

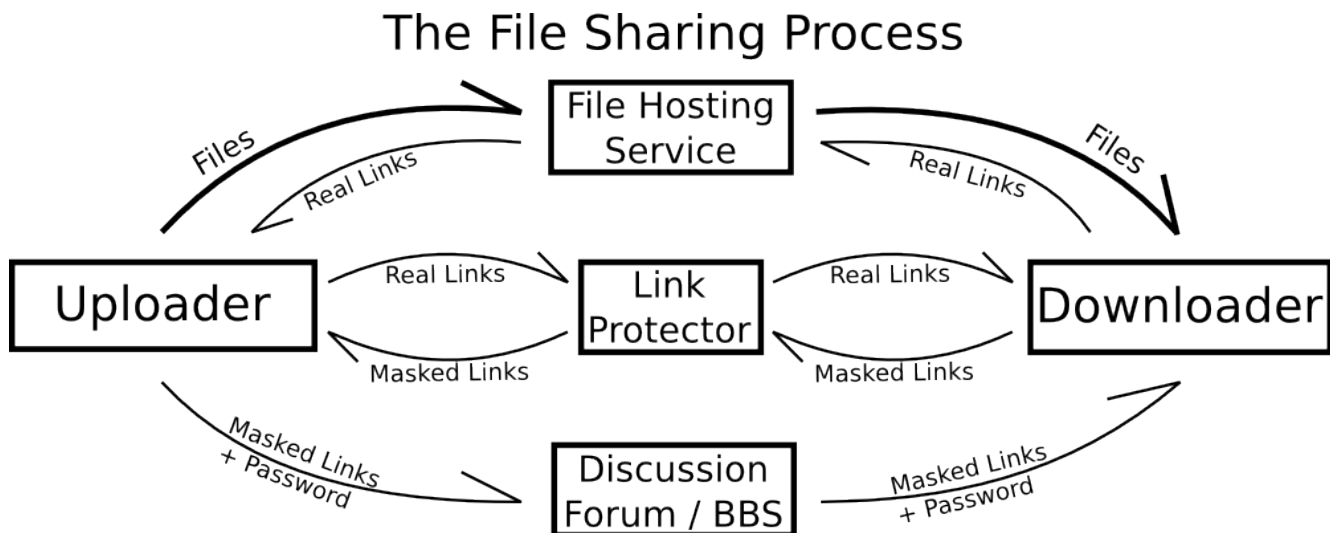
## The Reality of Modern File Sharing

Forget what you know about peer-to-peer networking. That was last century's debate. File sharing has moved on. File sharing is no longer about kids swapping music files. Those kids have grown up and are now sharing huge files of all types. From DVD and Blu-ray movies, to cracked software, anything you want can now be downloaded within minutes. This is being done through businesses called *file hosting services*. The sharers looked at the vulnerabilities and restrictions of the peer-to-peer model and evolved this new more robust regime.

### Hosting Service are not Peer-to-Peer Networks

A peer-to-peer (p2p) network exists as a web of interconnected computers run by persons wanting to share files with others running similar software. These networks function efficiently only where many connected users share identical files. This allows *swarm downloading*, whereby a single file is downloaded simultaneously from multiple peers. The downloading of rare, unique or simply unpopular files is difficult and places heavy requirements on those who do choose to share. Because files within p2p networks are stored only on user machines, when a user is offline the files he or she was sharing are also offline. Where only one user is sharing a particular file, that file is only available if that user remains online. Unique or simply unpopular files are not shared easily within p2p networks.

File hosting services function very differently. Hosting services allow users to upload files to dedicated servers within large commercial data centers. Each file is then assigned a link that is provided to the uploader. Thereafter this link can be used to download the file. It is these links that file sharers swap with each other. There are no direct file transfers between sharers. This structure has huge advantages over p2p networking. Download speeds are no longer limited by the speed of the uploader's internet connection. Popularity of files becomes irrelevant as all files are hosted in one central location where they can be simultaneously downloaded by hundreds. Most importantly, it no longer matters whether or not the original uploader remains online. These improvements make file hosting services an exponentially more powerful sharing tool than p2p.



