Copyright Protection in the Internet

White Paper

During the last years, file sharing of copyright-protected material, particularly in peer-to-peer (P2P) networks, has been a serious threat to the established business models of the content industry. There have been numerous discussions about possible countermeasures, some of which have already been implemented. This white paper aims to provide an as objective as possible assessment of the countermeasures for P2P from the perspective of a network device vendor with particular experience with Internet traffic management solutions.

Authors: Klaus Mochalski, Hendrik Schulze, Frank Stummer

Approach

Music, movie and software companies as well as legacy publishing houses claim billion-Euro revenue losses that have driven them to job cuts. The widespread illegitimate sharing of copyright-protected material thus has a negative economical impact both on a national and international scale. There are many, often contradictory statements about feasibility and effectiveness of countermeasures for P2P file sharing. Judgments are often driven by the interests of different groups, such as industry lobbyist and privacy activists, and their recommendations differ widely.

In this white paper the focus lies on technical solutions. The countermeasures can be classified into three categories:

1. Prevention of file transfers at application level, irrespective of content of single files

2. Detection of copyrighted and non-copyrighted material, prevention of file transfers at single file level or prosecution of infringers

3. Non-technical approaches

While some proposed measures are simply unfeasible, others could be implemented both from a technical and commercial perspective. Some of the measures can be combined. However, we will look at each of them individually and subject them to a reality check, that immediately rules out some approaches. Non-technical (political, legal and economic) solutions are regarded in the third part but not evaluated. The two feasible measures will be evaluated, based on the following criteria:

Technical feasibility:

Can the measure be implemented technically? What infrastructure is necessary? Is it already being used or is it still under development?

Effectiveness:

How effective and comprehensive is the measure?

Costs and time of implementation:

How expensive is the first installation? Who would pay or how can costs be shared? How long would the implementation take?

Operation and maintenance costs:

What maintenance does the measure require? What costs does it involve and who bears them?

Impact on Internet users:

Is there a privacy or data protection impact? Will it limit service availability?

Impact on Internet service providers:

What infrastructure changes are necessary? Does the measure involve additional workload when offering Internet service to customers? Do business models have to be changed?

Impact on content providers:

Which input is necessary from content providers? Is it possible to protect all titles? Do business models have to be changed?

Impact for governments and society:

Is a change in legislation necessary? Does political consensus appear to be achievable? Would the measure be politically unpopular? How severe would lobby groups react?

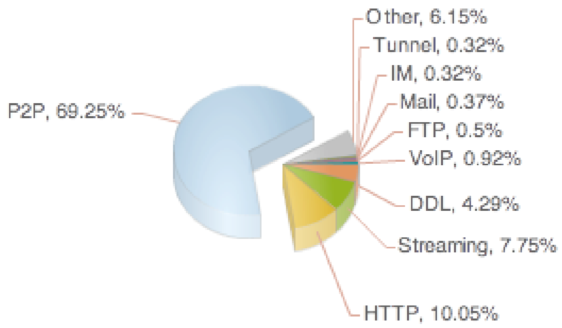
Assumptions

Many transfer methods can be used to share legal and illegal content. An optimal countermeasure should be able to solve the problem of copyright infringements effectively, efficiently and as comprehensive as possible. Although a 100% solution is generally impossible, most of the current methods for illegitimate content sharing (e.g. P2P networks, file hosting services, streaming services) should be covered. The legitimate use of all these services has to be taken into account.

Other distribution systems used for more serious criminal activities (e.g. pedophilia, terrorism, organized crime), such as closed private networks or the exchange of storage media via e-mail, are difficult to control with the proposed measures. They require criminal investigation methods, which means a higher effort but also a higher effectiveness.

Nearly all measures incur a more or less serious interference with Internet traffic and thus a violation of net neutrality. This paper will not discuss this issue because there are no clear and generally agreed definitions of net neutrality. While this discussion is important, it would go beyond the scope of this paper.

P2P file sharing is producing more traffic in the Internet than all other applications combined as shown in this figure taken from ipoque's Internet Study 2007.



