

# 基于 P2SP 的网络下载技术分析

马 晓<sup>1</sup>, 柴艳娜<sup>1</sup>, 李 佳<sup>2</sup>

(1. 长安大学 教育技术与网络中心, 陕西 西安 710064;  
2. 长安大学 信息工程学院, 陕西 西安 710064)

**摘 要:** P2SP 即“点对服务器和点”技术, 整合了 P2P 和 P2S 技术, 同时还通过资源检索技术把传统技术的服务器资源和 P2P 资源整合到了一起, 以达到更快更稳定的下载速度和更加丰富的下载资源。文中分析了 P2SP 技术原理及其整个下载系统的体系结构与实现, 对资源检索、负载均衡、断点续传及离线下载等核心问题进行深入研究, 同时对盗链问题进行分析。P2SP 技术的自身特点, 决定其可能被滥用并影响到网络稳定, 因此合理、高效、稳定地使用 P2SP 技术是一个重要的研究方向。

**关键词:** 点对服务器和点; 资源服务器; 负载均衡; 断点续传; 离线下载

**中图分类号:** TP31

**文献标识码:** A

**文章编号:** 1673-629X(2014)06-0187-05

**doi:** 10.3969/j.issn.1673-629X.2014.06.046

## Analysis of Network Download Technology Based on P2SP

MA Xiao<sup>1</sup>, CHAI Yan-na<sup>1</sup>, LI Jia<sup>2</sup>

(1. Education Technology and Network Center, Chang'an University, Xi'an 710064, China;  
2. School of Information Engineering, Chang'an University, Xi'an 710064, China)

**Abstract:** P2SP technology, Peer to Server and to Peer, not only integrates the P2P and P2S technology, but also gathers the classical server resources and P2P resources by search technology to get faster and more stable download speeds and more abundant resources. Analyze P2SP technology theory and the whole download system architecture and realization, research the core such as resource searching, load balancing, resuming download and offline download. Some features of P2SP may affect the network stability, therefore using P2SP efficiently and stably is an important research direction.

**Key words:** P2SP; resource server; load balancing; resuming download; offline download

## 1 P2SP 技术概念

### 1.1 P2S 技术与 P2P 技术

1) P2S (Peer to Server) 技术。

P2S 即“点对服务器”技术, 是最经典的一种信息传输技术。P2S 是典型的 C/S 结构, 从使用协议上可以分为 HTTP 和 FTP 两种类型<sup>[1]</sup>。终端用户通过这两种协议与服务端通信, 将资源下载到本地。浏览网站以及下载 HTTP/FTP 文件, 都是典型的 P2S 类型应用。P2S 的下载工具一般会对文件进行切分, 从而可以使用多线程的方法成倍地提高下载速度。

2) P2P (Peer to Peer) 技术。

P2P 正好与 P2S 相反, 它是点对点技术, 没有中心服务器, 依靠用户与用户之间的直连关系进行信息交

互。在 P2P 网络中, 每个节点都是对等的, 每个用户端既是一个节点, 同时也充当服务器的角色, 从其他节点下载资源的同时也上传资源给其他节点<sup>[2]</sup>。P2P 技术的一个知名实现 BitTorrent 通过“种子”(Torrent) 的方式进行资源传播。下载同一资源的人越多, 且下载后, 继续维持上传的状态, 就可以“分享”, 成为其他用户下载的种子, 下载该资源的速度也就越快。

### 1.2 P2SP 技术

P2SP (Peer to Server and to Peer), 即“点对服务器和点”技术。“点”(Peer) 即网络节点或终端, 通常可以理解为用户的个人电脑<sup>[3]</sup>。P2SP 整合了 P2P 与 P2S 技术。在稳定性和速度方面, P2SP 协议比传统的 P2P 或 P2S 有了极大的提高。迅雷及 QQ 旋风等下载工具都是 P2SP 类型的工具。P2SP 下载工具结合资源