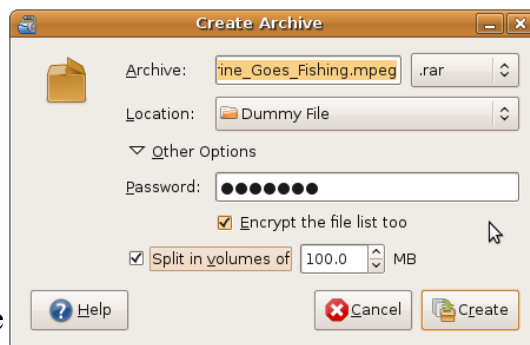
## Let's Start Sharing!

Rather than prattle on about details, I am going to show you how this stuff works. I am going to share a large movie file over my home internet connection. Anyone reading this article will be able to download this file, at any time. No magic software is required. By doing this I hope to shed some light on how complex file sharing has become and explain that intrusive enforcement regimes such as “deep packet inspection” are not the panacea their proponents claim.

The file I will share is called “A\_Wolverine\_Goes\_Fishing.mpeg”. I will first perform some prep work on the file, upload it to a hosting service called *Rapidshare*, then publish its whereabouts via a popular file sharing website, which I will not be naming. I will also take several measures to protect the file from copyright enforcement efforts. In reality this file is just 300mb of random data<sup>1</sup>, so there is no need to report me to the MPAA. For purposes of this exercise just pretend that the file is a major motion picture that I am looking to share illegally.

### Step One: Prepping the File

At three hundred megabytes (300MB) the file is too big. Rapidshare has a 200MB per-file limit if the file is to be made available to non-subscribers. To get around this I am going to cut the file into two 150MB pieces. For reasons I will explain later I am also going to encrypt the file. In the file sharing community the *rar* file format is used for both these tasks. I am using a Linux tool called *File Roller* but Windows users can perform the same tasks via *WinRAR*. Mac users ... I don't know anything about Macs, but I am sure there is a tool out there somewhere.



I now have two 150MB files. They are encrypted with the keyword “Rabbits”, which downloaders will need when reassembling the file. It is not unusual to see Blu-ray movies cut into fifty or more individual rar files, but such packages can take a week or more to prep and upload.

### Step Two: Uploading the Files

I have chosen to share my movie files via Rapidshare.com. There are others services that are arguably better/faster/cheaper services, but Rapidshare is currently top dog<sup>2</sup>. The upload process is simple. I login to the Rapidshare website, click the upload button and select the files I want to share. Rapidshare then assigns me a series of links. These links can now be used to download the files.

You are uploading:  
A\_Wolverine\_Goes\_Fishing.mpeg.part1.rar



84 % (128822 KB from 154113 KB)

Upload speed: 62 KB/Sec, Time remaining: 3:18 minutes

These are the links assigned to my files by Rapidshare:

[http://rapidshare.com/files/268545102/A\\_Wolverine\\_Goes\\_Fishing.mpeg.part1.rar](http://rapidshare.com/files/268545102/A_Wolverine_Goes_Fishing.mpeg.part1.rar)  
[http://rapidshare.com/files/268556496/A\\_Wolverine\\_Goes\\_Fishing.mpeg.part2.rar](http://rapidshare.com/files/268556496/A_Wolverine_Goes_Fishing.mpeg.part2.rar)

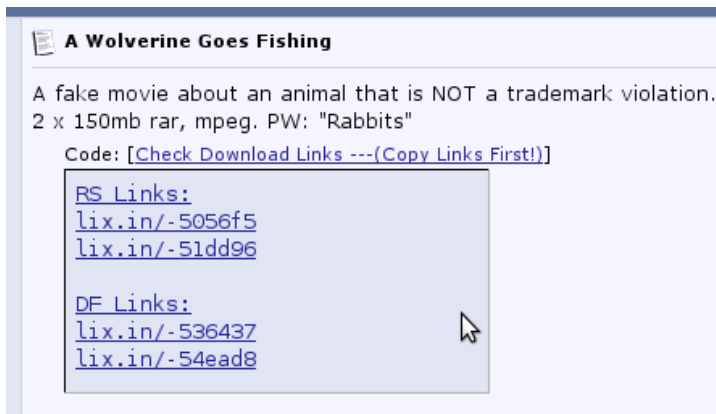
Feel free to download these files. Follow the links and click the “free user” option. You will have to wait fifteen minutes before downloading the second file.

