迅雷下载连接识别的方法与限制的实现

- 摘要

随着迅雷支持的下载协议的增加，有越来越多的网民使用迅雷作为自己的下载工具。但迅雷使用率的增长也给不同的网站造成了不同程度的影响，带宽有限的小型网站受到的影响较为严重。本文通过介绍和提出几个方法来限制迅雷对网站资源的大量占用，以帮助小型网站不再因迅雷的盗链致使流量被大量占用。

- 关键词

迅雷 下载 流量 限制

第一章前言

1.必要性

小型网站一般所占有的网络带宽和服务器资源有限，不足以提供较大流量的下载服务。但一般小型网站也会根据自身的内容，提供一些相关的下载，其中有可能有较大的文件。因为小型网站的访问量不高，因此通过正常途径获取下载地址的用户进行的下载一般不会对小型网站造成太大影响。但小型网站所提供的下载地址一旦被迅雷收录，就有可能造成大量与此网站无关的用户直接下载这些小型网站的资源，超出了小型网站所能承受的流量，这就会对小型网站的正常运

行造成严重影响。

2.预期目的

通过使用本文介绍和提出的一系列技术手段，希望能识别部分由迅雷发起的 HTTP 请求，并对其实行拒绝或抛弃处理，同时尽量避免对正常用户的浏览和下载造成影响，以达到限制迅雷对网站资源的下载，减少由迅雷引起的网络带宽和服务器资源的浪费，保证网站稳定运行的目的。

第二章 技术分析

1.请求识别

迅雷是一种不通过网页浏览这一环节就直接请求下载资源的下载工具。由于其跳过了网页浏览这一行环节，我们就可以通过一些手段将其与对网页进行过浏览操作的普通用户区分开来，从而识别出迅雷的请求。

另一方面，截止到 5.7.3.389 版本为止，迅雷所发送的用户代理（User-Agent）字符串一直没有变化，而这个用户代理字符串是一个干净的 Windows XP 系统所特有的。通过对每个请求所发送的用户代理字符串进行分析，也可以在一定的误差程度内识别出由迅雷发起的请求。

在 FTP 协议下，迅雷的表现方式也与其他 FTP 客户端有所不同，通过这些差别就可以识别出迅雷发起的 FTP 连接并拒绝。

2.限制请求

通过识别出迅雷的请求，就可以对该请求执行抛弃或拒绝操作，以限制迅雷对网站资源的下载。一般来说，对于通过程序读取保存在磁盘上的文件以后转发到客户端的下载方法，可以在程序中容易地实现对请求的限制。对于地址是直接公开的方法，也可以通过在更低的层次上编写服务器端扩展来对请求进行限制。

第三章HTTP方式下的具体实现

1.方法介绍及使用

这里介绍了几种在HTTP服务上防止迅雷下载的方法，并重点介绍用户代理判断方法。

1.用户代理判断方法

原理说明

对于每一个HTTP请求，客户端都会发送自己的用户代理，用户代理中一般包括了操作系统信息、浏览器信息以及其他的一些信息。由于不同的操作系统和浏览器版本的不同，不同的浏览器发送的用户代理也不尽相同。但是下载工具并不是浏览器，因此没有自己的浏览器信息。目前使用较为广泛的下载工具都默认发送了已经预定好的用户代理，这个用户代理一般仅带有操作系统和浏览器信息（多为IE），迅雷也是使用了这样的方法。

用户在使用电脑的过程中，常常会往电脑内安装一些软件或使用不同的浏览器，这样就有可能改变用户经常使用的浏览器所发送的用户代理。而与此同时，下载工具所发送的用户代理是不会变化的，这就造成了迅雷发送的用户代理的特殊性。

因为迅雷的用户代理是特殊的，我们就可以将其与其他软件发送的用户代理区分出来。通过观察，我们发现迅雷不同的版本最常使用的是以下 4 个用户代理（截止到 2008 年 4 月 8 日观察到的数据）：

用户代理（User-Agent）	版本
Mozilla/4.0 (compatible; MSIE 6.0; Windows+NT 5.1)	Xunlei V5.7.3.389 或更早
Mozilla/4.0 (compatible; MSIE 6.0; Windows+NT 5.0)	Xunlei V5.7.7.441 或到 Xunlei V5.7.9.466 之前（不含）
Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1;)	Xunlei V5.7.9.466 之后（含）
Mozilla/5.0 (compatible; MSIE 6.0; Windows+NT 5.0)	Web Xunlei

