

ファイル交換ソフト利用実態クローリング調査概要

2007年12月21日

1	目的	1
2	方法	1
2-1	データの収集	1
2-2	日時	1
2-3	利用したソフトウェア(クローラー)	1
2-4	無許諾コンテンツの流通状況	1
3	調査結果	3
3-1	Winny2	3
3-1-1	キー総量について	3
3-1-2	ノード量について	3
3-1-3	ファイル量について	3
3-1-4	無許諾コンテンツの流通状況	4
3-2	Share EX2	5
3-2-1	キー総量について	5
3-2-2	ノード量について	5
3-2-3	ファイル量について	5
3-2-4	無許諾コンテンツの流通状況	6
3-3	WinMX3	7
3-3-1	無許諾コンテンツの流通状況	7

社団法人 コンピュータソフトウェア著作権協会 (ACCS)

社団法人 日本レコード協会 (RIAJ)

日本国際映画著作権協会 (JIMCA)

1 目的

ファイル交換ネットワークの把握と分析

2 方法

2-1 データの収集

各ファイル交換ネットワークに対応した手法を用いて実際のネットワークを巡回（クロウリング）し、実際にネットワーク上を流通している情報を自動収集し、分析を行った。

2-2 日時

2007/9/28(金) 17:00 から 2007/9/29(土) 17:00 (24 時間)

2-3 利用したソフトウェア（クローラー）

Winny2 P2P FINDER (Winny) 2007 年 9 月 Version

Share EX2 P2P FINDER (Share) 2007 年 9 月 Version

WinMX3 WinMX3.31 + MXPie hosts/FILE FINDER (WinMX) 2007 年 9 月 Version

WinMX3のソフトウェアそのものをオートパイロットするシステムを用いて、WinMX3ネットワーク上に流通するファイルの検索とファイル情報を自動保存した。

WinMX はインデックスサーバーに対して、ファイル名をフルパスで送信するため、キーワードに偏らずに検索を行えるように、ドライブ名およびWindows ネットワークドライブ名にマッチするように設定を行った。

また、WinMXのサービスしていたFrontcode Technologies社は既にインデックスサービスを運営していない為、WPN(WinMX Peer Netork)サーバーはMXPie (<http://www.mxpie.info/>)のhosts書き換えによる互換サーバーへのアクセスを行った。

2-4 無許諾コンテンツの流通状況

ファイル交換ソフトネットワーク上で、権利者に無許諾で送信可能な状態におかれ、流通しているファイルの調査を行った。

ファイル交換ソフトのネットワーク上から自動収集したキー情報（ファイル名やサイズ、ハッシュ値（ファイル内容の識別情報）、ファイルの位置情報としてのIPアドレスなどが含まれた情報。一部のファイル交換ソフトは、利用者の持つファイルに関するキーを自動生成してネットワーク上に流通させ、検索等に利用させる）の総取得件数から下記手順にて重複件数を削除した上で、約2万件を無作為に抽出し、抽出データに含まれるファイル名を目視において確認し、各ファイルについて推定されるジャンル、権利の対象及び許諾の有無について調査した。

	工程
総取得件数	クローラーにより、IP、ポート番号、ファイル名、時間を収集。
重複件数の削除	で取得したデータのうち、IP、ポート番号、ファイル名が重複したデータを削除。

間引き後数	で取得したデータから 20,000 件を無作為に抽出。
アダルトキーワード除去	で抽出したデータのうち、アダルトキーワード(「18 禁」「無修正」など)を除外。
共通除外ワード適用	のデータから更に除外キーワード(「同人」「ハッシュュリスト」など)を除外。
キーワード抽出	のデータを各ジャンルのキーワードで分類。

3 調査結果

3-1 Winny2

3-1-1 キー総量について

1日間の総取得件数 146,362,055 件のうち、ノード情報（IP アドレスおよびポート）とファイル情報（ハッシュ値）について、一意な組み合わせの量をキー総量とすると、本調査では 1 日で 37,310,415 件の一意なキー情報が収集された。

3-1-2 ノード量について

IP アドレスとポート番号の一意な組み合わせをノードの量として算出した。本調査では、一日で 264,252 件の一意なノード情報を収集した。全数としてはおよそ 30 万ノード程度と推定。

