情報処理学会研究報告
IPSJ SIG Technical Report

P2P ファイル交換ソフトウェア環境 Share における効率的なファイル保持ノード特定手法

寺田真敏 [1)]　重本倫宏 [1)]　仲小路博史 [1)]　吉成愛子 [2)]
宮川雄一 [2)]　湯川隆司 [2)]　鵜飼裕司 [3)]　金居良治 [3)]

Winny, Share などの P2P ファイル交換ソフトウェア環境におけるファイル流通は匿名性が高いために，ファイル保持ノードを特定にしくい．このため，マルウェアや流出ファイルの削除，著作権上適切ではないファイルの監視など，P2P ファイル交換ソフトウェア環境にすでに流通してしまっているファイルに関して対策を取りにくい状況にある．本稿では，P2P ファイル交換ソフトウェア Share を対象に，クローリング手法を用いたファイル保持ノード特定方法を提案し，検証実験を通してその有効性を示す．

The Observation method of the nodes with complete cache files on Share P2P network

Masato Terada[1)], Tomohiro Shigemoto[1)]
Hirofumi Nakakoji[1)], Aiko Yoshinari[2)], Yuichi Miyagawa[2)]
Ryuji Yukawa[2)] Yuji Ukai[3)] and Ryoji Kanai[3)]

P2P file exchange software is spreading on the Internet. The requirements of observation method of the nodes with complete cache files on Share P2P network are increasing. In this paper, we describe the observation method of the nodes with complete cache files by file keep flag in P2P key information. Also, we show some experiment results for P2P network "Share" enforced on StarBED that is a Large Scale Network Experiment Environment.

1. はじめに

P2P(Peer to Peer)ファイル交換ソフトウェア利用が広がる中，ウイルス感染によるファイルの流出，著作権上適切ではないファイル交換による著作権侵害は続いており，その被害が顕在化している状況にある．Web サイトによるファイル掲載は掲載ノードを特定しやすいが，Winny，Share などの P2P ファイル交換ソフトウェア環境におけるファイル流通は匿名性が高いために, ファイル保持ノードを特定にしくい. このため，マルウェアや流出ファイルの削除，著作権上適切ではないファイルの監視など，P2P ファイル交換ソフトウェア環境にすでに流通してしまっているファイルに関して対策を取りにくい状況にある．そこで，本稿では，P2P ファイル交換ソフトウェア Share を対象に，クローリング手法を用いたファイル保持ノード特定方法を提案する．

本稿の構成について述べる．2 章で P2P ファイル交換ソフトウェア環境におけるファイル保持ノード特定に関する既存方式の概要と課題を示す．3 章で Share を対象としたクローリング手法を用いたファイル保持ノード特定方法を提案し，4 章でネットワーク実験環境 StarBED[1]上に構築した 120 台規模の閉域環境下での実験を通して，提案方式の有効性を示す．

2. 関連研究

P2P 通信技術については，ネットワーク上のトラフィック分散を実現する技術として期待されている一方，国内で普及している P2P ファイル交換ソフトウェア環境は，著作権上適切ではないファイルやマルウェアファイルなどの流通基盤と化している状況にある. 本章では, 国内における P2P ファイル交換ソフトウェア環境の利用状況と，P2P ファイル交換ソフトウェア環境におけるファイル保持ノードの特定に関する既存方式について述べる．

2.1 P2P ファイル交換ソフトウェア環境の利用状況

(1) P2P 利用者数と稼動ノード数

文献 2) (調査時期：2009 年 10 月)では，約 2 万名を対象としたアンケート調査結果から，ファイル交換ソフトウェアの利用者がインターネット利用者の 9.1%(2008 年 9 月の調査では 10.3%)となり，若干の利用者減少を報告している．また，クローリング

[1)] (株)日立製作所　システム開発研究所
System Development Lab. Hitachi Ltd.
[2)] (株)クロスワープ
CROSSWARP Inc.
[3)] (株)フォティーンフォティ技術研究所
Fourteenforty Research Institute, Inc.