表 3-1 不同版本的迅雷发送的用户代理

通过服务器端的编程，我们可以很容易地获取用户代理，将获得的用户代理与以上 4 个用户代理进行对比，就可以识别出迅雷的链接。

识别后的处理方法

在确定一个连接是由迅雷发起的以后，我们可以采取这样一些措施。

- 断开此连接
- 断开此连接，并将其 IP 屏蔽一段时间
- 重定向到提示页面

第一种方法是比较直接的，也是较为简单的。但如果迅雷不断地进行连接，对于一些下载量较大的资源来说，也会对服务器造成一定的压力。如果是这种情况，可以考虑使用第二种方法。

如果你还有别的考虑，也可以考虑使用第三种方法，或者是其他的这里没有提及的方法。

对正常用户的影响

虽然使用用户代理判断方法可以较为简单地识别出迅雷发起的连接，但这种方法存在的最大问题是会给小部分正常用户的正常访问操作造成影响。在这里说明了影响产生的原因，并结合实际分析影响的范围。

造成影响的原因

不排除有这样一些正常用户，他们使用的浏览器发送的用户代理和迅雷发送的用户代理相同。在这种情况下，使用这样一种方法来限制迅雷就会给这些用户带来影响。这种影响是无法完全避免的，但可以通过一些方法来减小对这些用户的影响。

影响范围

分析网站的访问日志，可以看到每一个 HTTP 请求所请求的文件和客户端发送的用户代理。通过分析网站的访问日志，就可以得出一个大概的影响范围。

这里采用的方法是，筛选出使用了与迅雷发送的用户代理一致的请求，根据这些请求所请求的文件进行判断。若请求的是大型文件，则认为是迅雷，否则认为是正常用户。

下面的数据给出了小樱之町论坛（http://bbs.bbxy.net/)从 2007 年8月 14 日到 2008 年 3 月 2 日的访问记录分析结果。

	Xunlei V5.7.3.389 或更早	Xunlei V5.7.7.441 或更晚	Web Xunlei	合计
用户代理异于迅雷的正常用户	78008	34470	0	112533
日志文件总大小（不区分迅雷）	8613569			
用户代理与迅雷相同的正常用户占总数的比例	0.906%	0.400%	0	**1.306%**

表 3-2 与迅雷发送相同的用户代理的访问分布

正常用户的数据是根据合计数据减去所有迅雷的数据得到的。在统计的时候日志文件总大小的数据已被人为缩小（2008 年 3 月的数据没有统计），但用户代理与迅雷相同的正常用户的统计数据样本来源没有缩小，因此实际上使用和迅雷相同的正常用户的比例会更小。

统计数据以 KB 为单位进行计算，这不会对结果造成不可忽略的影响。

若请求的文件扩展名为“zip”、“rar”或“mp3”则认为是大型文件。

减小影响的方法

普通用户的大部分操作是浏览网页而不是进行下载，而迅雷的主要操作是下载而不是浏览网页。因此，我们可以加上这样一个判断：

若一个疑为迅雷的连接请求的不是下载文件，则允许此连接，否则断开。加上这个判断以后，我们可以保证这部分用户除了下载文件以外，可以正常地使用网站所提供的其他服务。

考虑到迅雷连接到其他索引时绝大部分都是进行续传请求（即带 Range 的 HTTP 请求），我们可以仅限制疑为迅雷的连接的续传请求。这样可以保证普通用户正常进行无断点续传支持的下载操作，而限制迅雷通过索引链接进行的大部分下载。

应用举例

服务端编程的方法很多，可以利用动态网页技术，也可以使用 HTTP 服务程序所提供的接口。以下以 Apace、PHP 和 ISAPI 进行简单的举例。

使用 Apache 的 mod_rewrite 进行限制

mod_rewrite 是 Apache HTTP Server 的扩展模块。提供格式化的重写引擎对用户请求的资源地址的重写，可以使用多种环境变量。在这里，我们就使用 HTTP_USER_AGENT 这一环境变量并结合正则表达式来进行判断。