搜索技术和 P2P 网络,不再是传统的 P2S 类多线程下载,而是把传统服务端资源与 P2P 网络上人人分享的巨大资源进行有机融合,达到多来源多线程下载的目的。

图 1 是 P2SP 网络的组成结构,资源分布于客户端之间以及下载节点服务器之间。资源服务器保存着下载节点服务器所持有的资源索引,同时还维护着 P2P 网络中所有资源的索引,客户端之间组成 P2P 网络<sup>[4]</sup>。当用户试图下载某个特定资源时,资源服务器可以根据此资源的特征值搜索索引,告知用户所有可用下载来源(下载节点服务器或者客户端)。

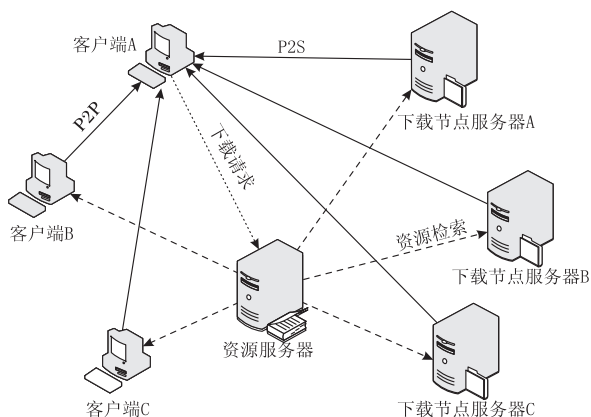


图 1 P2SP 的结构模型

P2SP 搜索发现技术根据传统服务器资源搜索以及上报,及 P2P 文件共享为基础,根据 Hash 加密算法,将所有拥有同类资源的客户端与服务器存入数据库中,当用户请求一个资源时,按照数据库中的记录,查找所有在线并拥有资源的客户端,将其返回给用户。

## 2 P2SP 下载系统

### 2.1 系统的总体结构

P2SP 下载系统由资源服务器、下载节点服务器和客户端三部分组成<sup>[5]</sup>。

#### 1) 资源服务器。

资源服务器担负着管理全网资源的职责,维护着所有资源的索引、用户列表、节点管理、下载节点间的业务请求以及各种下载状态信息。资源服务器本身并不存储资源。

#### 2) 下载节点服务器。

下载节点服务器是传统 P2S 的组成部分,本身持有资源。为了加速下载,首先下载节点由资源服务器获得拥有该资源的所有节点列表,然后根据距离远近,先后与这些节点建立连接并进行下载。下载节点服务器可以按照 CDN 技术进行布局,根据就近原则,将用户要下载的资源分发到距离用户最近、服务质量最好的节点上,从而最快地响应用户,从这个角度看下载节

点服务器之间类似于一个 P2P 网络<sup>[6]</sup>。另外,下载节点服务器之间需要时刻保持通信,确保用户的请求都能分配到可用的节点上<sup>[7]</sup>。

#### 3) 客户端。

即安装于用户端的下载工具,负责为用户提供人机操作界面,同时也是用户接入下载网络的代理终端。

### 2.2 系统的特点

#### 1) 智能资源选择。

选择合适的资源需要考虑两方面的因素:资源本身的服务能力和资源所处的网络环境。

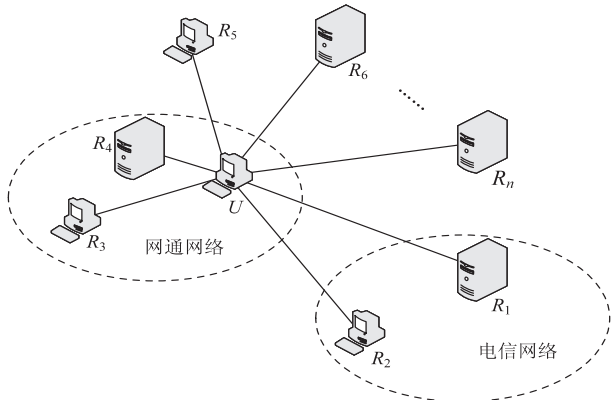


图 2 资源智能选择示意图

如图 2 所示, $U$  表示用户, $R_1, R_2, \dots, R_n$  表示候选资源,其中包括下载服务器资源和个人站点资源。在相同的网络中,下载服务器资源的服务能力要高于个人站点资源,在图 2 中, $R_4$  的服务能力要好于  $R_3$ ,用户  $U$  选择  $R_4$  可获得较高的下载速度;由于网间限速,选择与  $U$  同处的网通网络的资源比电信网络的资源更能获得较高的下载速度。