I have also uploaded my files to a separate service called *Depositfiles*:

<http://depositfiles.com/files/gtp941917>  
<http://depositfiles.com/files/bgewj8wf9>

### *Step Three: Publishing The Links*

There are many websites dedicated to the swapping of links. Most of these are run on small budgets, some ad-supported and others by member donations. These websites take the form of discussion forums where uploaders can publish their links and receive comments from downloaders. As file hosting services do not allow users to search or browse files directly, these sites provide the needed communications link between uploaders and downloaders. Some of these sites are invitation-only affairs while others are open to the public. I am not going to point a finger at the site I used for this exercise, but they are not very hard to find.



Notice that my posted links are not those assigned by Rapidshare. This is part of a ploy by me to avoid the robot armies of the copyright police. They are in fact one-off redirection links from a service called *Lix.in*.

### *Step Four: Walking Away*

Now that the files are uploaded and the links published, I walk away. Unlike file sharing via p2p networks there is no need for me to remain online. The files now live on

Rapidshare's servers. They can be downloaded by anyone at any time and at whatever speed their home connection will allow, but I do retain some control over the situation. I could have Rapidshare deleted the files, or I could remove the published links from the discussion forum.

That's it. That's how large files are shared. Please feel free to download them via the Rapidshare links. If you are asked, click the “free user” option. Once you have both files you can reassemble them via WinRAR. The encryption key is always “Rabbits”.

### **Keeping Things Under the Radar**

The files that I have uploaded will remain available so long as they are not removed by Rapidshare. There are two ways this can happen. The file could be detected by Rapidshare's internal copyright enforcement measures, or the file could be reported via a DMCA takedown notice sent by an

agent of the copyright holder.

### *Avoiding Rapidshare's Internal Security*

Rapidshare's maintains a growing database of MD5 hashes, essentially fingerprints taken from known illegal files. Any file matching a hash in that database is automatically removed. In theory this prevents users from repeatedly uploading illegal content and allows Rapishare to remove duplicate copies that may have already been uploaded by others. In reality this is only a minor inconvenience as the system can only target files with a matching hash. I did not upload the actual movie file, I uploaded a rar archive. Any alteration to this process, by changing the encryption keyword or adding a small text file, would radically alter the hash values of the resulting files and render them invisible to hash-based detection measures.

These are my files and their MD5 hash values:

<u>File</u>	<u>Size</u>	<u>MD5 Hash</u>
A_Wolverine_Goes_Fishing.mpeg	300 MB	c80f4a6b2b37dddce36e9bc436b07a65
A_Wolverine_Goes_Fishing.mpeg.part1.rar	150.5 MB	fb81fd894abd17efdf1a6cae08ea1564
A_Wolverine_Goes_Fishing.mpeg.part2.rar	150.3 MB	9af068ec20f86473a2bec9ad1e37a29e

If I were to change the rar encryption key to "Rabbits2" the result would be:

<u>File</u>	<u>Size</u>	<u>MD5 Hash</u>
A_Wolverine_Goes_Fishing.mpeg	300 MB	c80f4a6b2b37dddce36e9bc436b07a65
A_Wolverine_Goes_Fishing.mpeg.part1.rar	150.5 MB	ebec7768835b3417567057a49befc326
A_Wolverine_Goes_Fishing.mpeg.part2.rar	150.3 MB	9eb0d898ee6e3c280301279b850a5a19

Same file names, identical sizes, but the rar files now have radically different hash values.

There is talk of file hosting services taking steps to look inside rar archives. File encryption makes this process impossible and so is disallowed by many services who want to remain able to inspect all files stored within their systems<sup>3</sup>. Even if systems were deployed to open archives, there is no reason to expect that the files contained within would have stable hashes. Even an imperceptible alteration to a media file can change its hash. It would only take a slight tweak of the resolution, codec, length, or even metadata to again render the file invisible to automated systems. These automated systems may have a place in detecting exact mirror images of illegal files, but they cannot do much to curtail improper sharing.



### **ERROR**

This file is suspected to contain illegal content and has been blocked. After the file has been blocked for 7 days it will automatically be deleted, if the block is not removed by RapidShare. For this reason, a download of this file is currently not possible.

