

# Suche nach Urheberrechtsverletzungen in Internet-Tauschbörsen mittels digitaler Wasserzeichen

Konstantin Diener, Patrick Wolf, Martin Steinebach, Hans-Peter Wiedling

Fraunhofer IPSI  
Dolivostraße 15, 64293 Darmstadt  
kdiener@konstantin-diener.de  
patrick.wolf@ipsi.fraunhofer.de  
martin.steinebach@ipsi.fraunhofer.de

Hochschule Darmstadt  
h.wiedling@fbi.h-da.de

**Abstract:** Begünstigt durch die mittlerweile flächendeckend verfügbaren Breitbandanschlüsse wird das Internet verstärkt zur unberechtigten Veröffentlichung und Verbreitung digitaler Medien genutzt. Als Plattform hierzu dienen vor allem Internet-Tauschbörsen. Wir stellen einen zweistufigen Suchmechanismus vor, um diese unberechtigten Veröffentlichungen aufzudecken. Die erste Stufe stellt eine Metasuchmaschine dar, die potentiell unberechtigt veröffentlichte Medien anhand klassischer Kriterien wie Schlagwörtern in beliebig vielen Internet-Tauschbörsen sucht. Die eigentliche Aufdeckung der Rechteverletzungen findet anhand digitaler Wasserzeichen statt. Dies sind Informationen, die den Schöpfer des Werkes oder den Käufer eines digitalen Mediums identifizieren und direkt in die Medien eingebettet wurden. Diese Informationen werden in der zweiten Stufe des Suchvorgangs aus den im Internet gefundenen Medien ermittelt.

## 1 Motivation

In digitaler Form gespeicherte Medien sind heutzutage allgegenwärtig und haben ihre analogen Vorgänger weitgehend abgelöst. Zu dieser Verbreitung hat beigetragen, dass sich digitale Medien im Gegensatz zu analogen beliebig oft ohne Qualitätsverlust reproduzieren lassen. Eine zweite Eigenschaft dieser Medien ist es, nicht mehr zwingend an ein physikalisches Trägermedium gebunden zu sein. Dies ermöglicht ihre Übertragung über Datennetze wie das Internet. War diese Möglichkeit in den Anfangstagen des Internets noch eher von theoretischer Natur, werden heutzutage riesige Datenmengen in Internet-Tauschbörsen veröffentlicht [B06] (auch wenn es sich dabei vielleicht nicht um die größte Quelle für Inhalte handelt [BEPW03]). Dabei werden häufig rechtliche Rahmenbedingungen, insbesondere das Urheberrecht, ignoriert.

Von den Rechteinhabern werden daher verschiedene Strategien verfolgt, um dieses Phänomen zu erschweren oder sogar zu verhindern [SDL03]. Eine Methode hierzu basiert auf digitalen Wasserzeichen. Mit ihnen kann eine illegal vervielfältigte Mediendatei auf den ursprünglichen Kunden und somit zur Quelle der Vervielfältigung zurückverfolgt werden. Da digitale Wasserzeichen passiv sind, müssen sie allerdings von einer aktiven Komponente unterstützt werden: Markierte Medien müssen durch eine geeignete Suche im Internet beziehungsweise in Internet-Tauschbörsen gefunden werden. Aufgrund der großen, durch Internet-Tauschbörsen erreichbaren Datenmengen stellt die Suche in diesen Veröffentlichungsplattformen die zentrale Herausforderung dar [SW06].

Nach der einleitenden Beschreibung der grundlegenden Technologien werden die Metasuche in Internet-Tauschbörsen und Ermittlung der Wasserzeicheninformationen beschrieben.

## **2 Grundlagen**

In diesem Abschnitt stellen wir kurz die wesentlichen Technologien vor, die der Metasuche zu Grunde liegen.