Protocol distribution in Germany

1 Host- and Application-Level Traffic Management

Methods to prevent undesired file transfers on host and application level require manipulation of Internet traffic, thus either extension of existing or installation of new devices.

1.1 Blocking of IP Addresses/DNS Names

Access to individual Internet hosts is blocked based on IP address or host name. On one side, this allows to block access to central servers that offer copyright-protected material, such as file hosting services, for instance. On the other side, also individual subscribers can be blocked temporarily or permanently, if they are known infringers. In both cases, blocking needs to be implemented close to the network edge, for instance at the DSLAM (DSL termination) or CMTS (cable termination) level.

Reality Check

Dynamic IP addresses and network address translation (NAT) restrict technical feasibility. The implementation is possible at ISP level, where dynamic address mapping information is - at least in theory - available.

Conclusion

It is theoretically possible but infringers would have to be detected first (See category 2 for such measures). Blocking of IP addresses could be an additional measure in a combination of different measures, but is not the solvation of the problem itself.

1.2 Blocking of URLs

Access to URLs hosted by infringing Web servers can be blocked. This requires URL filters with a database of URLs that are to be blocked.

Reality Check

URL filters are widely available. Centrally hosted services such as Piratebay and even BitTorrent trackers could be blocked. An up-to-date list of URLs is a necessary prerequisite to make this measure effective. Unfortunately, it is nearly impossible to keep the URL database current. Affected sites could rapidly change URLs and propagate these changes. Ultimately, this would result in a never-ending cat and mouse game.

P2P networks with no central servers are totally off limits for this approach.

Conclusion

Blocking of URLs only makes sense as a complementary measure in specific cases such as web sites with criminal content.

1.3 Blocking of TCP and UDP Ports

All ports used by infringing applications will be blocked.

Reality Check

Port blocking systems are widely available, but many applications do not adhere to standards with regard to port usage anymore to avoid being detected by these systems.

Conclusion

This is an old idea that does not work at all anymore.

1.4 Black- and Whitelisting of Protocols and Applications

Block (blacklisting) and allow (whitelisting) specific applications based on layer-7 classification. Candidates for the blacklist would be applications that are mostly used for illegal content distribution who are trying to avoid monitoring by using encryption. Candidates for the whitelist would be applications or services that cannot be used for illegal content distribution or are monitored by other means.

Reality Check

Black- and whitelisting is technically feasible with current deep packet inspection (DPI) and behavioral analysis (BA) technology, which allows the reliable classification of protocols even if they are using encryption or obfuscation. However, blacklisting and – even more so – whitelisting would seriously impede innovation in the Internet, because usage of new protocols would be severely restricted. P2P networks are a good example. While the overwhelming proportion of exchanged content violates copyrights, the networks themselves are important new technologies. Not only copyright infringers use P2P but also scientists share their research data this way.

For a really comprehensive approach, a large-scale coordination at least at national, better transnational, level would be required. All ISPs would have to cooperate. Politically, this is hardly imaginable in most nations, not only in light of the current net neutrality debate. It would mean a return to the very roots of the Internet, only with strictly enforced standards. Ultimately, this measure would bring the Internet to a grinding halt.

An example is ipoque's BitTorrent tracker whitelisting, that allows access to guaranteed legal BitTorrent content, while blocking access to all other P2P content. This approach works because nearly all legal P2P content is distributed over BitTorrent using dedicated and controlled BitTorrent trackers. Operators can deploy such systems to limit access to and from their networks to improve security and prevent illegal file sharing activities. This can be important because, depending on national jurisdiction, network operators may be liable to their users' activities or are at least required to cooperate with legal authorities and private infringement monitoring firms.

Conclusion

Black- and whitelisting of applications cannot be deployed throughout the Internet. However, it is a good option for individual network operators, such as companies or universities, to control their network usage. Traffic management systems with both whitelisting and blacklisting for various applications and even content-awareness are available on the market.