手法を用いて収集したデータを元に Winny 稼動ノード数 18 万台以上(／日)[*]，Share 稼動ノード約 21 万台以上(／日)と推定しており，それぞれ 9 割以上が日本からのアクセスであるとしている．

(2) ファイル流通状況

(a) 著作権上適切ではないファイル

文献 2)では(調査時期：2009 年 10 月)，Winny ネットワーク上のファイルは約 513 万 5 千件(／日)存在し，流通するファイル全体の約 47.6%が著作物，Share ネットワーク上のファイルは約 69 万 7 千件(／日)存在し，流通するファイル全体の約 52.7%が著作物と推定している．また，文献 3)では(調査時期：2008 年 1～2 月)，Winny ネットワーク上に流通しているファイルの約 6 割強が著作物と推測されるファイルであることと，映像系ファイル(64%)が多く流通しており，そのほとんどが許諾の無いと推測されることを報告している．

(b) マルウェアファイル

文献 3)では，Winny ネットワーク上ではマルウェア単体での流通は稀であり，9 割以上がアーカイブファイル(zip，lzh など)に混入して流通していること，収集した全ての zip，lzh，rar コンテンツ中，19.0%がマルウェアを含むコンテンツであることを報告している．

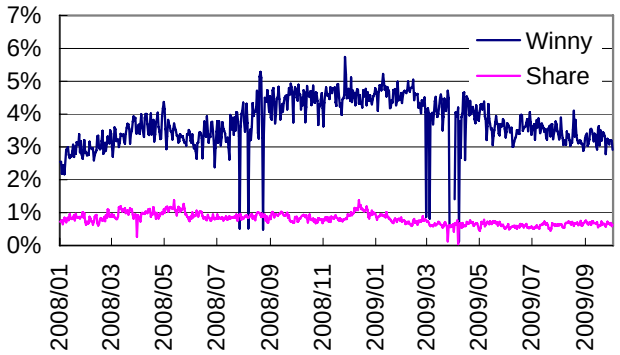

図 1：情報漏えいファイルの流通比率

*) 2009 年 10 月の調査では、Winny 稼動ノード数として約 40 万台相当の一意なノード情報を収集している。ただし、インデックスポイゾニングされたノード情報が流布していたために、稼動ノード数の推定が困難としている。このため、2008 年 9 月の調査結果を記載した。

(c) 情報漏えいファイル

著者らのサンプリングによるクローリング調査によれば，情報漏えいファイルに見られる特徴的なファイル名のみを抽出した場合，Winny ネットワーク上には約 3.7%(／日)，Share ネットワーク上には約 0.8%(／日)のファイルが流通していると推定される(図 1).

2.2 ファイル保持ノードの特定

Winny ネットワーク上のファイル保持ノードの特定については，広域を探査しながらファイル保持ノードを特定する方法と，特定ノードに対象を絞ってファイル保持を判定する方法がある．前者の広域を探査しながらファイル保持ノードを特定する方法は，クローリング調査において，ファイルの所在情報が記載されたキーの出現回数が多い場合には，該当する実ファイルを保持している可能性が高いことに着目した手法である[4][5]. 後者の特定ノードに対象を絞ってファイル保持を判定する方法は，経路情報に 5 ホップ分の架空の経路をあらかじめ書き込んだ検索クエリ(0 ホップ検索クエリ)を特定ノードに送信し，その応答から該当する実ファイルを保持しているかどうかを判定する[6].

Share ネットワーク上のファイル保持ノードの特定については，クローリング調査において，ファイルの所在情報が記載されたキーの出現回数が多い場合には，該当する実ファイルを保持している可能性の高い傾向のあることは示されているが[4]，Winny ネットワーク上のファイル保持ノードの特定のような方法論が確立されていない．本稿の目的では，P2P ファイル交換ソフトウェア Share を対象としたファイル保持ノードを特定する方法を提案することにある．

3. 提案手法

本章では，Share を対象としたクローリング手法において，ファイルの所在情報が記載されたキーを取得する際に得られる特定のフラグ(以降，ファイル保持フラグ)の参照により，該当する実ファイルの保持ノードを特定する手法を提案する．

3.1 クローリング手法を用いたファイル保持ノードの特定

P2P モデルで構成されたファイル交換システムには，Napster[7]のようにノード情報やファイルの所在を中央サーバで管理するハイブリッド P2P ファイル交換システムと，Winny，Share，Gnutella のように，全ての処理を P2P で行なうピュア P2P ファイル交換システムがある．ピュア P2P の場合，不特定多数のノードと能動的に通信する必要があり，全てのノードは他ノード情報を保持している．クローリング手法は，ノード

が保持している他ノード情報一覧を取得するという操作を，取得した他ノード情報を用いて繰り返していく事で，ピュア P2P を構成するノードを調査するという方法である(図 2)．また，ファイルの所在を示すキー情報を同時に収集することで，ファイルの流通状況やファイル保持ノードに関する情報を収集できる．