Apache 配置代码

```
RewriteCond %{HTTP_USER_AGENT} ^Mozilla/4\.0\ \(compatible;\ MSIE\ 6\.0;\ Windows\ NT\ 5\.1\)$ [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^Mozilla/4\.0\ \(compatible;\ MSIE\ 6\.0;\ Windows\ NT\ 5\.0\)$ [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^Mozilla/4\.0\ \(compatible;\ MSIE\ 6\.0;\ Windows\ NT\ 5\.0\)$ [NC]
RewriteCond %{HTTP_USER_AGENT} ^Mozilla/4\.0\ \(compatible;\ MSIE\ 6\.0;\ Windows\ NT\ 5\.1;\ \)$ [NC]
```

```
RewriteRule        ^/download/.*                                        /
[NC,F]
```

代码 3-1　　使用重写模块通过用户代理识别并限制迅雷下载文件

给 Apache 应用这段设置以后，Apache 会将发送了迅雷的用户代理的连接重定向到非下载目录，从而阻止迅雷进行下载服务器上的资源。

使用 PHP 进行限制

在 PHP 中，我们可以直接在程序中读取文件，然后通过程序发送到客户端，同时不公开文件的 URL。这样我们就可以对用户隐藏文件的地址，用户必须通过 PHP 来下载文件，而不能直接向服务器请求。在 PHP 中我们只要验证用户是否使用了迅雷，如果是，则断开链接，不是，则返回文件数据到客户端。

代码的实现很简单，首先是判断用户代理，接着根据判断结果决定是否允许用户下载。

PHP 代码

```php
<?php
// 如果是迅雷则退出
if(isThunder()) {
    forbiddenThunder();
}
// 下载主程序
/* 函数定义 */
// 发送 HTTP 403 错误并给出提示，退出程序
function forbiddenThunder(){
    header('HTTP/1.1 403 Forbidden');
    header('Tip: Please DO NOT download VIA Thunder');
    exit;
}
// 根据 UserAgent 判断是否为迅雷
function isThunder(){
```

```
    $ua = trim($_SERVER['HTTP_USER_AGENT']);
    return  $ua == 'Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)' ||
            $ua == 'Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.0)' ||
            $ua == 'Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; )' ||
            $ua == 'Mozilla/5.0 (compatible; MSIE 6.0; Windows NT 5.0)';
}
?>
```

代码 3-2　　使用 PHP 通过用户代理识别并限制迅雷下载文件

在用户试图下载资源时，先由 PHP 程序通过用户代理判断这是否是由迅雷发起的连接。如果判断结果为该连接可能是迅雷，则返回 HTTP 403 错误；如果不是迅雷发起的连接，则继续往下执行，返回文件的二进制流。

使用 ISAPI Filter 进行限制

ISAPI Filter（以下简称 ISAPI）是可以应用于 Internet Information Service（万维网信息服务，以下简称 IIS）的服务器端扩展。用户请求到达 IIS 之后，ISAPI 都会有机会处理这个请求。利用 ISAPI 我们就可以在 IIS 确定处理用户的请求之前识别出迅雷发起的连接并断开。

在 ISAPI 的 OnPreprocHeaders 事件里我们可以获得此连接发送的用户代理，在这里我们就可以将其与迅雷的用户代理进行对比，并据定是否断开此连接。

C++ 代码

```
DWORD CXBFilter::OnPreprocHeaders(CHttpFilterContext* pCtxt,
    PHTTP_FILTER_PREPROC_HEADERS pHeaderInfo){
    // …… 变量定义省略
    pCtxt->GetServerVariable("HTTP_USER_AGENT", UserAgent, &BufSizeUA);
    //屏蔽迅雷，若符合迅雷的用户代理且有Range请求，或者请求了一个音乐或压缩包之类的可能会很大的文件，返回给它
```

```
    bIsThunder =
                !strcmp(UserAgent, "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)")
        ||    !strcmp(UserAgent, "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.0)")
        ||    !strcmp(UserAgent, "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; )")
        ||    !strcmp(UserAgent, "Mozilla/5.0 (compatible; MSIE 6.0; Windows NT 5.0)");
    if( bIsThunder )
    {
        //这里就是屏蔽了，返回一个提示给它，然后结束连接
        pCtxt->AddResponseHeaders("Tip: 请不要使用迅雷进行下载\r\n", 0);
        pCtxt->ServerSupportFunction(SF_REQ_SEND_RESPONSE_HEADER,          "403
Forbidden", NULL , NULL);      // 设置HTTP头
        return SF_STATUS_REQ_FINISHED;      // 返回结束连接的状态码
    }
    // 返回正常的状态码
    return SF_STATUS_REQ_NEXT_NOTIFICATION;
}
```