1.5 Injection of Counterfeits

In P2P networks, each user also is a content provider. This makes it easy to inject counterfeited copies of files into the network. Poor hashing algorithms in the eDonkey P2P network, for a long time had allowed to offer same-size files with different content but equal hash values. Today, this is no longer possible due to new hashing algorithms.

Reality Check

Current file sharing networks use strong protection mechanisms against the injection of counterfeited copies. It was common practice to inject fake files (i.e. files with misleading names) into file sharing networks – and to a lesser degree still is today. This has significantly decreased the content quality particularly in the eDonkey network. As the main effect, these measures have driven file sharers to the BitTorrent network, that is nearly immune against injection of fake files, mainly because content distribution is organized through Web-based torrent directories such as thepiratebay.org.

Conclusion
The injection of counterfeits is no effective countermeasure anymore.

1.6 Exploitation of Vulnerabilities in File Sharing Software

Attacks on file sharing networks using implementation and protocol vulnerabilities to derail their proper function.

Reality Check

As for any computer system, attacks are possible, and there are commercial providers offering this as a service. An attack on eDonkey, for instance, may have the effect that the downloaded file is larger than the original, and the download never finishes. There are similar attacks for BitTorrent.

Conclusion
As this kind of attack is based on vulnerabilities, the common risk with such measures is that the vulnerabilities will be fixed. The short-lived effectiveness makes the measures expensive and only worthwhile for a limited number of highly valuable files (e.g. newly released movies and computer games, expensive software).

2 File-Level Traffic Management

The following methods are able to detect a single file and find copyright protected content in it. Once they detect copyright protected content, there are two possible actions: either to prevent the file transfer directly; or to prosecute infringers afterwards. In the first case, the exchanged file's hash value (or another kind of fingerprint) is compared to a database or classification system. Based on the verdict, the file transfer is either blocked or allowed. Both blacklisting and whitelisting are possible to block or allow file transfers. In the second case, the file transfer is allowed, but recorded for later prosecution.

2.1 Fingerprinting

Fingerprinting is a method that uses parts of a file to generate a fingerprint. A transferred file's fingerprint can then be compared to a database of reference files and classified with a high reliability.

Reality Check

Although fingerprinting is already used commercially, the technology is still under development. Certain parts of it have been used for a long time, for instance the detection of patterns in pictures. There are no independent benchmarks for its accuracy and reliability.

The major advantage of fingerprinting over other technologies is that modified copies of an original file are still detected as copies of it, independent of the modifications. Due to its computational complexity, fingerprinting does not work in real-time for high-speed networks. Also, even though ever more file and compression formats are supported, fingerprinting is blind to encrypted archive files (e.g. password-protected ZIP files), and these are becoming more and more popular. Large-scale deployment of fingerprinting technology would push the popularity of all kinds of encryption and render the whole technology useless as a countermeasure.

Conclusion
Fingerprinting systems do not operate in real-time and cannot be deployed on a large scale. They do not work with encrypted communication or encrypted files. Still, fingerprinting can be useful for an offline search of particular files and the involved IP addresses in captured network traffic. In addition, its ability to correlate modified copies with the original file can be used to build a more complete database of file hashes, which can then be used with file hash-based measures as described below.

2.2 File Hash-Based Identification and Blacklisting

Each file in file sharing networks has a unique ID. In P2P networks, this is the file hash, and in file hosting systems, this is the URL. For each title, there often exists a number of files with different hash values, e.g. modified copies or different file formats. This is caused by different users offering the same title. Consequently, the number of file hashes is significantly higher than the number of unique titles circulating in file sharing networks. In practice, however, only a limited number of such copies are propagated throughout the P2P network. The common ratio between a title and its copies usually is about 1:3-6.

Reality Check

Traffic managers are capable to maintain file hash databases with at least one million entries and to selectively block or allow individual file transfers. While both black- and whitelisting are in theory possible, only blacklisting would be politically viable. Whitelisting, i.e., the controlled admission of validated files only, would be a serious infringement of freedom of speech making it all but politically impossible.