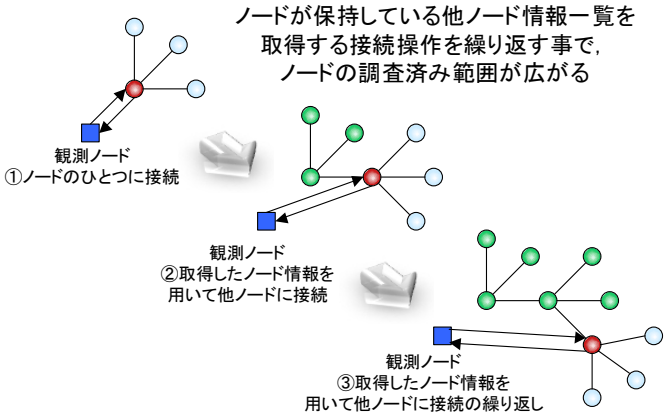

図 2：クローリング手法の概要

観測ノードには，図 3 に示すモジュールから構成された Share ネットワーク用観測ツールを搭載し，クローリング手法を実現している．観測エンジンは，接続と同時に各種情報を送信するモードに移行させる開始要求を送信した後，ノード情報やファイル名を含むキー拡散コマンド(コマンド番号 0)を処理する．また，得られたキー情報から他ノード情報を取り出し，接続操作を繰り返すことで，Share ネットワーク上で交換されているファイルのプロパティが含まれるキー情報を網羅的に収集し，これら収集したキー情報をデータベースに格納する．観測ツールで観測可能な項目を表 1，表 2 に示す．

提案方式は，Share ネットワーク用観測ツールによって観測可能な項目のうち，用途不明なフラグが，ファイル保持に関連していることに着目する．

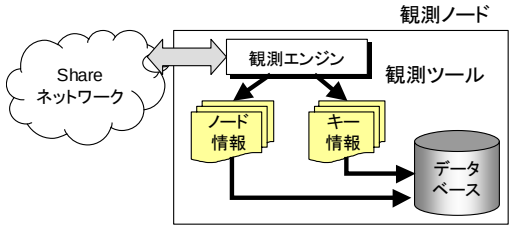

図 3：観測ツールのモジュール構成

表 1：ノード情報として観測可能な項目

分類	項目
ノード情報	IP アドレス
	ポート番号

表 2：キー情報として観測可能な項目

分類	項目
キー情報	キーの IP アドレス
	キーのポート番号
	キーの更新日時
	ファイルのサイズ
	ファイルのハッシュ値
	ファイルの名称
	用途不明なフラグ(以降，ファイル保持フラグ)
	ニックネーム
	ID
付加情報	接続ノードの IP アドレス
	接続ノードのポート番号
	接続ノードからのキー取得時刻

4. 提案手法の検証

本章では，検証実験の目的ならびに検証環境について述べる．

4.1 目的

本検証実験の目的は，用途不明なフラグがファイル保持に関連していること，クロ

ーリング手法を通して，実ファイルの保持ノードを特定できることを閉域環境下での検証実験を通して示すことにある．

(1) **キャッシュファイル種別とファイル保持フラグとの関係**

Share ノード上のキャッシュフォルダに格納されるキャッシュファイル種別を表 3 に示す．本検証項目では，キャッシュファイル種別毎に，ファイル保持フラグに格納された値との関係を明らかにする．

(2) **クローリング手法によるキー取得時刻とファイル作成時刻との整合性**

キャッシュファイル種別とファイル保持フラグとの関係が明らかになったファイルにおいて，観測ツールによるキー取得時刻と作成されたファイルの最終更新日付の視点からファイル保持フラグの有効性を確認する．

表 3：Share EX2 のキャッシュファイル種別

分類		説明
UP キャッシュ		アップロードファイルのキャッシュ (キャッシュブロック保有率=100%)
完全キャッシュ		ダウンロード操作によりキャッシュブロック保有率=100%となったキャッシュ
部分キャッシュ		ダウンロード操作によりキャッシュブロック保有率=0%以上100%未満であるキャッシュ
拡散 (diffuse) キャッシュ	完全	キャッシュブロック保有率=100%となった拡散キャッシュ (キャッシュモニタでは，拡散キャッシュ(完全)と完全キャッシュとの判別はできない)
	部分	キャッシュブロック保有率=0%以上 100%未満である拡散キャッシュ
非保持		上記以外(キャッシュファイルを保持していない状態)

4.2 検証環境

4.2.1 ネットワーク構成

StarBED のグループ H に属するノード(CPU Intel QuadCore Xeon X3350 2.66GHz，メモリ 8GB，HDD160GB) 170 台に，WMware ESXi を用いて 6 仮想ノード／物理ノードとし，計 1,020 台のノードから成るネットワークを構築した．このうち 120 台の仮想ノードに Windows XP と表 4 に示す設定をした Share EX2 を稼動させ，Share ノードによる P2P ネットワークを構成した(図 4).