代码 3-3　　使用 ISAPI 通过用户代理识别并限制迅雷

这段代码的作用是，在请求交由 IIS 正式处理之前，先有 ISAPI 对用户代理进行识别，如果发现此连接是由迅雷发起的话，就结束连接，否则交由 IIS 继续处理。

2.链接定期更换方法

原理说明

在某一个人使用迅雷下载了服务器上的资源以后，迅雷就会把这个资源的地址索引到迅雷的索引数据库里。在这以后，从其他地方下载同一个资源的用户就会从迅雷的索引服务器上得到该资源在你服务器上的地址。也就是说，迅雷收录的是用户从你的服务器上下载这个资源时使用的链接，其他的没有经过你的下载页的用户也只能是获得这一个链接。如果这时候我们将下载链接进行更改，在没有用户使

用迅雷通过修改后的链接下载资源之前，迅雷索引的就会保持为先前的链接。这样，迅雷用户通过迅雷获得的下载链接实际上已经成为一个无效地址，也就无法从你的服务器上下载资源了。

可能造成的影响

如果一个资源文件大小很大，使用速度较低网络的用户可能就不能在一个下载目录重命名周期内完成下载。虽然用户也许可以通过修改下载地址继续下载，但并不是所有的用户都会进行这样的操作，且这样操作也会给用户带来麻烦。

3.网页 Cookies 判断方法

原理说明

在用户访问浏览网站的普通网页时，可以给用户写入一个 Cookies，以记录该用户曾经访问过我们的网站。同时，当一个连接请求下载的时候，检查此连接的客户端有没有标识用户曾经访问过网站的 Cookies，如果有，则允许下载，否则拒绝下载。

可能造成的影响

部分用户可能已经通过正常途径访问了网页，并得到了一个标识已访问过网页的 Cookies。但这些用户下载时使用的下载工具没有向服务器发送 Cookies，这时用户就不能正常下载到资源。

4.IP 绑定方法

原理说明

针对每一个用户，根据其 IP 给出不同的下载地址，并将 IP 地址和下载地址绑定。每一个用户访问下载页面的时候，动态地根据用户的 IP 生成一个下载地址。接收到下载请求的时候，检查其请求的下载地址是否对应其 IP，对应则允许下载，否则拒绝下载。

可能造成的影响

部分用户的 IP 变动极快，可能每几分钟 IP 就会变动一次。这样的用户在 IP 地址改变以后，他的 IP 地址就和他所得到的下载地址失去对应，因而无法正常下载。对于普通的 PPPoE 拨号用户，一般他们的 IP 在计算机重新启动之后都会改变，如果这些用户在下载过程中重新启动了计算机，他们就必须重新下载。这对用户来说是很不方便的。

第四章FTP 方式下的具体实现

1.FTP 指令序列识别方法

名词定义

FTP 指令序列

在本文中，FTP 指令序列指一个 FTP 连接从连接建立到开始接

收首个下载文件的区间内，所发送的 FTP 指令中指令名部分所组成的队列。

原理说明

任何一个 FTP 连接需要下载文件，从该连接建立开始，都要通过一系列 FTP 指令向服务器发送请求，才能完成期望的操作。我们注意到，几乎每一种 FTP 客户端都有自己独特的 FTP 指令序列，迅雷也不例外。

从理论上来说，对于大部分 FTP 客户端发起的连接，即使某一个连接没有通过 CLNT 指令发送客户端信息，我们也可以通过该连接的指令序列来确定其客户端。

对于迅雷来说，我们可以实现截获其 FTP 指令序列，并在服务器端进行相应的设置。在迅雷试图从 FTP 服务器下载资源时，服务器就可以通过指令序列识别出迅雷的连接，从而阻止下载。

具体实现

迅雷的 FTP 指令序列

截止到 2008 年 4 月 8 日，我们观察到的迅雷 FTP 指令序列有如下 4 种：

- 下载文件位于根目录，直接下载

  USER PASS TYPE SIZE PASV RETR

- 下载文件位于根目录，续传下载

  USER PASS TYPE SIZE PASV REST RETR

- 下载文件位于非根目录，直接下载

  USER PASS CWD TYPE SIZE PASV RETR

- 下载文件位于非根目录，续传下载

  USER PASS CWD TYPE SIZE PASV REST RETR

具体实现方法

对于不同的 FTP 服务端软件，有不同的设置方法。具体可以参考 FTP 服务端软件的用户手册。这里仅简单介绍 Serv-U 的方法。

对于 Serv-U，可以借助其开发的插件接口来开发相应的插件。Serv-U 开放了截获每一个指令事件的接口，通过这个接口，我们可以截获每一个连接的 FTP 指令序列。在某一个连接请求下载时，将此连接的指令序列与已知的迅雷的 FTP 指令序列进行对比，如果相同，则可以认为该连接是由迅雷发起的，这时就可以不响应此连接的下载请求，还可以根据情况屏蔽此 IP。