File hash-based measures do work effectively with unencrypted and public sharing services. Encrypted communication and private file sharing networks can only be controlled by criminalistic methods involving a high effort. However, the vast majority of copyright infringements happen in open services, as the public availability is the key success factor for such services.

Technical feasibility:
This measure can be implemented using currently available traffic management systems based on deep packet inspection deployed at network access or peering points. A central management system would control their operation and also maintain the file hash or fingerprint database for files to be detected and blocked.

Effectiveness:
This measure can be effective for a whole country or even larger regions, if all relevant access or peering points are covered. Blacklists have to be continually maintained.

Costs and time of implementation:
An installation at peering points would cost approximately 1.50 Euros per network user, and an installation at the access points about ten times this amount. Costs could be borne by the ISPs (who would also benefit from additional functionalities provided by the traffic management systems), or by the government and thus by the taxpayer. Implementation should be feasible in less than one year.

Operation and maintenance costs:
In addition to the maintenance of the traffic management systems (e.g. firmware updates, hardware maintenance), regular distribution of file hash databases is necessary, for instance every 48 hours. This database has to be maintained at a central entity. Based on experience, a single person is able to manually track 1,000 titles. Assuming that 10,000 English titles have to be monitored, 10 staff would be able to do the job. Other language regions would require about 1-5 staff. The hash databases can be used internationally. One worldwide copyright agency could be operated by the content industry, for example. Annual costs would be less than 1 Euro per user.

Impact on Internet users:
There is no noticeable change for Internet users apart from copyright-protected files not being downloadable anymore. There are no privacy or data protection issues, as no subscriber IP addresses or any other personal details are being tracked.

Impact on ISPs:
ISPs have to deploy traffic management systems at a sufficient number of access or peering points. The devices can be implemented with bypasses, which secure uninterrupted network connectivity. There is no change in the services offered and business models in general.

Impact on content providers:
Content providers can – and have to – provide a list of titles they want protected. Depending on the level of service provided through the discussed central agency, they could either provide the names of titles or a list of relevant hash values. Because the list of titles that can be monitored is finite, it needs to be continually updated and cleaned from outdated entries. It should focus mainly on popular and current titles. One possible model would be that content providers pay a per-title fee to the monitoring agency. Current business models could be effectively protected with this measure.

Impact on governments and society:
The implementation of a nationwide (or even international) protection system requires sufficient and enforceable rules – either through legal or industry regulations. The implementation of such a system is new territory for most countries and would certainly trigger fierce debates involving the content industry, privacy and data protectionists, and consumer protection groups. In several countries there are ongoing discussions about this countermeasure.

Conclusion
Blacklisting based on file hashes or other file IDs can provide a viable way to severely limit the distribution of copyright-protected content.

2.3 Signing of Transmitted Content with a Legally Binding Digital Signature

Each shared file needs to be digitally signed through personal certificates. With this signature, a person or legal entity sharing a file certifies the ownership and consent for sharing of the file. In case of misuse, the sharer can be easily prosecuted. All files without a valid signature would be blocked.

Reality Check

Different kinds of signing technologies are available and in use for several applications. Due to the fact, that they do not have a legal or central entity it is nearly impossible to enforce it for public file sharing networks. Also, file hosting services can easily avoid a prosecution by moving to another country.

Conclusion
A modification of file sharing networks with signing techniques would very effectively solve the problem. It is a good solution for several applications. But the implementation is not possible for widespread, public file sharing networks or services.

2.4 Watermarking and Investigation of Seeders

It is technically possible to armor each title with a digital watermark that would allow tracking its way through the transmission chain. The aim is to prosecute infringers of copyrights. For example this method is already in use to determine the movie theater where a movie was filmed off the screen and then put into a file sharing network.

Reality Check

Conclusion
It appears very likely that watermarks would be erased or destroyed if used on a large scale, resulting in yet another cat and mouse game that cannot be won by the copyright owners.

Making watermarking an effective measure requires full control over the entire production and distribution chain up to the screen and speaker. For CDs and DVDs, using current technology, this is not possible.

