

Двуранговая файлообменная сеть общего назначения

Никулин Дмитрий

Класс: 11

Секция информатики

ГОУ СОШ №564, г. Санкт-Петербург

Введение

В последнее время все более актуальными становятся приложения, основанные на идее peer-to-peer - или, кратко, P2P. Действительно, большинство данных в сети Интернет сейчас передается по схеме клиент-сервер[1], из-за чего возникает проблема неравномерного распределения нагрузки в глобальной сети. С каждым годом эта нагрузка увеличивается экспоненциально. В связи с этим, возникает необходимость в принципиально новых моделях сетевого взаимодействия. Одной из таких моделей и стал P2P: исчезновение привилегированных узлов в лице серверов и уход от централизованности к равноправному взаимодействию стали его отличительными особенностями.

За последние 10 лет были разработаны многочисленные приложения, так или иначе взаимодействующие друг с другом по принципу P2P. Одни завоевали огромную популярность среди пользователей Интернета. Главным примером здесь, бесспорно, должен служить BitTorrent[2], однако существуют также и другие, менее известные, но, тем не менее, весьма популярные сети, такие как Gnutella[3] и eDonkey[4]. Другие стали нишевыми продуктами, широко известными в узких кругах. Здесь в качестве примеров можно назвать сети Freenet[5] и I2P[6], чьей главной задачей стала максимальная анонимность пользователей. Однако, существующие на данный момент файлообменные сети обладают рядом недостатков. В частности, все они не имеют встроенной системы поиска по содержимому. Либо имеют, но работает она крайне медленно, или же результаты поиска не являются полными. Мы предлагаем прототип файлообменной сети, решающей данную проблему, а также включающей в себя ряд других интересных концепций.

Описание работы

Все существующие подходы к поиску в файлообменных сетях основаны, как правило, на централизованной индексации сети ограниченным числом серверов или на обмене между пользователями списками файлов, которыми они готовы поделиться. Таким образом, нарушается либо принцип децентрализованности, присущий P2P, либо поиск становится возможен лишь в сильно ограниченном пространстве файлов, находящихся у ближайших соседей, либо любой поисковый запрос вызывает огромную нагрузку на сеть. Мы предлагаем принципиально иную систему поиска.

В нашей сети каждый узел хранит у себя набор документов, описывающих некоторую единицу информации: видеофайл, программу или музыкальный трек - которой у него может и не быть. Подобный документ мы называем метафайлом. В метафайле, помимо названия описываемого объекта, хранятся его категория, краткое описание, а также некоторое количество служебной информации. Кроме того, формат метафайлов включает в себя систему тегов, которая несколько расширяет общепринятую, позволяя более эффективно структурировать информацию.

Для обеспечения высокой доступности файлов и метафайлов используется разработанная нами система распределения данных. Она позволяет узлам при сравнительно небольших объемах хранимой служебной информации иметь доступ к большому числу ресурсов сети. При этом доступность ресурса пропорциональна его популярности: чем чаще узлы нуждаются в некотором объекте, тем легче его получить. Кроме того, вышеуказанная система дает возможность легко и быстро скачивать ресурс по его идентификатору.

Наша сеть учитывает особенности текущей инфраструктуры сети Интернет, в связи с чем вводится разделение узлов сети на два типа: "открытые" и "закрытые". "Закрытые" узлы не могут принимать соединения, а могут лишь устанавливать их самостоятельно. "Открытые" узлы могут как принимать, так и устанавливать соединения, однако, в силу специфики "закрытых" узлов, соединяются преимущественно с ними.

Для написания прототипа мы использовали исключительное свободное программное обеспечение: язык программирования Python, библиотеку асинхронного сетевого взаимодействия Twisted, библиотеку для построения графических интерфейсов GTK и библиотеку локального поиска

Whoosh. В скором времени, мы планируем выпустить клиент для нашей сети под свободной лицензией.

Заключение

В работе построен прототип файлообменной сети, сочетающей в себе достоинства современных p2p-сетей и при этом лишенной многих их недостатков. Представлен ряд идей, претендующих на инновационные в данной области. Предложено решение для одной из фундаментальных проблем подобных сетей, а именно для вопроса поиска данных по сети. Сделан обзор текущих технологий поиска и демонстрация преимуществ нашего подхода. Обозначены основные направления дальнейшей работы.

Литература

- ^ Sandvine Global Internet Phenomena Report — Fall 2011: [http://www.sandvine.com/downloads/documents/10-26-2011_phenomena/Sandvine Global Internet Phenomena Report - Fall 2011.pdf](http://www.sandvine.com/downloads/documents/10-26-2011_phenomena/Sandvine%20Global%20Internet%20Phenomena%20Report%20-%20Fall%202011.pdf)
- ^ The BitTorrent Protocol Specification: http://bittorrent.org/beps/bep_0003.html
- ^ Gnutella Protocol RFC: <http://rfc-gnutella.sourceforge.net/rfc-gnutella.zip>
- ^ eDonkey Protocol Specification: <http://kent.dl.sourceforge.net/pdonkey/eDonkey-protocol-0.6.2.html>
- ^ The Freenet Project: <https://freenetproject.org/>
- ^ I2P Anonymous Network: <http://www.i2p2.de/>