3-1-3 ファイル量について

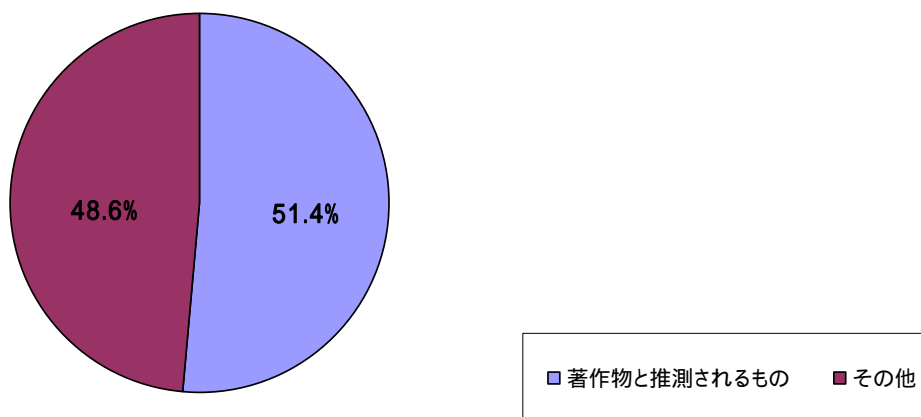
Winny 上ではファイルの情報はファイル本体から算出されるハッシュ値を用いて管理されているため、一意なハッシュ値の数を算出する事で Winny 上で流通しているファイルの量を推定できる。本調査では、一日で 4,846,175 件の一意なハッシュ値を収集した。全数としてはおよそ 550 万件程度と推定。

3-1-4 無許諾コンテンツの流通状況

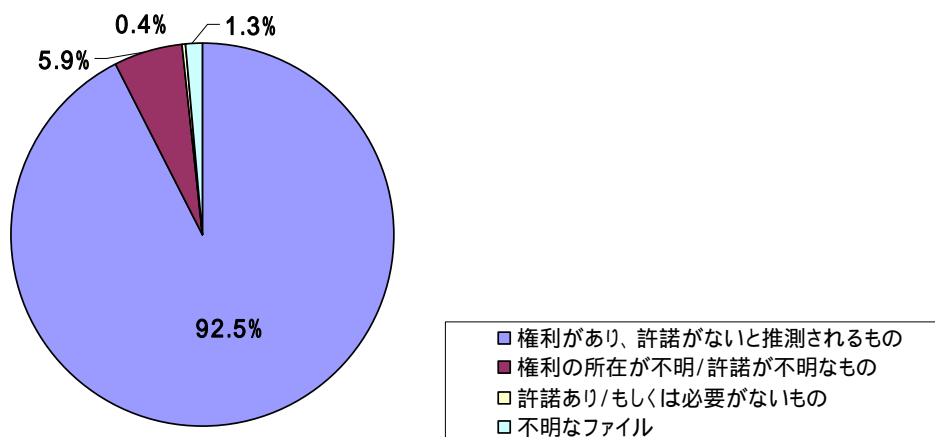
2万件の抽出データのうち、51.4%が何らかの著作物であると推測される。残りの48.6%は、アダルトや同人など本調査では権利の所在が確認できないもの。

著作物と推測されるファイルのうち、著作権があり、かつ許諾がないと推測されるものが92.5%となっている。

【図3-1 無許諾コンテンツの流通状況～本調査対象】



【図3-2 無許諾コンテンツの流通状況～権利の対象性】



3-2 Share EX2

3-2-1 キー総量について

1日間の総取得件数 272,72,190 件のうちノード情報（IP アドレス）とファイル情報（ハッシュ値）について、一意な組み合わせの量をキー総量とすると、本調査では1日で 11,012,210 件の一意なキー情報が収集された。

3-2-2 ノード量について

IP アドレスとポート番号の一意な組み合わせをノードの量として算出した。本調査では、一日で 201,845 件の一意なノード情報を収集した。全数としてはおよそ 21 万～22 万ノード程度と推定。

3-2-3 ファイル量について

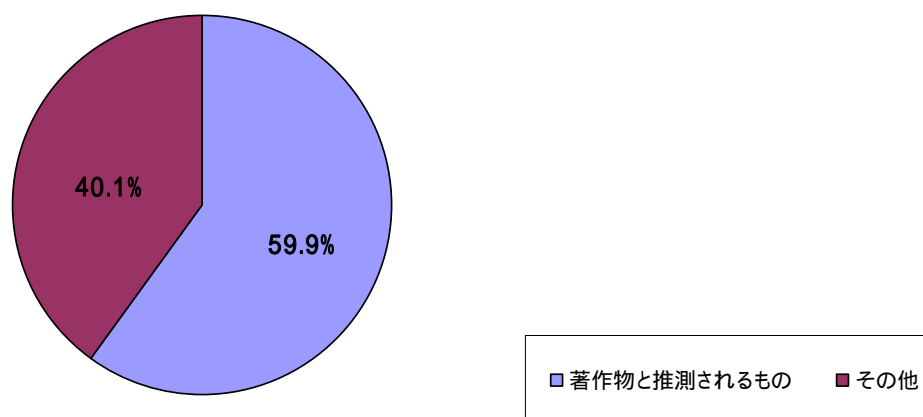
Share 上ではファイルの情報はファイル本体から算出されるハッシュ値を用いて管理されているため、一意なハッシュ値の数を算出する事で Share 上の流通しているファイルの量を推定できる。本調査では、一日で 549,124 件の一意なハッシュ値を収集した。全数としてはおよそ 65 万～70 万件程度と推定。