2.5 Monitoring of Copyright Infringements

Active and Passive Monitoring

In case of active monitoring a monitor participates in the P2P network as an active client and tries to download copyright-protected files. It can only find files it is explicitly looking for. Files to be monitored are usually provided by the copyright owners. Only data of persons sharing protected files are collected. No other traffic is monitored.

Passive monitoring inspects the complete Internet traffic, ignoring all uninteresting traffic and looking only for exchanges of copyrighted titles. It causes severe privacy and data protection concerns as it has, potentially, access to all data, including e-mails, web traffic, etc. The two methods – active and passive monitoring – are totally disparate technologies.

Clearing Instance

Prerequisite for monitoring of file sharing networks is a central instance that provides the mapping between the IP address (along with the time of its recording during the monitoring process) to the personal identity behind this address. The success of this measure depends on the clearing instance.

In Germany, for instance, the mapping could only be acquired through an official prosecution process. The flood of prosecutions has overwhelmed state attorneys. France, in contrast, has chosen to implement a much simpler and very promising process. The government has decided to implement a clearing instance, called HADOPI, and all parties are required to cooperate with it.

Besides copyright violations in P2P networks, the clearing instance can also provide personal data in case of other infringements, such as libels, agitation or similar. The mapping from IP addresses to personal data requires to store connection records over a certain period of time, which is a controversy in some countries.

Active Monitoring

Active monitoring only works for P2P networks. Automated clients try to download copyright-protected files from these networks, or also offer them for download by others. It is difficult for P2P users to detect these clients because they work just as ordinary P2P programs. Clients can be modified so that they only download, but do not upload, any files to avoid spreading copyrighted material.

Active monitoring can be conducted from virtually anywhere in the world, covering P2P networks independent of the infringer's location. It also works just as good for encrypted P2P networks because the monitor participates as an ordinary peer.

Reality Check

Active monitoring has garnered a bad reputation because content providers have in the past often tried to criminalize copyright infringers and imposed ridiculous penalties as a deterrent. In addition, there have been flawed lawsuits with verdicts about persons with no Internet access. Careful investigation along with adequate penalties are necessary to improve the reputation of this measure, even more so as it bears the potential to solve the copyright problem in P2P networks.

Technical feasibility:

Different systems are available and have been in operation for some years. Different P2P networks, among them the most popular services, are covered.

Effectiveness:

Such systems can detect infringements for one or more countries – nationally and internationally. The location is not important. Especially automatic detection systems work highly efficiently and produce court-proof evidence data. This measure is very difficult to circumvent.

Costs and time of implementation:

A single appliance could cover 2,000 to 10,000 titles and would cost 1 to 10 Euros per title. Costs could be covered by a fee from the industry, subsidies from the government or the penalties from infringers. A high organizational effort is necessary to implement the processes to prosecute infringers. This could be done by the copyright holders, a lobby organization, or a state authority.

Operation and maintenance costs:

Depending on the requirements, detection of infringers costs 1 to 10 Euros per title and year. A per-case fee would be possible, too. The cost of prosecution depends on the legal requirements and organizational set-up, but could be very efficiently done with automatic processes. The content providers or a lobby organization would have to feed the system by providing the titles or the shared files for these titles. Based on experience, a single person is able to manually track 1,000 titles.

Impact on Internet users:

Only infringers are detected and tracked, other users are not monitored by the system. Privacy and data protection issues have to be clarified by law or regulations.

Impact on ISPs:

There is no impact on the network infrastructure of ISPs, as the appliances can be located anywhere. ISPs play an important role in matching the detected IP addresses to personal names and addresses. They could possibly involved even more in the prosecution processes, e.g. by supporting an automatic infringement notification system or by executing penalties (e.g. disabling of Internet access).

Impact on content providers:

Content providers can – and have to – provide a list of titles or files that they want protected.

Impact on governments and society:

Active monitoring systems are possible and in operations in several countries. In other countries discussions about such systems and the necessary laws are ongoing.

Conclusion

Active monitoring is already in operation in several countries for some years and will be implemented in other countries, too. However it is a prosecution of infringements and is therefore a (powerful) answer to the problem, not a full solution.