P2SP 系统对海量数据的挖掘利用,充分考虑用户的网络条件、服务器的服务能力以及当时的网络负载,实时迅速地选择最佳资源路径<sup>[8]</sup>。

#### 2) 下载服务器的负载均衡。

海量数据下载对于单个下载服务器来说,负荷过重,传统的方法提供多个下载地址,供用户自己选择,其缺点有如下两点:

① 将下载的复杂性转移给用户,用户一个个尝试,直到能够下载为止;

② 从上至下的选择,上面的服务器的压力过大,下面的基本用不上,服务器的利用率和用户的下载速度受损。

P2SP 负载均衡技术自动根据当前服务器的压力进行负载均衡:

①  $U$  的下载会自动地在下载服务器的列表中选择负载最低的服务器;

② 在保证  $U$  本身下载速度的前提下, $U$  可能使用  $U$  的下载服务器;

③ 防止  $U$  不提供下载服务,大量盗用  $U$  资源,自动做网站进行分级,降低甚至取消对贡献小的或者不贡献的网站加速。

2.3 断点续传技术

断点续传技术按照一定的规则,将一个文件拆分为若干个很小的部分,然后多线程分批次下载,每个小块完成后都会合并到文件中<sup>[9]</sup>。在下载过程中断后,重新开始下载时,通过文件字节的长度来判断下载的起点,然后重启断点续传的过程,直到完成下载为止。

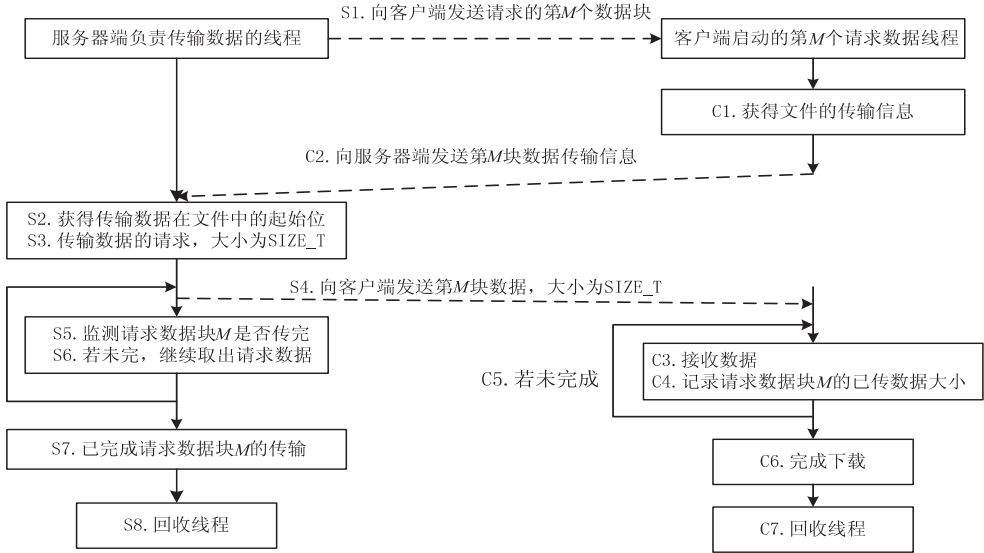


图3 断点续传过程示意图

当线程接收到应答后,便发送开始数据传输的请求。以 HTTP 传输协议为例,此时首先向服务器发送 GET 请求,在 GET 请求中加入 Range 域,告知服务器其所要数据的起始位和结束位。当服务器接收到该请求后,便向客户端发送 Range 所指定范围的数据及其他信息(文件传输的位置和已发送的字节数)。线程继续向服务器发送这种 GET 请求,直到其负责的下载任务完成,这就是 HTTP 分段下载。

实际上,在下载过程中还会新建一个临时文件用于记录下载过程的各种信息,比如下载文件的大小、每个线程所负责下载的数据段以及已下载的部分等,以保证下载程序因为某种原因(如网络阻塞、断电等)突然终止,便可根据该临时文件来恢复现场,继续之前的下载过程。这样,已下载完成的部分便可避免重复下载,特别是对于大文件,这是非常重要的。如果下载突然中断,已下载的部分将被写入这个临时文件。在下载完成的时候,该临时文件会被真正的资源文件所代替。

