**发明名称：一种基于混合协议的大规模文件分发的方法和装置**

**一、 发明目的**

**1.1解决的技术问题以及应用该发明的产品或项目：**

近十几年，基于HTTP/HTTPS协议的文件分发取代FTP等协议，成为主流的文件分发方式，该方式具备协议实现简单、跨平台兼容性好、并发下载等优点。但是随着分发规模的扩大，传统的基于HTTP/HTTPS的文件分发方法，当面临当前亿万级别的分发规模时，遇到如下问题：

1. 对分发源站的压力大，所有分发目的端均需要从源站下载完整文件，分发规模越大，对源站的压力越大。
2. 带宽费用高，目前带宽通常有流量计费和峰值带宽计费两种模式。不管哪种模式，费用和分发规模以及文件大成正比，分发规模的扩大带来成本的极速提高。
3. 分发速度慢，因为源站带宽和性能限制，分发速度是分发规模的反比，所以同样随着分发规模的扩大，分发速度大幅减慢。

为解决在亿万级别规模的文件分发场景下的如上三个问题，本文设计了一种基于混合协议的大规模文件分发的方法以及装置，并应用在公司内部变更系统部署包分发场景中，未来将会应用在云存储、网盘等场景，从而大幅节约成本，提高分发速度。

**1.2 现有技术的实现方案：**

现有技术为解决如上三个问题，有三种实现方案：

1. CDN方案。其原理是将文件先分发到成百上千个CDN边缘节点上，然后再就近分发到各个分发目的地端，从而降低源站压力并提升分发速度。同时因为边缘节点的带宽价格一般低于源站，所以也有一定的成本节约。CDN方案的优点是无需对原来的分发目的端进行修改，依然基于HTTP/HTTPS协议。
2. P2P 方案。P2P点对点协议的原理是将文件拆成固定大小的Block，然后让已经下载到部分Block的分发目的端向其他尚未下载Block的分发目的端进行传输。从而不仅降低的源站压力、节约源站带宽，同时随着分发规模扩大，分发速度反而予以提升。
3. Hybird P2P-CDN方案。P2P方案存在需要持续做种以及在部分网络环境下被封禁限制的问题，故有Hybird P2P-CDN方案，即CDN节点除提供HTTP/HTTPS下载外，也作为P2P的分发源持续做种，从而兼取CDN和P2P方案的优点。

**1.3 现有技术的缺点：**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 技术方案 | 源站压力 | 带宽费用 | 分发速度 | 其他优点 | 其他缺点 |
| 直接源站下载 | 极高 | 极高 | 极低 | 协议简单 | 无 |
| CDN方案 | 低 | 高 | 中 | 无需修改分发协议 | 依赖于CDN厂商的服务稳定性 |
| P2P方案 | 低 | 低 | 高 | 无 | 需要修改分发协议  需要源站持续做种机制  部分网络环境兼容性差，下载成功率低  分发启动速度较慢 |
| Hybird P2P-CDN方案 | 低 | 中 | 极高 | 适应多种网络环境  分发启动速度快  无需源站持续做种 | 需要修改分发协议  技术上依赖于CDN厂商解决方案，无法自主可控。  价格上依赖于CDN厂商的定价，成本高于P2P方案。 |
| 本文混合协议方案 | 低 | 低 | 高 | 除Hybird P2P-CDN方案优势外，本文方案无需依赖于CDN厂商的解决方案，并兼容原有的任何服务商的CDN或者源站分发方式。  从而不仅下载成功率高、下载速度快外，成本也是方案中最低的。 | 需要修改分发协议 |

。

**二、发明内容（本发明技术方案的详细阐述）**

* 1. **本发明提供的完整技术实现方案**

下面详细描述本发明的实施例，实施例的示例在2.2附图中示出。下面通过参考附图描述的实施例是示例性的，旨在解释本发明，而不能理解对本发明的限制。

如图1所示，本发明实施例的大规模文件分发系统，包括如下组件：控制服务器1，多个中继服务器2，数据库3和多个客户端4。其中控制服器务1和数据库3位于IDC机房或者云平台上；中继服务器2可以位于IDC机房、云平台上，也可以位于CDN边缘节点上；多个客户端4位于分发目的端的计算机设备上。而文件分发的原始数据位于数据源服务器5上，我们将使用文件分发的用户称为用户0。