表 4：Share EX2 の設定(全ノード共通)

項目	設定値
送信速度	10,000K バイト/秒
受信速度	10,000K バイト/秒
最大キー保持数	10,000 件
クォータ機能使用有無	有
クォータ機能使用時	削除サイズ 1,000M バイト 削除実行サイズ 100M バイト
自動ダウンロード設定	自動ダウンロードキーワードをランダムに割り当て

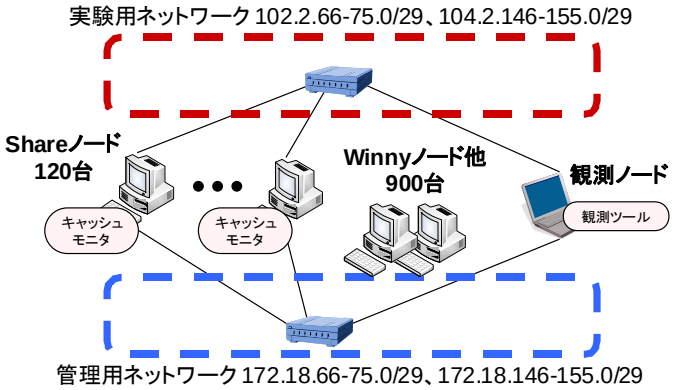

図 4：ネットワーク構成

4.2.2 ツール

本検証実験において使用したツールの概要について述べる．

(1) **キャッシュモニタ**

Share には，ノード上のキャッシュフォルダに格納されているキャッシュファイルを管理するインデックス情報がある．キャッシュモニタは，このインデックス情報を参照するツールで[8]，表 3 に示すキャッシュファイルの分類と共に，ダウンロード済

みサイズ，ファイルの最終更新日付などの情報を取得できる．

(2) 観測ツール

図 3 に示すモジュールから構成された Share ネットワーク用観測ツールである．観測ツールが出力するファイル保持フラグには，ON か OFF が格納される．

4.3 検証方法

本節では，4.1 節で提示した目的に対する検証方法について述べる．

(1) キャッシュファイル種別とファイル保持フラグとの関係

検証実験開始時と終了時に，キャッシュモニタを用いて取得した個々のキャッシュファイル種別に，観測ツールを用いて取得したファイル保持フラグの値を対応付ける．

(2) クローリング手法によるキー取得時刻とファイル作成時刻との整合性

観測ツールによるキー取得時刻と，検証実験終了時に，キャッシュモニタを用いて取得した個々のキャッシュファイルの作成時刻とを比較する．

5. 検証結果

本章では，検証実験の結果について述べる．

5.1 キャッシュファイル種別毎の状態

検証実験開始時と終了時に，キャッシュモニタを用いて取得した個々のキャッシュファイル種別毎の件数を表 5 に示す. 120 台から構成した Share ネットワーク上には，ファイルのハッシュ値を用いて識別した一意のファイルは 2009 件，“IP アドレス＋ファイルのハッシュ値”を用いて識別した一意のファイルは開始時に 2,511 件，終了時 4,724 件となった．Share ネットワークによって，検証実験中にファイル 2,213 件が複製されたことになる．

表 5 に記載したファイル消失とは，開始時にキャッシュモニタで参照可能なインデックス情報として存在していたキャッシュファイルが，終了時にインデックス情報から消失していたことを示す。内訳に記載された非保持とは，開始時にキャッシュモニタで参照可能なインデックス情報にキャッシュファイルが存在しておらず、検証実験中に、新たなキャッシュファイルとしてインデックス情報に登録されたことを示している．

5.2 キャッシュファイル種別とファイル保持フラグとの関係

観測ツールを用いて取得したキー情報は 2,419 件であり，このうち，ファイル保持フラグ ON は 2,272 件，OFF は 147 件であった．また，取得したキー情報に含まれている，“キーの IP アドレス＋ファイルのハッシュ値”を用いてキー情報を識別した場合，重複は発生していなかった．