### **2.1 Digitale Wasserzeichen**

Digitale Wasserzeichen sind eine verhältnismäßig neue Technologie, die in den letzten Jahren im Schutz von Multimedia-Daten Einzug gehalten hat [CMB02]. Durch sie können Informationen untrennbar mit Medien verknüpft werden. Wir verstehen unter einem digitalen Wasserzeichen ein transparentes, nicht wahrnehmbares Muster, welches in das Datenmaterial, wie beispielsweise Bild, Video, Audio oder 3D-Modelle mit einem Einbettungsalgorithmus unter Verwendung eines geheimen Schlüssels eingebracht wird [D00]. Zu jedem Einbettungsalgorithmus existiert auch ein Detektions- bzw. Abfragealgorithmus, der das Wasserzeichen wieder aus dem markierten Datenmaterial auslesen kann, wenn der geheime Schlüssel vorliegt. Wasserzeichen können so beschaffen sein, dass die eingebettete Information auch nach Veränderungen am Trägermedium wie etwa verlustbehaftete Kompression oder Skalierung noch erhalten bleibt. Solche Wasserzeichen werden als robust bezeichnet.

Urheberrechtsverletzungen bei digitalen Medien sind deswegen schwer aufzuklären, da Kopien digitaler Güter perfekt (also ununterscheidbar) sind und sich sehr leicht verbreiten lassen. Die Quelle einer unberechtigten Veröffentlichung lässt sich somit nicht ermitteln. Mit Hilfe von digitalen Wasserzeichen können von einem Original unterscheidbare Kopien erzeugt werden. Dies wird auch aktives Fingerprinting genannt. Dabei wird etwa beim Kauf eines digitalen Mediums in einem Onlineshop, die (eindeutige) Kundennummer des Käufers robust in das Medium eingebracht. Wird das Medium im Internet veröffentlicht, kann der ursprüngliche Kunde an Hand der Kundennummer identifiziert werden.

Digitale Wasserzeichen sind allerdings passive Schutzmechanismen. Durch sie kann daher kein aktiver Schutz erfolgen, der z.B. den Missbrauch von Medien verhindert. Wird dies notwendig, müssen digitale Wasserzeichen durch weitere Mechanismen ergänzt werden, die anhand der Wasserzeichen entscheiden und reagieren können. Die in den Abschnitten 3 und 4 vorgestellte Metasuche stellt einen solchen aktiven Mechanismus dar.

## 2.2 Peer-to-Peer Tauschbörsen

Peer-to Peer (kurz P2P)-Netzwerke versuchen, Ressourcen zu nutzen, die auf den Peer-Rechnern im Internet zur Verfügung stehen. Diese Ressourcen können Plattenspeicher, Rechenleistung, Inhalte usw. sein. Zu diesem Zweck versuchen sie, diese Peers sowohl als Clients als auch als Server zu nutzen, so dass alle Rechner im P2P-Netzwerk gleichberechtigt in ihrer Funktion sind. Dabei haben P2P-Netzwerke das Problem, die Fluktuation der Peers, dynamische IP-Adressen, Firewalls, NAT und Menge der Peers zu bewältigen. Zum heutigen Zeitpunkt beschäftigen sich die meisten P2P-Netzwerke mit dem so genannten Filesharing, d.h. dem Dateiaustausch zwischen zwei Rechnern.

Wo notwendig, dient in dieser Arbeit Gnutella als konkretes Beispiel für ein solches Netzwerk. Das Gnutella-Netz gilt als das erste große, komplett dezentralisierte P2P-System, das mit Hilfe des Internets funktioniert. Es verbindet viele bekannte Technologien, wie TCP/IP, CDMA und das verlustbehaftete Routen von Nachrichten. Das Gnutella-Protokoll ist mittlerweile frei verfügbar. Jede Applikation, die dieses Protokoll implementiert, ist kompatibel zum Gnutella-Netzwerk.

Das Netzwerk baut sich aus sogenannten Nodes auf, deren Anzahl sich dynamisch verändern kann. Auf diese Weise ist das Netzwerk einer permanenten Veränderung unterworfen. Um sich als Node im Gnutella-Netzwerk zu beteiligen, stehen dem Benutzer verschiedene Clientprogramme zur Verfügung, beispielsweise Bearshare oder Limewire [Limewire]. Es findet keine eindeutige oder dauerhafte Identifikation eines Nodes innerhalb des Netzwerkes statt, was einen gewissen Grad von Anonymität verspricht und somit die Aufdeckung von Verstößen zusätzlich erschwert.

