної технології, а також наведені приклади об'єднання бібліотек, різних філіалів та їх користувачів в єдину мережу.

В любой человеческой деятельности невозможно обойтись без информационных технологий, не являются исключением и современные библиотеки. Из одиноких зданий, библиотеки давно превратились в мощные библиотечные комплексы, зачастую имеющие сложную территориально-распределенную структуру и высокую степень информатизации. Быстрыми темпами развивается инфраструктура и коммуникационная связь как между филиалами одной библиотеки, так и между соседними комплексами. Растет поток передаваемой, обрабатываемой и хранимой электронной информации. Многие книги переводятся в электронный формат, заводятся электронные базы учета книг, создаются локальные сети и идет необратимая автоматизация рабочего процесса.

Одной из немаловажных задач автоматизации библиотек является оперативный обмен данными между ее филиалами и другими библиотеками. В роли данных могут выступать как различные справочники и каталоги, используемые в автоматизированных системах (например, кодификаторы, шифраторы), так и данные, являющиеся непосредственным продуктом деятельности библиотек (базы данных, электронные каталоги, статистические данные, электронные издания и аудиовизуальные материалы). Обеспечить такой обмен, призвана система репликации данных, которая является необходимым инструментом в

информационной системе библиотек при передаче, хранении и обмене данными. Под репликацией понимают процесс приведения данных электронных таблиц двух баз данных (БД) в идентичное состояние [1]. Данные обычно реплицируются для повышения надежности и увеличения производительности.

Одной из разработок является концепция создания на базе библиотечного комплекса единого репликационного центра, описанная в статье «Проблемы репликации данных в информационных системах библиотек [2]. Такой подход к репликации неоднозначен и содержит ряд отрицательных моментов. Предложенное решение задачи, требует дополнительных расходов на приобретение оборудования для единого репликационного центра. Кроме того, значительно увеличивается объем сетевого трафика и, как следствие, общая нагрузка на всю сеть в целом.

Для решения перечисленных проблем, разрабатывается концепция репликационной сети, построенной на основе сетевой технологии peer-to-peer (P2P). Подобная технология позволяет решить проблему объединения разрозненных данных филиалов, и других библиотек в единую репликационную сеть, не прибегая к использованию дополнительных сетевых ресурсов.

В P2P, так же как и в классических сетях, имеющих топологию клиент-сервер, применяется такая же по своей сути распределенная модель, но имеющая очень важную отличительную черту. В архитектуре P2P не существует понятия клиента или сервера. Каждый объект в сети, назовем его реер (от англ. равный по положению), имеет одинаковый статус, это означает, что этот объект может выполнять как функции клиента (отсылать запросы) так и сервера (получать ответы).

В отличие от архитектуры клиент-сервер, такая организация позволяет сохранять работоспособность сети при любом количестве и сочетании доступных узлов. Применяя P2P технологию, можно решить множество вопросов и проблем, охватив широкий спектр услуг:

- телевидение, телеконференции;
- распределение ресурсов для разгрузки серверов;
- поддержка системы доменных имен;
- создание распределительных хранилищ данных и систем устойчивых к атакам типа «Отказ в обслуживании».

Все это является наиболее важным для нас, так как позволяет решить проблему репликации данных.

Единую репликационную сеть можно представить в виде модели классической P2P сети, которая, к сожалению, имеет ряд досадных недостатков: проблема безопасности и жесткая зависимость от каких-либо внешних факторов. Важным моментом безопасности является доступность постороннего вмешательства в P2P сеть. Например, ресурсы сети могут быть доступны хакерам по ряду причин: открытость протокола доступа, его использование, подмена и представление ложной информации, открытость передачи данных, что и позволяет осуществить перехват данных. Внешними факторами могут быть умышленными и случайными действиями, которые не создают последствий в работе сети. Возможны проблемы отключения провайдером сети или каких-либо несанкционированных действий с коммуникационными каналами. Не исключена вероятность отключения электричества в узлах связи.

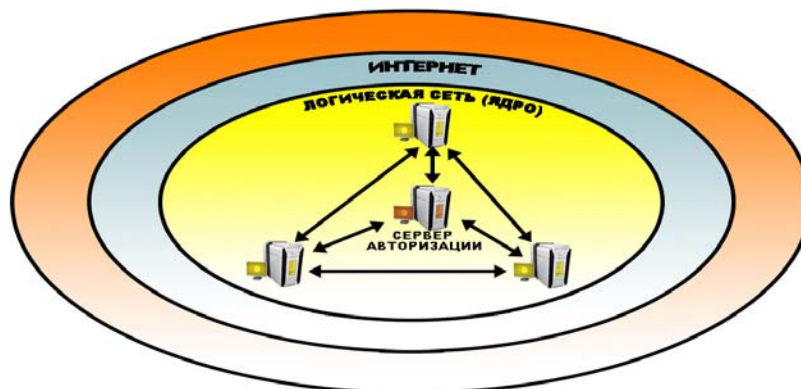


Рис. 1 Схема гибридной реер-to-реер сети

Эти проблемы решаются с помощью закрытия протокола, шифрования данных, создания жесткой системы авторизации и аутентификации, посредством введения в сеть особых серверов. Однако с вводом в P2P сеть сервера авторизации, она перестает быть одноранговой в ее привычном понимании. Поэтому вводится понятие гибридная peer-to-peer сеть (см. рис. 1).