観測ツールを用いて取得したファイル保持フラグの値と，検証実験開始時と終了時に，キャッシュモニタを用いて取得した個々のキャッシュファイル種別との対応付けを表 6 に示す．ファイル保持フラグが ON である 2,272 件のうち，2,252 件(99%)がキャッシュブロック保有率=100%に関するファイルであることを示している．

表 5：キャッシュファイル種別毎の状態

ファイル種別		開始時	終了時	種別遷移の内訳(件数)	
UP キャッシュ		2,009	2,009	－	
完全キャッシュ		358	1,909	完全キャッシュ	356
				部分キャッシュ	39
				拡散キャッシュ(部分)	2
				非保持	1512
部分キャッシュ		125	647	部分キャッシュ	81
				非保持	566
拡散キャッシュ	完全	0	0	－	
	部分	19	159	拡散キャッシュ(部分)	17
				非保持	142
ファイル消失		－	(7)	完全キャッシュ	(2)
				部分キャッシュ	(5)
計		2,511	4,724	－	

表 6：キャッシュファイル種別毎の状態

ファイル保持フラグ	件数	内訳(件数) 開始時	終了時	
ON	2,272	UP キャッシュ	UP キャッシュ	430
		完全キャッシュ	完全キャッシュ	357
		部分キャッシュ／非保持	完全キャッシュ (表 7)	1,465
		完全キャッシュ	ファイル消失	2
		非保持	完全キャッシュ (但し，ファイル未作成)	18
OFF	147	拡散キャッシュ(部分)	完全キャッシュ (表 8)	2
		拡散キャッシュ(部分)	拡散キャッシュ(部分)	14
		部分キャッシュ	完全キャッシュ	0
		非保持	完全キャッシュ (表 9)	12
		非保持	拡散キャッシュ(部分)	119

5.3 クローリング手法の調査時刻とファイル作成時刻との整合性

開始時に完全キャッシュでなく，終了時に完全キャッシュとして構成されたファイルは，この検証実験期間中に，キャッシュファイル種別が変わっていることになる．ここでは，表 6 の中から「部分キャッシュ／非保持→完全キャッシュ」,「拡散キャッシュ(部分)→完全キャッシュ」,「非保持→完全キャッシュ」の 3 つについて，観測ツールによるキー取得時刻と，検証実験終了時に，キャッシュモニタを用いて取得した個々のキャッシュファイルの作成時刻とを比較する．

- 部分キャッシュ／非保持→完全キャッシュ

  ファイル保持フラグ ON のキー情報のうち, 観測ツールによるキー取得時刻が, キャッシュファイル作成時刻よりも後のキーが 726 件，前のキーが 739 件となった(表 7)．約半分のキーがファイル作成時刻よりも早い，すなわち，完全キャッシュファイルができあがる前に，観測ツールはファイル保持フラグ ON のキーを取得する可能性があるという結果が得られた．

表 7：部分キャッシュ／非保持→完全キャッシュ

項目	時刻	件数	内訳(件数)	
完全キャッシュ (終了時) 1,465 件	取得時刻≧ファイル作成時刻	726	部分→完全	19
			非保持→完全	707
	取得時刻＜ファイル作成時刻	739	部分→完全	20
			非保持→完全	719

キー取得時刻が，キャッシュファイル作成時刻よりも前のキー739 件について，キー取得時刻～キャッシュファイル作成時刻の時刻差の分布を図 5 に示す．約 3 割のキーがキー取得からキャッシュファイル作成までに 60 分以上要している．

- 拡散キャッシュ(部分)→完全キャッシュ

  ファイル保持フラグ OFF のキー情報 2 件は, いずれも観測ツールによるキー取得時刻が，キャッシュファイル作成時刻よりも前となっている(表 8).

- 非保持→完全キャッシュ

  ファイル保持フラグ OFF のキー情報 12 件のうち 1 件は，観測ツールによるキー取得時刻(15:31:46)が, キャッシュファイル作成時刻(15:31:28)よりも後ではあるが，ほぼ同時刻であった(表 9).