## 3 Metasuche in Internet-Tauschbörsen

Internet-Tauschbörsen dienen in unserem Kontext als Peer-to-Peer-Plattform zur dezentralen Bereitstellung digitaler Medien. Zur Aufdeckung von Urheberrechtsverletzungen gilt es zunächst, *alle möglichen* Tauschbörsen nach unberechtigten Veröffentlichungen zu durchsuchen. Eine solche Metasuche muss in der Lage sein, allgemeine Suchanfragen nach Medien entgegen zu nehmen und sie spezielle Anfragen der angeschlossenen Tauschbörsen zu übersetzen. Herunter geladenes Material kann dann in einem zweiten Schritt nach Wasserzeichen durchsucht werden (siehe Abschnitt 4).

Es existieren derzeit eine Reihe verschiedener Tauschbörsen, die sich zudem in einem steten Wandel befinden – ältere Vertreter wie Napster oder KaZaA werden durch neuere Ansätze wie Torrent-Netzwerke abgelöst. Die einzelnen Tauschbörsen unterscheiden sich zum Teil grundlegend in der zugrunde liegenden Organisationsphilosophie. Das für die Metasuche interessante Prinzip der Suchschnittstelle ist allerdings bei allen nahezu identisch: Es handelt sich um eine auf (textuellen) Metadaten basierende Suche. Wie Abbildung 1 zeigt, besteht die Eingabe der Suchschnittstelle aus einem Satz von Metadaten, welche die gesuchten Medien beschreiben. Als Ausgabe liefert die Suchschnittstelle eine Menge von Suchergebnissen, von denen jedes aus (weiteren) Metadaten und einer Zugriffsmöglichkeit auf den Dateinhalt beziehungsweise die Medienessenz besteht.

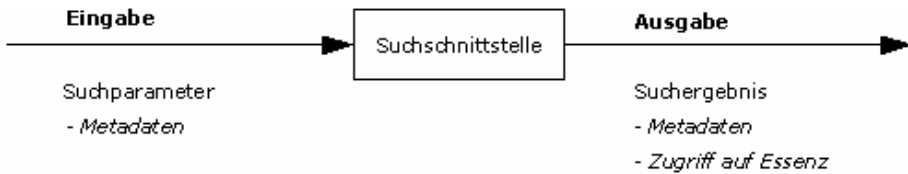


Abbildung 1: Suchschnittstelle von Internet-Tauschbörsen

Die Metasuche in Internet-Tauschbörsen funktioniert nach demselben Prinzip wie die Suchschnittstellen der gängigen Internet-Tauschbörsen. Im Folgenden soll ein einheitlicher Satz von Metadaten sowohl für die Formulierung einer Suchanfrage als auch für die Beschreibung eines Suchergebnisses definiert werden. Dabei wird eine möglichst kleine, aber trotzdem aussagekräftige Menge gewählt, damit die Abdeckung aller vorhandenen Suchschnittstellen sichergestellt ist:

- **Suchanfragen** unter zu Hilfenahme von: Schlagwörter, Medientyp, Mediengröße und deren Verknüpfung mittels logischer Operatoren
- **Suchergebnis** enthält: Adresse des bereitstellenden Netzknotens, Dateiname, Dateigröße, Medientyp

Es ist zu bemerken, dass sich die Suchschnittstellen der verschiedenen Internet-Tauschbörsen im Umfang der Metadaten unterscheiden, die als Suchparameter akzeptiert werden und derer, die in Form eines Suchergebnisses geliefert werden. Gleiches gilt auch für die Art des Zugriffs auf Metadaten und Medienessenz. Das heißt, dass für jede spezielle Tauschbörse eine dedizierte Suchkomponente existieren muss, an die der eigentliche Suchvorgang delegiert wird (vgl. Strategiemuster [GHJV96]). Dazu müssen sie sich nur noch bei der zentralen Metasuchschnittstelle, die die allgemeinen Suchanfragen entgegen nimmt, registrieren. Dies kann etwa im Rahmen eines größeren Frameworks geschehen (vgl. [D06]). Für das dort beschriebene Framework haben wir eine Suchkomponente implementiert, die Medien in Gnutella-Netzwerken auf Basis der Limewire API [Limewire] sucht. Dazu wurde die Limewire API erweitert, so dass direkter Zugriff auf Medienströme möglich wurde.