Гибридная сеть – это такая сеть, в которой все множества клиентов подключены к друг другу, но в ней присутствует сервер, выполняющий четко определенные функции. Другими словами, в одноранговую сеть добавляется сервер, через который проходят операции авторизации, поиска, контроля версий. При недоступности узла, либо части сети, либо целого сегмента сети происходит ее самоорганизация. Под самоорганизацией сети понимают *синтез* модели, т. е. изменение количества узлов и структуры связей между ними.

Данные в библиотечных сетях обычно хранятся на серверах, и репликация данных так же происходит на сервера. При применении гибридной P2P сети данные могут располагаться непосредственно на компьютерах сотрудников библиотеки. А сервер сам проведет индексацию данных и составит таблицы метаданных, призванных облегчить дальнейший поиск требуемой информации.

Сотрудники библиотеки могут авторизоваться на сервере и войти в единую сеть библиотек и филиалов. После этого, возможно будет произвести запрос информации на сервер, и если данные, удовлетворяющие условиям запроса, были найдены, то клиент будет перенаправлен непосредственно на тот компьютер, на котором эти данные хранятся. Таким образом, данные реплицируются только по запросу, и реплицируются не целиком, а частично.

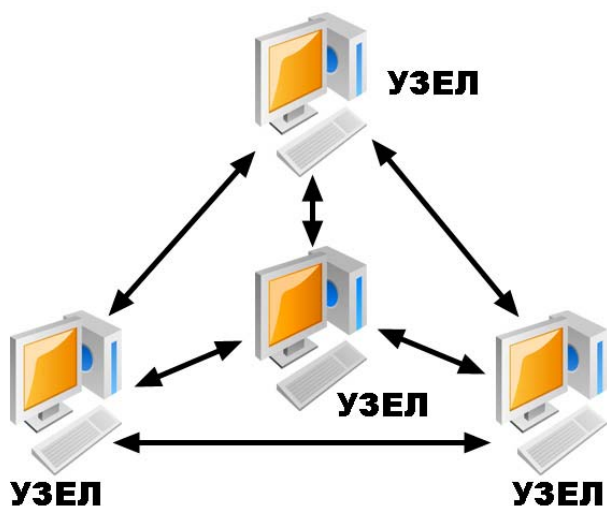


Рис. 2 Схема обмена данными в P2P сети

Еще одно преимущество такой сети при репликации данных, это удобство для пользователей библиотеки. Пользователь, войдя в сеть библиотеки, становится равноправным клиентом в сети библиотеки, то есть ресурсы, находящиеся на его компьютере, становятся доступны для других клиентов. В данном случае, под ресурсами будет пониматься информация, которая поступила на его компьютер в результате запросов в сети библиотек. Также, если пользователь не обнаруживает данных в библиотеке, где он записан, автоматически попадает на ресурсы другой библиотеки, где эта информация имеется.

Таким образом, устраняется главный недостаток концепции «репликационного центра» – огромная ресурсоемкость сети. Теперь роль центра репликации будет выполнять сервер авторизации, на который, помимо этой функции будут еще возложены обязанности сбора метаданных с клиентских машин, хеширования данных контроля версии и авторского права. Согласно терминологии, подобный сервер принято называть *трекером*. Сам же обмен реплицированными данными будет происходить согласно протоколу P2P как показано на рис. 2

Таким образом, гибридная P2P сеть удовлетворяет требованиям задач репликации. Отличие от всех предложенных ранее методов, заключается в том, что отсутствует единый центр репликации и обязательная репликация как таковая – она происходит только по запросу требуемых данных.

## Литература

1. Таненбаум Э., ван Стеен М. Распределенные системы: принципы и парадигмы / Э. Таненбаум, М. ван Стеен. – Спб.: Питер, 2003. – 876 с.
2. Халабия М. Л., Халабия Р. Ф. Проблемы репликации данных в информационных системах библиотек / Халабия М. Л., Халабия Р. Ф // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: материалы конф., Судак, 10-18 июня, 2006. – Электронный ресурс: <http://gpntb.ru/win/inter-events/crimea2006/disk2/084.pdf>