### *Avoiding the Copyright Police*

The most serious threat to my uploaded files is the dreaded DMCA takedown notice. Upon receipt of such notice Rapidshare will delete my files, add their hashes to the database, and sanction my account for violating our agreed terms of service. Furthermore they could hand my information over to investigators, but that has yet to happen on the basis of a simple copyright violation<sup>4</sup>. But before any of this can occur, investigators must provide Rapidshare with a link to my files. This might seem straightforward given that I have posted my links in a public forum, but there are thousands of such locations and relatively few investigators. For this reason, investigators rely on automated systems to locate offending links. These can be dealt with.

The simplest tool for finding offending links is the basic search engine. If you Google

“Rapidshare Wolverine” you will find file sharing links to the copyrighted movie. Nearly all of them will have been detected and removed. If they appear in a Google search they are far to discoverable and, if working, may in fact link to malicious sites<sup>5</sup>. To avoid my links being discovered so easily I used a type of *link protector* called a *URL redirection service*. These services are designed to protect links from detection and reporting. The particular service I used is called Lix.in and allows me to publish working links without using words “Rapidshare” or “Wolverine”. Without these keywords, my links should remain safe from discovery via search engines.

Advanced copyright investigators often automate the entire process of discovering, testing, and reporting links. Automated *bots* crawl the web testing even protected links to see if they mask links to copyrighted material. Lix.in lets me use a *CAPTCHA* challenge to make sure that my true Rapidshare links are only divulged to flesh and blood humans. While certainly an annoyance for downloaders, this simple tool saves my links from bot-based reporting schemes.



There are countless other link protection tools. I cannot hope to explain them all. The only way to keep up is to visit various file sharing sites and see them in action for yourself.

### **Legitimate Use of File Hosting Services**

File hosting services are the most powerful way to distribute large files. There are countless non-piracy reasons for wanting to do this. File hosting services are the answer to anyone faced with the question of how to distribute a wedding video to a few hundred email addresses. One need not subscribe to a hosting service to accomplish such basic tasks. For instance Depositfiles.com allows anyone to upload and share very large files without cost, but free downloads come with a popup. Rapidshare's free upload option does not use popups and instead places limits on the number of times a file can be downloaded. These free programs will always be sales pitches for the premium packages, but service levels constantly improve as the various services fight for market share.

There are also commercial applications to consider. Many small businesses must intermittently distribute large files. Software updates are a prime example of this. File hosting services provide an alternative to maintaining costly distribution servers. Creative payment plans such as Rapidshare's

*Trafficshare* allow uploaders to cover bandwidth costs and allow downloaders to bypass the restrictions normally placed on non-subscribers. Such systems have potential as valuable alternatives to traditional distribution models.

## **Proposed Technological Countermeasures**

### *Packet Inspection*

The internet is a series of tubes, and *Deep Packet Inspection* (DPI) is a pipe dream. The automated scanning of traffic by ISPs with the goal of identifying the improper sharing of copyrighted material will never work. There is only one word to know: encryption. DPI relies on the presumption that internet traffic can be inspected, that an automated system could scan the files being transmitted and compare them to a database of known illegal and/or copyrighted files, but encryption protocols are specifically designed to close the door on such inspection.

At the moment most file sharing are not end-to-end encrypted and this gives DPI false hope. Today it seems possible to scan file transfers for known illegals, but the sharers will switch to encrypted protocols within hours of any such scheme becoming public. The p2p community has already prepared the *BitTorrent protocol encryption*. File hosting services would simply move from normal web interfaces to SSL/TLS encrypted pages. The only cost would be slightly higher CPU loads for all involved. The cost to the ISPs burdened with implementing and servicing this ineffective scheme would run into billions.

The final nail in the DPI coffin is the originality of uploads. The files I uploaded were my own creation. The movie file may have been a copyright violation, but the encrypted container I put it in is new to the world. No packet inspection regime could have detected the file within without first knowing the keyword. Therefore hash-based DPI will never be able to detect improper uploads. Absent some draconian ban on the private use of encryption, DPI should be considered a dead end for copyright enforcement.