Neben dem eigentlichen Aufruf der Suchschnittstelle der assoziierten Internet-Tauschbörse hat eine Suchkomponente zwei weitere Aufgaben: die Übersetzung von Suchparametern und Suchergebnissen. Übersetzung meint hier die Abbildung allgemeiner Metasuchparameter auf die konkreten Parameter der Internet-Tauschbörse sowie die Abbildung der im Suchergebnis enthaltenen speziellen Metadaten auf allgemeine. Abbildung 2 stellt die Aufgaben einer Suchkomponente schrittweise dar.

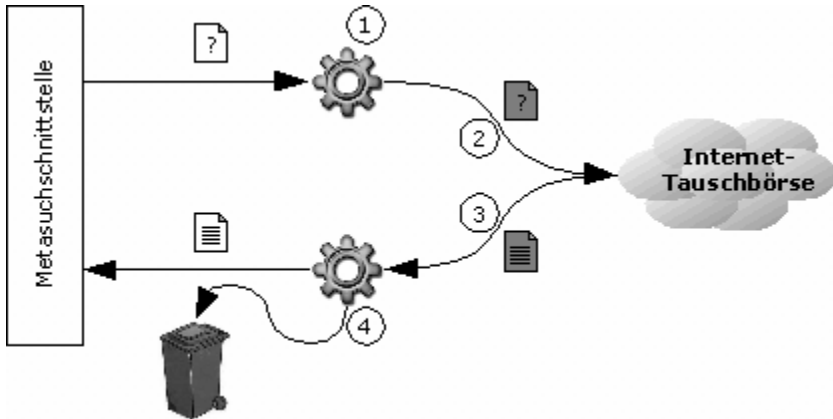


Abbildung 2: Aufgaben einer Suchkomponente in schrittweiser Darstellung

Die einzelnen Schritte sollen im Folgenden genauer beschrieben werden:

1. Die Suchkomponente transformiert die von der Metasuchschnittstelle stammende allgemeine Suchanfrage in eine Suchanfrage für die assoziierte Internet-Tauschbörse. Alle allgemeinen Metadaten werden auf die durch die Internet-Tauschbörse zur Formulierung von Suchanfragen unterstützten Metadaten abgebildet. Dabei werden alle allgemeinen Metadaten ignoriert, für die es keine passende Abbildung gibt. Sie sind erst in einem späteren Schritt (4) wieder relevant.
2. Die Suchkomponente sendet die formulierte Suchanfrage an die Suchschnittstelle der Internet-Tauschbörse. Dies kann beispielsweise wie in unserer Implementierung mit einer API wie Limewire geschehen.
3. Die Suchkomponente empfängt synchron oder asynchron ein Suchergebnis von der Suchschnittstelle der Internet-Tauschbörse.
4. Die speziellen Metadaten des Suchergebnisses müssen nun von der Suchkomponente auf die Metadaten eines allgemeinen Metasuchergebnisses abgebildet werden. Wie genau diese Abbildung aussieht, ist teil der Implementierung der Suchkomponente. Da im Suchergebnis zumeist ein größerer Umfang an Metadaten zur Verfügung steht als bei der Formulierung von Suchanfragen, wird die Konformität des Suchergebnisses mit der Suchanfrage nochmals geprüft. Ist diese Prüfung erfolgreich, handelt es sich um ein gültiges Suchergebnis, das an die Metasuchschnittstelle zurückgegeben werden kann – andernfalls wird es verworfen.

Der vorgestellte Prozess stellt eine zweistufige Metasuche in Internet-Tauschbörsen anhand von Metadaten dar: Im ersten Schritt werden so viele Metadaten der Metasuchanfrage wie möglich als Eingabe für die Suchschnittstelle der Internet-Tauschbörse verwendet. Nach Eintreffen eines Suchergebnisses werden die darin enthaltenen, umfangreicheren Metadaten für eine zweite Prüfung verwendet.

Im folgenden Abschnitt wird beschrieben, wie in einem der Metasuche nachgelagerten Schritt die Essenz eines Mediums zur Suche nach digitalen Wasserzeichen verwendet werden kann.