3-2-4 無許諾コンテンツの流通状況

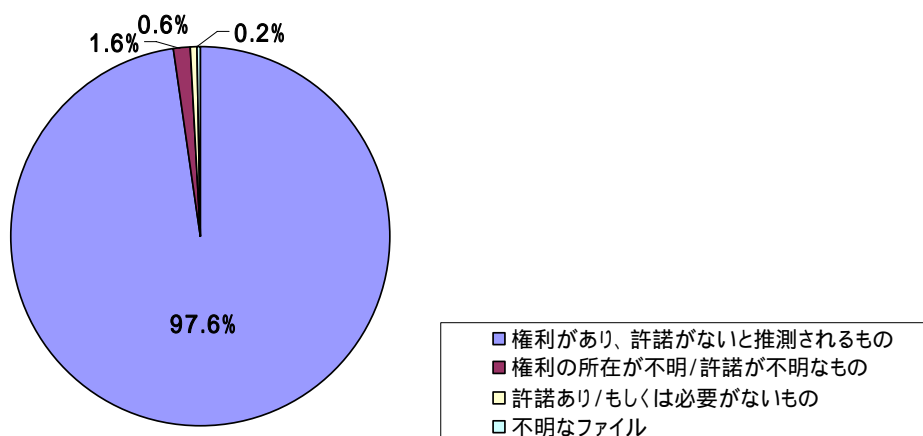
2万件の抽出データのうち、59.9%が何らかの著作物であると推測される。残りの40.1%は、アダルトや同人など本調査では権利の所在が確認できないもの。

著作物と推測されるファイルのうち、著作権があり、かつ許諾がないと推測されるものが97.6%となっている。

【図3-3 無許諾コンテンツの流通状況～本調査対象】



【図3-4 無許諾コンテンツの流通状況～権利の対象性】



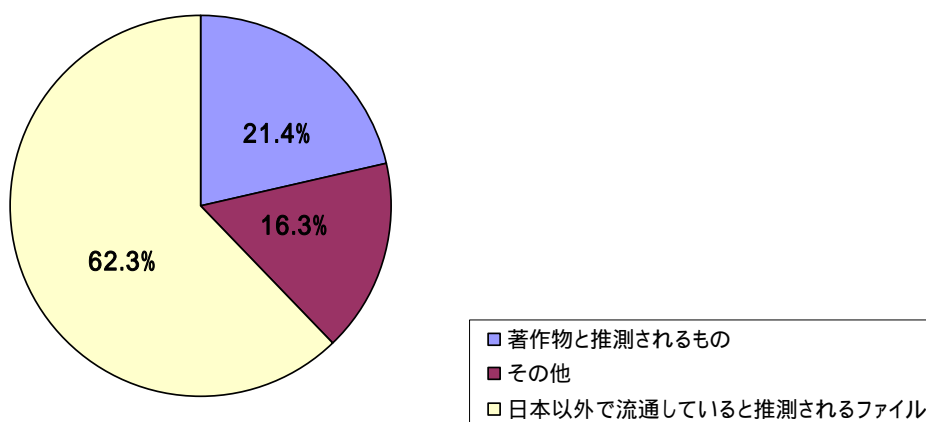
3-3 WinMX3

3-3-1 無許諾コンテンツの流通状況

2万件の抽出データのうち、21.4%が何らかの著作物であると推測される。残りの78.6%のうち、16.3%はアダルトファイルや同人など、62.3%は主として日本以外で流通している（ファイル名が外国語で記載されている）コンテンツと推測され、本調査では権利の所在が確認できないもの。

著作物と推測されるファイルのうち、著作権があり、かつ許諾がないと推測されるものが96.5%となっている。

【図3-5 無許諾コンテンツの流通状況～本調査対象】



【図3-6 無許諾コンテンツの流通状況～権利の対象性】