### *Packet Shaping (Throttling)*

Rather than attempt to block illegal traffic, many ISPs have started restricting the bandwidth available for file sharing. Proponents describe this as a necessary network management tool to safeguard bandwidth for more sensitive application such as VoIP and streaming video. Detractors see an attempt to avoid the costly infrastructure improvements needed to actually expand overall network capacity. Either way, the practical impact on file sharing has and will be minimal.

Packet shaping requires an ISP to differentiate between file sharing and non-sharing traffic. While p2p sharing involves specialized protocols, traffic to and from file hosting services is indistinguishable from normal web traffic. Both use standard html protocols for file downloads and restrictions based on file size or metadata would be dramatically overinclusive and easily defeated. The only option then is to restrict traffic associated with known sharing services and websites. In practice this means restricting traffic associated with undesirable services like Rapidshare in favor of more politically-correct sharing services like YouTube. This line drawing runs afoul of any concept of neutrality and will surely trigger litigation on behalf of any aggrieved service.

### *Internet Filtering/Blocking*

ISPs sometimes block access to certain internet locations and services. In recent years this blacklisting has become the preferred option for dealing with undesirable internet content, particularly

child pornography. An effort to expand blacklisting to cover the various file hosting services would be very difficult. File hosting services are legitimate businesses. They will not take kindly to anyone putting up blocks between them and their paying customers. Even if implemented, sharers would simply subscribe to any number of proxy services to bypass the blocks and shield their activities from their local ISPs. Then the bypassing schemes are also targeted. So new bypassing techniques are invented, and blocked, until the blocking effort starts to look like China's *Great Firewall*.

Filtering can also target individual files, as opposed to entire websites. For instance Australia's infamous blacklist included specific Wikipedia entries, rather than the entire website<sup>6</sup>. A similar filter could be setup to block access to specific files within hosting services. To do this one would have to create a list of known illegal files and their links. But if one knows the location of the illegal files, why not just report their presence and have them removed? Removal by the hosting service will always be more efficient than a nationwide filtering attempt. File-specific filtering of files stored on hosting services is therefore a waste of resources.

### **New Politics**

While the p2p file sharer is largely without friends, those who use file hosting services have many. Sharing files via a p2p network means opening up one's files to be viewed and downloaded by others. Residential ISPs view this as operation of a *file server* and forbid it via term of service agreements. The mere act of sharing can be treated as a violation, regardless of copyright law. The use of file hosting services does not involve operation of a file server and so is not in itself a TOS violation. Secondly, while a p2p sharer ties up valuable bandwidth whenever he or she is sharing, those uploading to file hosting services do not. They only upload a shared file once. ISPs are therefore far less irritated sharing via hosting services than with p2p activity.

Residential ISPs might also have a darker financial motivation. Ten years ago as broadband was just taking off, there was little choice in service plans. Users got whatever bandwidth was available and everyone paid the same amount for the privilege. Today customers are offered a wide range of premium broadband packages. Higher speeds for higher prices means higher profits for ISPs. If most internet traffic is related to file sharing, and there is plenty of evidence that this is true<sup>7</sup>, one must question whether or not ISPs are in fact in bed with file sharers. If illegal sharing was somehow dissuaded, what would happen to the demand for those premium packages?

The greatest friends of the file sharer are the file hosting services. There are perhaps a dozen popular services already, and new players enter the market regularly. Any service viewed as too willing to participate in copyright investigations will loose customers to those viewed as more protective. A German court recently forced Rapidshare to hand over the IP address of a German man who had repeatedly uploaded Metallica's latest album prior to its release date<sup>8</sup>. This sent shockwaves through the file sharing community. Rapidshare moved quickly to safeguard their reputation by publishing a plain-English privacy policy<sup>9</sup>. This policy states that Rapidshare does not collect, let alone divulge, information regarding downloads. Only uploaders need worry. This has calmed the masses and sharing via Rapidshare has continued almost unabated.

### **Websites and Discussion Forums are not Torrent Trackers**

Thepiratebay.org is more than a website, it is a *torrent tracker* and as such serves two roles. It is both a place for human users to announce torrents and a needed database of peers/ The actual tracer is accessed not by humans, but by p2p client software needing to locate peers hosting particular torrents. Popularity matters when it comes to trackers. A popular tracker will index connection information for a

larger number of peers, resulting in greater torrent availability and faster sharing. Without a centralized tracker, torrent-based sharing falls apart as the p2p software can no longer find peers. This integral database of peers means that torrent trackers are privy to the IP addresses of everyone sharing torrents indexed with the tracker. If you want to know who is sharing what, trackers have that information.