## **4 Aufdeckung von Rechtsverstößen durch digitale Wasserzeichen**

Das Ziel des vorgestellten Suchprozesses ist die Aufdeckung unberechtigter Veröffentlichung und Verbreitung digitaler Medien mittels digitaler Wasserzeichen. Alle Suchparameter, die durch vorhandene Internet-Tauschbörsen zur Verfügung gestellt werden, wurden im vorhergehenden Abschnitt in Form einer Metasuche abstrahiert.

Die Suche nach digitalen Wasserzeichen gehört nicht zum Umfang der unterstützten Parameter. Sie findet in einem zweiten Schritt statt: Das Ergebnis der Metasuche ist eine Menge möglicherweise unberechtigt veröffentlichter Medien. In diesen Medien muss nun nach eingebetteten digitalen Wasserzeichen gesucht werden, um je nach eingebetteten Wasserzeichen entweder die Datei eindeutig als unberechtigt vervielfältigte Kopien zu identifizieren oder aus den gefundenen Informationen auf die Quelle der Kopien zu schließen.

### **4.1 Detektion der Wasserzeichen**

Die Detektion digitaler Wasserzeichen hat zwei in diesem Kontext relevante Eigenschaften: sie findet zumeist auf unkomprimierten Daten statt und erzeugt bei einigen Wasserzeichenverfahren Pseudo-Daten als Ergebnis, auch wenn kein Wasserzeichen im Medium vorliegt. Dementsprechend muss das Auslesen eines Wasserzeichens in einem automatisierten System über das Ausführen des Detektors auf die Mediendatei hinausgehen. Die potentiell notwendigen Schritte werden in Abbildung 3 dargestellt. Die gefundene Mediendatei wird hier im Schritt der „Aufbereitung“ in ein für den Wasserzeichendetektor handhabbares Format umgewandelt. Oft handelt es sich hier um eine Wandlung von einem komprimierten Format wie jpg oder mp3 in Rohdaten wie Bitmap oder PCM. Danach wird die Detektion des Wasserzeichens durchgeführt, welche als Ergebnis die eingebetteten Daten ausgibt. Hier kann das Resultat entweder aus einem gefundenen Wasserzeichen, Pseudodaten oder einer Nachricht über das Nicht-Vorhandensein eines Wasserzeichens bestehen. Die ersten beiden Fälle werden im Schritt der „Validierung“ unterschieden. Die Validierung kann beispielsweise durch die Anwendung regulärer Ausdrücke auf den detektierten Daten stattfinden.

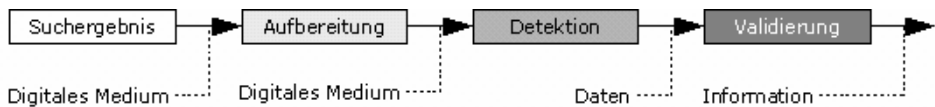


Abbildung 3: Inhaltsbasierte Suche am Beispiel digitaler Wasserzeichen

Die Verarbeitung der gefundenen Medienessenzen muss sich nicht auf die Detektion von Wasserzeichen beschränken.

#### 4.2 Erweiterung des Suchkonzeptes auf eine allgemeine inhaltsbasierte Suche

Der Prozess der Aufbereitung, Detektion und Validierung stellt eine mehrstufige inhaltsbasierte Suche in digitalen Daten dar. Die erste Stufe sucht in unserem Fall in Mediendaten nach digitalen Wasserzeichen, die zweite Stufe in Wasserzeichendaten nach Pseudo-Daten. Dieser Prozess lässt sich verallgemeinern, um auch andere mehrstufige inhaltsbasierte Suchvorgänge zu unterstützen – etwa Ähnlichkeitsanalysen von Bildern anhand von Kanten oder Schwerpunkten. Damit kombiniert dieser Ansatz eine metadatenbasierte Suche mit einer inhaltsbasierten. Durch die Trennung von inhaltsbasierter und metadatenbasierter Suche kann also inhaltlich gesucht werden, ohne dass dies die einzelnen Tauschbörsen selbst unterstützen.

Die dabei (durch die inhaltsbasierte Suche) zu bearbeitende Datenmenge ist durch die vorgeschaltete Metasuche bereits so weit eingeschränkt, wie es die zu Grunde liegenden Tauschbörsen zulassen. Inwieweit diese Reduktion der anfallenden Datenmengen ausreichend ist, hängt vom Einsatzgebiet der inhaltsbasierten Suche ab – im konkreten Fall digitaler Wasserzeichen hat es sich als attraktiver Ansatz herausgestellt.