3 P2SP 下载服务的策略分析

3.1 P2SP 下载分析

基于 P2SP 的下载流程如图 4 所示,其中包括获取

断点续传不但可以节约下载时间,还可以减少不必要的网络开销,提升用户体验。

图 3 为文件分割后的第  $M$  块传输的过程,即断点续传的流程。客户端获知文件的大小后,按照设定的分块策略对文件进行切割,从而为每个线程分配一部分下载任务<sup>[10]</sup>。然后,逐一创建线程,每个线程均尝试与目标服务器连接,当服务器接收到连接请求后,会返回一个应答。

原始资源链接、获取网络资源列表、下载资源数据等几个过程。

1) 获取原始资源链接。

用户在网络上利用搜索引擎或者专业的下载站点获得资源链接,下载工具在获得该链接后,便开始向资源所在服务器发起 HTTP/FTP 请求,从原始资源地址获取数据。

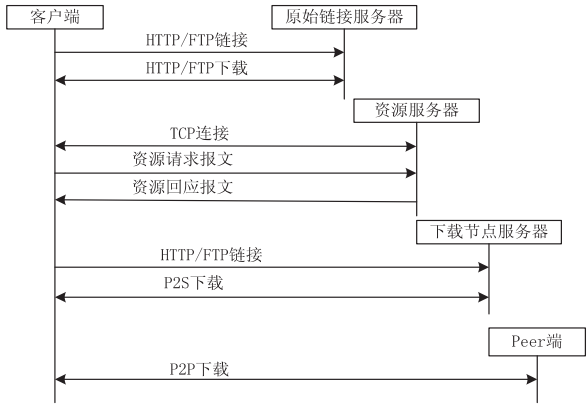


图4 P2SP 下载机制

2) 获取网络资源列表。

客户端根据原始资源的相关信息计算 Hash 值,此值是该资源的唯一标识。通过 Hash 值向资源服务器发送资源请求报文(POST 报文,如表 1 所示),请求网

络上具有该资源的其他节点服务器列表和在线 Peer 列表<sup>[11]</sup>。表 1 中资源 ID 用来标识资源的唯一性，Peer ID 用来标识客户端的唯一性。

表 1 POST 报文字段信息表

POST 报文字段信息
资源 ID(Hash 值)
资源原始链接地址
引用地址
Peer ID

资源服务器根据资源请求报文信息在其数据库中搜索,查看是否存在与用户指定文件匹配的资源记录:

① 若无,则将报文中的相关信息作为一条新的资源记录添加至数据库,为下一个用户加速;

② 若有,资源服务器向客户端返回资源回报文,包括节点服务器链接地址列表和在线 Peer 列表,列表中包含有资源的 HTTP 链接和 FTP 链接等信息。

3)客户端从下载节点服务器获取资源。

客户端从响应报文中得知资源所在的节点服务器,便尝试和这些节点服务器地址建立 HTTP/FTP 链接,开始下载数据。

4)客户端从在线 Peer 端获取资源。

客户端获取 Peer 在线列表信息后,向在线的 Peer 端相应的端口请求 TCP 和 UDP 连接,如果连接成功,则从 Peer 端下载数据,加速资源的下载。

5)停止或退出。

当暂停或退出时,客户端会向服务器端发送停止报文,服务器返回确认。

在下载还没完成时,在放置资源的路径下,下载工具开始生成临时文件,比如迅雷会为每个正在下载的资源生成一个名为“资源名称.td.cfg”的临时文件,保存着下载过程的所有信息,如表 2 所示。当下载完成时,cfg 文件会自动删除。

表 2 CFG 文件信息表

文件扩展名	文件中的信息
td.cfg	提供下载站点的 URL 链接、文件存放的位置、文件名、已下载的数据信息等
	提供下载的 Peer 端的 ID、MAC、IP 及端口信息等