以一次典型的文件分发业务流程为例，将整个流程拆分为如下步骤（参见图1）：

步骤一（图例中的step1），用户0向控制服务器1发送指令，要求将数据源服务器5上的文件分发到多个客户端4上，并同时提供分发配置。

步骤二（step2），控制服务器1创建配套的中继任务和分发任务，并持久化到数据库3中。

步骤三（step3），控制服务器1将分发任务下发到目标的多个客户端4上，并根据客户端的分布和数据，挑选多个中继服务器4，将中继任务下发到这些中继服务器4上。

步骤四（step4），中继服务器2从数据源服务器4中拉取需要分发的目标文件，在拉取的同时提供HTTP协议和BitTorrent协议的数据下载服务。

步骤五（step5），客户端4根据分发配置，和中继服务器2以及其他的客户端4组成BitTorrent下载网络（简称BT协议）进行数据下载，如果因为客户端4所在网络限制BT协议或者BT下载速度过慢，还可以辅以从中继服务2或数据源服务器4上通过HTTP协议下载文件作为BT协议的补充。分发过程中，客户端4和控制服务器1进行交互，从而实现下载、上传以及源服务器压力的全局控制。

在上述各个步骤中，本发明实现了三处技术方案，从而让本发明比其他文件分发有明显的优势。具体技术方案如下：

1. 客户端下载速度和质量保证技术： 客户端默认使用BT协议下载，当下载速度低于预设的配置值时，客户端可以通过HTTP协议从中继服务器2或数据源服务器4上进行辅助下载，从而保证下载速度满足预设要求。
2. 数据源服务器4压力精细化控制技术： 控制服务器1可以配置对数据源服务器4的全局下载并发数、下载带宽、总下载流量。在下载过程中，控制服务器1向客户端4分配并发度、带宽、流量配额，从而保证全局限制得到满足。
3. 下载快启动技术：在步骤四中，中继服务器2从数据源服务器4中拉取需要分发的目标文件的同时，就可以提供数据流给客户端4用于下载，所以客户端4下载启动时间较快。

**2.2 附图说明**

**（为了方便理解发明内容，请通过流程图或者系统框图的方式阐明技术方案的实现步骤或者结构。）**

**2.3 本技术方案对产品的帮助**

通过本技术方案，对大规模文件分发场景有如下帮助：

1. 对源站/CDN的压力进行精细化控制，从而有效控制源站/CDN成本
2. 带宽成本远低于CDN方案，同时可以和CDN配合，精细控制带宽比例，实现成本和分发效率的最佳平衡。
3. 平均下载速度、下载启动速度均显著高于其他方案，同时还可以提供最低下载速度保障。
4. 网络环境兼容度高，兼容多种网络。
5. 和CDN服务商无关，可配合任意CDN甚至无CDN的服务。可不对原分发网络进行任何修改，旁路加速，从而让分发技术方案切换的成本降到最低。

**2.4、针对2.1中的技术方案，是否还有别的替代方案同样能完成发明目的，如有，请列出。**

参考1.2和1.3，现有技术方案的缺点已经列出。

**三、业界相关产品及现有技术检索：**

**3.1 与该技术相关的竞争对手或相关产品：**

**（请列出竞争对手的名称、相关产品的名称，对于相关产品也可以采用截图等方式）**

各CDN厂商，如阿里云、网宿、腾讯云等。

**3.2 相关现有技术，重点是已公开专利及科技论文**

**（如专利/论文/标准，可以用附件、链接、文献名称及出处等方式提供）**

* **请列出使用的中英文关键词（或组合）：**

Hybird P2P CDN

Content Distribution

混合文件分发

* **跟本发明相关的现有专利或者论文名称及链接：**
  + Xu, D., Kulkarni, S. S., Rosenberg, C., & Chai, H. K. (2006). Analysis of a CDN–P2P hybrid architecture for cost-effective streaming media distribution. *Multimedia Systems*, *11*(4), 383-399.
  + Jiang, H., Li, J., Li, Z., & Bai, X. (2009). Efficient large-scale content distribution with combination of CDN and P2P networks. *International Journal of Hybrid Information Technology*, *2*(2), 4.
  + El Dick, M., Pacitti, E., & Kemme, B. (2009, March). Flower-CDN: a hybrid P2P overlay for efficient query processing in CDN. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology* (pp. 427-438).
  + Munther, A., Noori, S., Osman, A., Hussain, A., Jasim, I., & Shanoon, A. (2012). Peer-to-peer video conferencing using hybrid content distribution model. *Journal of Computer Science*, *8*(7), 1134.
  + Jiang, H., Li, J., Li, Z., & Bai, X. (2008, December). Performance evaluation of content distribution in hybrid CDN-P2P network. In *2008 Second International Conference on Future Generation Communication and Networking* (Vol. 1, pp. 188-193). IEEE.
  + Afergan, M. M., Leighton, F. T., & Parikh, J. G. (2012). *U.S. Patent No. 8,332,484*. Washington, DC: U.S. Patent and Trademark Office.

* **跟上述专利或者论文内容的核心区别点：**

见1.3，和Hybird P2P-CDN核心区别点是本文技术方案属于旁路运行，和CDN服务商无强绑定；同时可以对带宽进行精细控制，使成本低于 Hybird P2P-CDN 方案。