### 5 Zusammenfassung und Ausblick

In der vorliegenden Arbeit stellen wir ein Konzept zur Verfolgung von Urheberrechtsverletzungen in P2P-Tauschbörsen auf der Basis digitaler Wasserzeichen vor. Dabei dienen die digitalen Wasserzeichen als Mechanismus, digitale Medien mit Zusatzinformationen anzureichern, anhand derer später die ursprünglichen Käufer identifiziert werden können. Werden so markierte Medien in P2P-Netze eingestellt, können diese aufgespürt und untersucht werden, um eine Verbindung mit den Käufern herzustellen und bei ihnen eine Verletzung des Urheberrechts zu beanstanden.

Wir zeigen, wie die Suche nach und die Analyse von digitalen Medien automatisiert werden kann. Da die Suche nach Wasserzeichen eine inhaltsbasierte Suche darstellt, die nicht im Umfang aktueller Tauschbörsen vorhanden ist, müssen zunächst Medien, die möglicherweise unberechtigt veröffentlicht wurden, in Tauschbörsen gefunden werden. Dies geschieht über eine Metasuche anhand von Metadaten. Metasuchanfragen werden dabei an mit speziellen Tauschbörsen assoziierte Suchkomponenten delegiert. Die dabei gefunden Medien werden dann in einem zweiten Schritt einer inhaltsbasierten Suche unterzogen, die eingebettete Wasserzeicheninformation ausliest.

Somit entsteht ein System, welches selbständig unterschiedliche P2P-Netze anhand einer beschreibenden Metadaten nach potentiellen Urheberrechtsverstößen durchsuchen kann. Dadurch wird eine deutlich effizientere und umfassendere Suche nach illegalen Kopien in entsprechenden Netzen möglich. Das hier vorgestellte Konzept ist bereits für das Gnutella-Netzwerk umgesetzt, weitere Netze werden derzeit bearbeitet. Ziel ist es, eine zentrale Steuerung zu entwickeln, die parallel in verschiedenen Netzen nach Urheberrechtsverletzungen suchen kann und dabei die Suche hinsichtlich von Redundanz und Downloadgeschwindigkeit selbständig optimiert.

## Literaturverzeichnis

- [B06] GfK Gruppe (im Auftrag der Deutschen Phonoverbände); Brenner-Studie 2006, <http://www.miz.org/artikel/brennerstudie2006.pdf>, April 2006
- [BEPW03] Peter Biddle, Paul England, Marcus Peinado, Bryan Willman; The Darknet and the Future of Content Protection; Lecture Notes in Computer Science, Volume 2696, Jan 2003 (ACM Workshop on DRM), Pages 155 - 176
- [CMB02] Cox, Miller, Bloom; Digital Watermarking, Academic Press, San Diego, USA, ISBN 1-55860-714-5, 2002
- [D06] Diener; Technologische und ökonomische Aspekte des internetgestützten Auffindens von mit digitalen Wasserzeichen versehenen Medien – Masterarbeit, Fachbereich Informatik, Hochschule Darmstadt, 2006
- [D00] Dittmann; Digitale Wasserzeichen, Springer Verlag, Berlin, ISBN 3-540-66661-3, 2000
- [GHJV96] Gamma, Helm, Johnson, Vlissides; Entwurfsmuster – Elemente wiederverwendbarer objektorientierter Software, Addison-Wesley, München, ISBN 3-8273-1862-9, 1996
- [Limewire] <http://www.limewire.org/>, Stand: 06.2006
- [SDL03] Steinebach, Dittmann, Lang; Konzepte zur Vermeidung oder Verfolgung von Urheberrechtsverletzungen in Netzwerken auf der Basis digitaler Wasserzeichen, Competence in Content, Ralf Schmidt (Hrsg.), Tagungsband 25. Online-Tagung der DGI, S. 113 – 125, ISBN 3-925474-58-x, 2003
- [SW06] Steinebach, Wolf; Mark – Search – Retrieve Challenges and strategies for watermarking-based copyright protection, eingereicht zu: Proceedings of AXMEDIS2006 - 2nd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution, Leeds, UK, December 2006



# **Pervasive University**