在下载恢复时,根据 CFG 文件中的信息,采用断点续传技术进行下载。

3.2 离线下载分析

离线下载也称为代理下载,其实就是下载工具的服务器代替用户先行下载。离线下载利用离线服务器的强大带宽,将目标资源代理下载到离线空间,用户再从离线服务器高速下载到本地<sup>[12]</sup>。

离线下载的基本流程如下:

(1)用户通过客户端或 Web 界面向资源服务器提交了一个下载请求。

(2)资源服务端接收合法请求后,首先查询用户提交的下载链接是否被下载过:

① 若无,开启多次线程实施下载或用 P2P 方式下载;

② 若有,把已下载的数据文件(或链接),放入离线服务器的用户在线空间。

(3)下载完成后,用户在线登录到离线服务器的在线空间,取回下载的文件。其间也可以采取 P2P 方式,从已下载或正在下载相同文件的用户那里取得数据。

离线下载针对于冷门资源或者资源较少的文件,待服务器端用户下载完成后,用户还需要用下载软件从服务器上下载文件,相对于直接下载,增加了下载资源的速度,节约了时间。

4 盗链问题

4.1 盗链技术

盗链是一种非授权的资源获取手段,资源提供者自身不提供资源,通过技术等各种手段绕过其他第三方提供者的资源页面,直接盗取其资源链接,提高自身的浏览量和点击率。

盗链损害了资源原始拥有者的利益,却因为泛滥的盗链而给很多骗子带来可观的收益,同时又因为盗链鱼目混杂,用户往往下载的不是自己想要的资源,更有甚者下载了病毒或木马,增加了用户计算机安全风险。

4.2 反盗链措施

对于盗链问题,可以采取以下措施来减少或避免,维护自己的合法权益。

1)更改文件属性。

对于 HTTP/FTP 类型的下载资源,修改资源的存储路径即可避免被盗;P2SP 通过 Hash 值判断是否为同一文件,通过修改文件的路径、大小等属性,可以改变文件的 Hash 值,从而避免 P2SP 的文件下载。

2)动态服务端技术。

动态服务端技术是目前最常用的一种反盗链技术,其资源下载链接是一个动态页面,如 ASP、JSP 和 PHP 等。当用户请求下载时,该动态页面会检查用户的请求是否授权、是否有效<sup>[13]</sup>。用户不能够直接下载资源文件,只有在通过授权后才能下载。

5 结束语

相比传统的 P2S 和 P2P 下载,基于 P2SP 的下载



系统在资源利用率、下载速度和稳定性上有了显著提高,但是也不可避免地会占用大量的带宽,容易堵塞网络。同时,崇尚自由免费的分享精神,以及一直存在的盗链问题,导致资源得不到有效的管理和约束,更容易引发网络安全和版权问题<sup>[14]</sup>。因此,在保证用户安全和资源提供者利益的前提下,如何最大限度利用网络带宽,如何安全有效的管理资源和分发资源将是P2SP技术所亟需解决的问题。

参考文献:

[1] Wikipedia. P2S[EB/OL]. 2013-06-22. <http://zh.wikipedia.org/wiki/P2S>.

[2] Wikipedia. Peer-to-peer[EB/OL]. 2013-06-22. <http://en.wikipedia.org/wiki/Peer-to-peer>.

[3] Wikipedia. P2SP[EB/OL]. 2013-06-22. <http://zh.wikipedia.org/wiki/P2SP>.

[4] 陈淑英. 迅雷 P2SP 架构及服务策略的分析[D]. 北京: 北京交通大学, 2007.

[5] 侯峰. 基于 P2SP 技术的下载系统分析-以迅雷为例[J]. 河南财政税务高等专科学校学报, 2011, 25(5): 88-90.

[6] 马鸿雁, 宋海歌. 基于 P2SP 的流媒体技术的研究[J]. 电脑学习, 2009(4): 107-108.

[7] 许冬海. 基于 CDN 的 P2SP 下载系统的研究与实现[D].

(上接第 186 页)

[3] Dobre C. Context-aware platform for integrated mobile services[C]//Proc of international conference on emerging intelligence data and web technologies. Tirana; IEEE, 2011: 198-203.