The discussion forums where links are shared are absolutely different. Discussion forums are not part of the actual uploading and downloading of files. Instead of hosting a database for p2p software to exchange data, discussion forums are places where real people meet to exchange links with other real people. Normally they take the form of websites, but links and passwords can be distributed via any medium whatsoever, even legal journals. Where one acquires links is irrelevant, as is the popularity of any given forum. Discussion forums are also not always privy to IP addresses. The poster of a link to a forum is often not the actual uploader of the file. Many persons attempt to take false credit for important uploads by reposting the links of others. Others post links on behalf of uploaders who want to remain anonymous. The most careful sharers use proxies to mask their true IP addresses from forum administrators. Ignorance can be bliss when the copyright police come knocking. Discussion forums may be dens for illegal file sharers, but that does not mean that they are of any import to the sharing process.

An attack on a discussion forum would do more harm than good. The shutting down of a discussion forum will not remove offending files from a hosting services, nor will it provide reliable evidence regarding uploaders. A determined attack could instead push the forum deeper into the realms of anonymity. The bandwidth needs of discussion forums are miniscule. Take my post as an example. Between the links and the password it comes in at just over one hundred characters. Thousands of such posts could be stored on a site buried within the TOR anonymity network, a place beyond the reach of the DMCA<sup>10</sup>. Discussion forums are also used by copyright investigators looking to discover illegal links. The more well hidden the discussion forum, the more difficult it will be to report illegal links. The smart copyright owner might want to take a hands-off approach and instead view discussion forums as surveillance tools.

## **Conclusions**

Do not listen to anyone pitching a product, service or legal strategy purporting to eliminate file sharing. The sharing of files via hosting services is far more complex than peer-to-peer networking, and both evolve constantly. The next steps are already being taken. Proxy schemes are working protect uploaders, encryption protocols are masking files from inspection, and the dark market of anonymous payment schemes allow sharers to avoid leaving paper trails. The file hosting services are also protecting themselves through careful choice of jurisdiction<sup>11</sup>. The legal community needs to move on from the comfortable peer-to-peer debate and come to grips with the new reality of sharing via file hosting services.

-Richard Abbott  
[Rabbit@shaw.ca](mailto:Rabbit@shaw.ca)  
[Oregonrabbit@hushmail.com](mailto:Oregonrabbit@hushmail.com)

- 1 The file is in fact a TrueCrypt volume. The key is “Rabbits”. Call it the first easter egg ever hidden in a legal journal. The first five people to crack the egg will win a prize.
- 2 <http://arstechnica.com/old/content/2008/09/p2p-growth-slowing-as-infringement-goes-deeper-undercover.ars>
- 3 [http://www.mediafire.com/acceptable\\_use\\_policy.php](http://www.mediafire.com/acceptable_use_policy.php)
- 4 <http://torrentfreak.com/rapidshare-shares-uploader-info-with-rights-holders-090425/>
- 5 Identity thieves often use sharing-related links promising movies/music to lure unsuspecting users. Always be careful when clicking on links. Just because it says “Rapidshare.com” doesn't mean that is where it will take you.
- 6 I would have added a link to the leaked Australian blacklist, but re-publication of that list might be a crime. You will have to go out and find it yourself. I am sure there is some WIKI-type site collecting such LEAKS.
- 7 [http://www.ipoque.com/resources/internet-studies/internet-study-2008\\_2009](http://www.ipoque.com/resources/internet-studies/internet-study-2008_2009)
- 8 <http://arstechnica.com/tech-policy/news/2009/04/rapidshare-hands-over-user-info-in-germany-users-panic.ars>
- 9 <http://rapidshare.com/privacypolicy.html>
- 10 The Onion Router (TOR) is an anonymous communication layer that can mask the true location of websites. Called *hidden services*, these websites cannot be tracked to a particular ISP or country.
- 11 Megaupload.com is based in Hong Kong, but as recently as 2009 has blocked access via Hong Kong IP addresses. It is speculated that this decision was made to remove deny legal standing to local media companies.