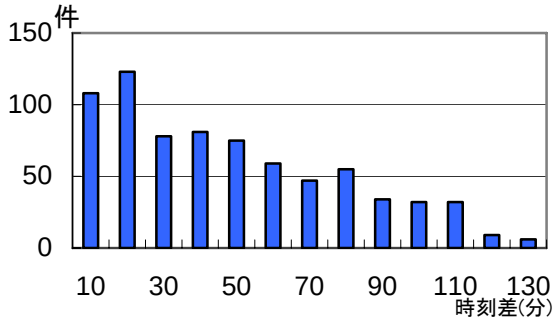

図 5：キャッシュファイル作成時刻－キー取得時刻の分布

表 8：拡散キャッシュ(部分)→完全キャッシュ

項目	時刻	件数
拡散キャッシュ(部分) →完全キャッシュ 2 件	取得時刻≧ファイル作成時刻	0
	取得時刻＜ファイル作成時刻	2

表 9：非保持→完全キャッシュ

項目	時刻	件数
非保持 →完全キャッシュ 12 件	取得時刻≧ファイル作成時刻	1
	取得時刻＜ファイル作成時刻	11

5.4 考察

ファイル保持フラグ ON となったキー情報については，いずれもキャッシュファイル種別が，アップロードキャッシュ，完全キャッシュに関わるものである．このことから，ファイル保持フラグは，広域を探査するクローリング手法において，Share ノードが復元可能なファイルを保持しているか否かを特定する際に利用できる．

文献 4)では，クローリング手法によるキー調査において，ファイルを保持していないキー情報を観測したと報告している．本検証実験でも，完全キャッシュファイル作成時刻よりも前に，ファイル保持フラグ ON をなったキーを取得するという類似の状況が発生した．このことから，1 回だけのファイル保持フラグ取得で復元可能なファイル保持ノードを特定することはできず，ファイル保持フラグ ON となったキーの出現回数や専用ダウンロードツールによる直接確認を併用することにより，ファイル保持ノードの特定精度を高める必要がある．

6. おわりに

本稿では，StarBED 上に 120 台の Share ノードによる P2P ネットワークを構成し，ファイル所在情報(キー情報)に付随する用途不明なフラグ(ファイル保持フラグ)がアップロードキャッシュ，完全キャッシュなど復元可能なファイルの保持に関連していることを検証を通して明らかにした．また，広域を探査するクローリング手法において，ファイル保持フラグが Share ノードが復元可能なファイルを保持しているか否かを効率的に特定する手法として利用できることを示した．

専用ダウンロードツールを併用した広域での検証でも，ファイル保持フラグを用いた本提案方式が，リモートから Share ノードが保持するファイルの特定に有効であることを示している[9]．今後の課題は，マルウェアや流出ファイルの削除，著作権上適切ではないファイルの監視などの運用を踏まえた本提案方式の適用検討などが挙げられる．

謝辞 大規模ネットワーク実験環境 StarBED を本実験環境として利用するにあたりご協力を頂いた独立行政法人情報通信研究機構北陸リサーチセンター，ICT 研究開発機能連携推進会議(HIRP)の関係者各位に深く感謝致します．また，StarBED 上の実験環境構築にあたり，有益な助言と協力を頂いた北陸先端科学技術大学院大学ならびに，独立行政法人情報通信研究機構北陸リサーチセンターの篠田陽一教授，三輪信介氏，宮地利幸氏，中井浩氏，安田真悟氏に深く感謝致します．

本研究は総務省から委託を受けた「ネットワークを通じた情報流出の検知及び漏出情報の自動流通停止のための技術開発」の支援を受け実施している．本研究を進めるにあたって有益な助言と協力を頂いた関係者各位に深く感謝致します．

参考文献

[1] StarBED Project, http://www.starbed.org/
[2] 社団法人コンピュータソフトウェア著作権協会:第 8 回「ファイル共有ソフト利用実態調査」, (2009 年 12 月 12 日), http://www2.accsjp.or.jp/research/research09.php
[3] 寺田真敏 他:P2P ファイル交換ソフトウェア環境における情報流通対策向けデータベースの検討，情報処理学会 CSEC 研究報告 Vol.2008 No.71, pp.123-128 (2008)
[4] 寺田真敏 他:クローリング手法を用いた P2P ネットワークの観測，情報処理学会 CSEC 研究報告 Vol.2007 No.48, pp.51-56 (2007)
[5] クロスワープ:Winny ネットワークにおける安全で効率的な著作権侵害監視について(2007), http://www.crosswarp.com/info/070118.pdf
[6] 吉田雅裕 他：Winny ネットワークに対するインデックスポイズニングを用いたファイル流通制御方式，情報処理学会論文誌 Vol.50 No.9, p.2008-2022 (2009)
[7] Napster, http://www.napster.com/
[8] ShareCacheList, http://p2p-db.net/
[9] クロスワープ：Share ネットワークにおける安全で効率的な著作権侵害監視手法について (2009)，http://www.crosswarp.com/info/090706.pdf