[4] Zhang Yanshuai, He Hong. Intelligent business assistant based on context-aware computing platform[C]//Proc of 5th international conference on pervasive computing and applications. Maribor; IEEE, 2010: 382-387.

[5] Devaraju A, Hoh S. Ontology-based context modeling for user-centered context-aware services platform[C]//Proc of international symposium on information technology. Kuala Lumpur, Malaysia; IEEE, 2008: 1-7.

[6] Yuan Soe-Tsyr, Peng Kai-Hsiang. Location based and customized voice information service for mobile community[J]. Information Systems Frontiers, 2004, 6(4): 297-311.

[7] Giaffreda R, Barria J. Collaborative context-awareness and reasoning for optimized service delivery[C]//Proc of IEEE 65th vehicular technology conference. Dublin; IEEE, 2007: 252-256.

南昌: 江西师范大学, 2011.

[8] 张智, 刘涤. P2P 与 Web 服务技术融合的研究[J]. 计算机技术与发展, 2011, 21(8): 105-108.

[9] 刘业, 刘林峰. 结构化 P2P 网络可用性增强问题研究[J]. 计算机技术与发展, 2012, 22(3): 222-225.

[10] 崔水龙. 基于多线程的网络文件下载[D]. 长春: 吉林大学, 2009.

[11] Ranjan R, Harwood A, Buyya R. A study on peer-to-peer based discovery of grid resource information[C]//Proceedings of the 2nd IEEE/ACM international symposium on cluster computing and the grid. [s. l.]; IEEE Computer Society, 2002.

[12] Filali I, Bongiovanni F, Huet F, et al. A survey of structured P2P systems for RDF data storage and retrieval[C]//Proc of large-scale data- and knowledge-centered systems III. [s. l.]; Springer, 2011: 20-55.

[13] Rowstron A I T, Druschel P. Pastry: scalable, decentralized object location, and routing for large-scale peer-to-peer systems[C]//Proc of IFIP/ACM international conference on distributed systems platforms. London, UK; Springer-Verlag, 2001: 329-350.

[14] Biddle P, England P, Peinado M, et al. The darknet and the future of content distribution[C]//Proc of 2002 ACM workshop on digital rights management. [s. l.]; [s. n.], 2002.

[8] 陈媛嫒, 刘正捷. 移动情境感知及其交互研究[J]. 计算机应用研究, 2011, 28(12): 4420-4425.

[9] Mao Hongyan, Jiang Ningkan, Su Wen, et al. A context aware modeling framework for pervasive applications[C]//Proc of international conference on cloud and service computing. Shanghai; IEEE, 2012: 40-44.

[10] 张剑. 移动环境下情境感知群管理-以 Mobile Life 项目为例[J]. 图书馆学研究, 2012(21): 20-24.

[11] Afridi A H, Pakistan P. Crowd sourcing in mobile: a three stage context based process[C]//Proc of IEEE ninth international conference on dependable, autonomic and secure computing. Sydney; IEEE, 2011: 242-245.

[12] 孟昭鹏. 基于情境感知的移动学习平台设计与实现[D]. 天津: 天津大学, 2012.

[13] 郭斌, 张大庆, 於志文, 等. 数字脚印与“社群智能”[J]. 中国计算机学会通信, 2011, 7(3): 53-59.

[14] 牟权, 叶保留, 陆桑璐. 基于云计算的普适服务集成平台技术研究[C]//第六届和谐人机环境联合学术会议. 南京: 南京大学计算机软件新技术国家重点实验室, 2010.

基于P2 SP的网络下载技术分析

作者：

马晓，柴艳娜，李佳，MA Xiao, CHAI Yan-na, LI Jia

作者单位：

马晓, 柴艳娜, MA Xiao, CHAI Yan-na (长安大学 教育技术与网络中心, 陕西 西安, 710064)

，李佳, LI Jia (长安大学 信息工程学院, 陕西 西安, 710064)

刊名：

计算机技术与发展

ISTIC

英文刊名：

Computer Technology and Development

年，卷(期)：

2014 (6)

本文链接：[http://d.g.wanfangdata.com.cn/Periodical\\_wjfz201406046.aspx](http://d.g.wanfangdata.com.cn/Periodical_wjfz201406046.aspx)