可能造成的影响

造成影响的原因

无法确定所有的 FTP 客户端的指令序列都是唯一的，因此有可能存在与迅雷的 FTP 指令序列相同的 FTP 客户端。若这种客户端存

在，那么在我们限制迅雷下载的时候，这种客户端也会被错误地限制 。

已发现的影响

截止到 2008 年 4 月 8 日为止，我们只发现了一个这样的影响。快车 V2.0 Beta 6 版本的 FTP 指令序列与迅雷在下载 FTP 根目录时发送的 FTP 指令序列相同。这意味着，如果不加处理，在限制迅雷的同时快车 2.0 Beta 6 也会被错误地限制。

解决方法

我们可以仅限制迅雷下载非 FTP 根目录下的文件来消除这个影响。同时为了到达完全限制迅雷的目的，可以不在 FTP 根目录存放任何文件。这样处理以后，就可以限制迅雷而不给快车造成影响。

第五章其他

一些离题的问题

有人说，使用上面的方法对迅雷进行限制以后，虽然迅雷已经不能下载我的资源了，但是我的服务器还是收到了很多迅雷的请求，虽然都被一一拒绝，但请求数量之大，已经和分布式拒绝服务攻击或者 CC 攻击没有什么不同了。对于这种情况，我们认为，这都是过去没有限制迅雷下载的时候，某人通过迅雷成功地下载了你的服务器上的资源，迅雷才索引到了你的资源。从限制迅雷下载开始，你的服务器上新增的资源就不会被迅雷成功下载，也就不会被索引了。也许过一

段时间以后，迅雷长期不能从你的服务器上下载资源，就会把你从索引中删除的。如果实在等不了这么长的时间，那么可以尝试修改HTTP下载为非标准 HTTP 端口，FTP 端口也修改为非标准的。比如把 HTTP 下载的端口改成 2081，FTP 端口改成 2021。

致谢

HTTP请求识别的用户代理方法是在互联网上看到的（原文：apache 防迅雷下载/盗链，作者：est），在此特别感谢。

另外感谢以下人员对反迅雷研究的支持：

江苏南京 白水山言

广东广州 包子研究员

作者信息

网络称呼：greensea

城市：广西百色

电子邮件：dengyi910@163.com

个人主页：http://www.gsea.com.cn/

结语

其实文中提到的很多方法都是很容易被迅雷绕过的。但是，迅雷的下载方式毕竟是和普通的下载方式不同的，既然不同就一定会有差异之处，只要能找到这个差异，就可以识别出迅雷。这世上有很多问

题都是看似不可能，但只要用心去想，未必不会有解决方法。其实在开始写这篇文章的时候，FTP 指令序列识别的方法我还没有想出来，当时也是认为在 FTP 上是没有办法识别出迅雷的。可现在呢，竟然想出了这样一个 FTP 指令序列识别的方法，当时我也是比较吃惊的，没想到在 FTP 上竟然也能较为准确地识别出迅雷。所以说，这世上没有一开始做不到的事情，只有你没有认真努力去做的事情。

在写下这段文字的时候，我已经开发出了 Serv-U 和 ISAPI 的反迅雷插件，它们分别名为 GSSXI 和 GSIXI，其中的 GS 就是我的名字的缩写，中间的 S 和 I，S 是 Serv-U 的意思，I 是 ISAPI 的意思，后面两个字母 XI 是 Xunlei Immune 的意思。Gene 6 的插件也现在也已经有人开发出来了。关于这几个插件的信息，都可以在我的主页上找到。直接地址是：http://www.gsea.com.cn/gs/fanxunlei/。

最后谢谢你能看完这篇文章，希望本文能对你有用。另外，我没有仔细地检查文中的错字和其他可能的错误，如果有错还请原谅。

相信你能做到，你就一定能做到。不尝试的话，怎么能断言做不到呢？愿你能发现更好的识别迅雷的方法，与大家交流，碰撞出思想的火花（如果迅雷依然我行我素的话）。