Passive Monitoring

Passive monitoring uses network probes installed at appropriate network locations to investigate Internet data flows. Technically, these probes could utilize the same infrastructure deployed in many countries for lawful interception purposes (i.e. the interception of telecommunications by law enforcement agencies and intelligence services in accordance with local law). As a major advantage, monitoring is not limited to P2P, but all communication can be scrutinized for copyright infringements – with the exception of encrypted traffic.

Reality Check

Passive monitoring is technically possible, but implies monitoring of every network user’s traffic, treating everyone as a potential suspect.

Conclusion

This approach is politically unfeasible in most countries.

2.6 Penalization of Copyright Infringements

The investigation of copyright infringements resulting in penalties is a measure that is already – or will be soon

– implemented in some countries. It is only used for P2P networks, and it is limited to the most popular networks such as eDonkey and BitTorrent. Depending on local legislation, some methods only investigate uploaders while others look for downloaders too. In the investigation process, profiles are created that comprise the kind and number of infringing titles to avoid prosecution of petty crimes. The penalty can be a cease and desist order along with a payment (as in Germany, for instance), or it can be the deactivation of the infringer's Internet access, usually after a number of warning messages (as planned in France and Great Britain).

3 Non-technical Solutions

Besides the technical methods discussed above, there are many different non-technical approaches to solve the problem of copyright infringements. The following approaches are the most important of the non-technical solutions. They are widely discussed.

3.1 Culture Flat Rate

The so-called "culture flat rate" is a model where each consumer pays a monthly flat fee for content usage – similar to the public service broadcasting as in the UK or in Germany. This flat rate would allow every paying customer to legally download arbitrary content from the Internet. A clearing entity would take care of distributing the collected fees to copyright owners. This measure would give legal access to huge music, movie and other content collections for everyone.

3.2 Digital Rights Management (DRM)

DRM with its potential to control the content distribution chain has been the content industry's preferred solution for a long time. The basic idea is quite simple: all copyright-protected titles are encrypted or armored with a signature and can only be played back – and not copied – on certified devices that obey copyright laws.

Currently, content providers argue for a relaunch of DRM methods. Should the implementation be successful, this measure has the potential to prevent the opportunistic distribution of copyrighted material. In the past, any DRM mechanism was hacked or otherwise circumvented. This is highly likely to happen to new systems as well. A simple example of an almost uncontrollable situation is the acoustic recording of music, which delivers perfectly acceptable audio quality for most users.

3.3 Improved Offerings and Pricing Models

It has often been argued that the losses the content providers have suffered during the last years are not only due to Internet file sharing. Another reason for them could be the missing alternative offers on the Internet or, rather, offers with a bad price-content quality ratio.

However, the industry has developed new and better offers during the last years. Sales figures for online content have increased dramatically. Researchers found that many infringers would buy or in fact bought the legal titles if they could find a decent offer – in price, quality and accessibility.

4 Summary and Conclusion

First, and most importantly, content providers need to provide other high-quality, well priced and easily accessible online content. New business models are inevitable. In the long run, this will make illegitimate sharing of copyright-protected material through the Internet a lot less interesting. Until then, two of the discussed countermeasures promise to be the most effective and viable ones: hash-based detection of copyrighted files and the prevention of their transfer in the network; and the active monitoring combined with the prosecution of infringers. For institutional network operators (e.g. universities, companies), traffic management solutions with whitelisting of desired applications and content is also possible, but this is no option for national or international deployments.

About ipoque

ipoque is the leading European provider of deep packet inspection (DPI) solutions for Internet traffic management and analysis. Designed for Internet service providers, enterprises and educational institutions, ipoque's PRX Traffic Manager allows to effectively monitor, shape and optimize network applications. These include the most critical and hard-to-detect protocols used for peer-to-peer file sharing (P2P), instant messaging (IM), Voice over IP (VoIP), tunneling and media streaming, but also many legacy applications. For further information see www.ipoque.com.