

**ファイル共有ソフトの利用に関する調査
～クローリング調査～**

報 告 書

2013年2月

一般社団法人コンピュータソフトウェア著作権協会

目 次

I. 調査概要	1
1. 調査実施日	1
2. 調査手法	1
(1) 基礎情報	1
3. データの抽出	1
(1) フィルタリング	1
(2) 権利の対象性の算出方法	2
II. 調査結果	3
1. Winny	3
(1) コンテンツの流通状況	3
(2) 権利の対象性	5
(3) 検出ノードの国・地域	5
2. Share	6
(1) コンテンツの流通状況	6
(2) 権利の対象性	8
(3) 検出ノードの国・地域	8
3. PerfectDark	9
(1) コンテンツの流通状況	9
(2) 権利の対象性	11
(3) 検出ノードの国・地域	11

I. 調査概要

1. 調査実施日

2013年1月18日 17:00～1月19日 17:00 (24時間)

2. 調査手法

調査は、2(1)のP2Pネットワークに対応した手法を用いてネットワークを巡回(クローリング)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、分析する形で実施した。

Winny、Share、PerfectDarkそれぞれのプロトコルを利用したクローラを用いて、特にキーワードを設定することなくWinnyネットワーク上に流通するキー情報(ノード情報、ファイル情報)の自動収集を行った。複数のクローラを用いる事で、24時間でほぼネットワークの全域をクローリングできる性能を確保している。

(1) 基礎情報

	Winny	Share	PerfectDark
利用したソフトウェア	P2P FINDER(Winny) 2011年7月Version	P2P FINDER(Share) 2011年7月Version	P2P FINDER(PerfectDark) 2011年7月Version
設定情報	総スレッド数 1,200	総スレッド数 4,000	総スレッド数 400

3. データの抽出

(1) フィルタリング

ファイル共有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。

調査を行うにあたり、総取得件数からノード(IPとポート)およびファイル名が同一なデータを取り除いた後、(調査対象データ)、2万件をランダムに抽出した。2万件の調査対象データから、アダルト系キーワード、共通除外キーワードを含むデータを除外し、データを目視にて確認し、各ファイルについて推定されるジャンル、権利の対象および許諾の有無について調査した。

	工程
①総取得件数	クローラにより IP、ポート、ファイル名、時間を取得
②重複件数の削除	①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 (PerfectDarkについてはファイルネーム、化身 ID)
③間引き後件数	②で取得したデータを 20,000 件になるようにランダムに抽出
④アダルトキーワード除去	③で抽出したデータのうちファイル名にアダルトコンテンツと想定されるキーワードがあるデータを除外
⑤共通除外キーワード除去	④のデータのファイル名に共通除外キーワードがあるデータを除外
⑥キーワード抽出	⑤のデータを各ジャンルのキーワードで抽出

(2) 権利の対象性の確認方法

3. (1) で抽出したデータを目視にて以下のジャンルに分類を行った。

著作物と推定されるもの	音楽、映像、プログラム、書籍、その他（音声、画像など）に分類されるもので、本調査で権利の所在が確認できるもの
アダルト	成人を対象に提供されているコンテンツで、本調査で権利の所在が確認できないため、権利の対象に関する調査を行わなかったもの
同人	同人サークルによって制作されたコンテンツで、本調査で権利の所在が確認できないため、権利の対象に関する調査を行わなかったもの
不明ファイル	ファイルのタイトルからは、コンテンツの内容が確認できないもの
危険ファイル	ファイルのタイトル、拡張子からウイルスなどと推定されるもの

II. 調査結果

1. Winny

本調査でIPアドレスとポート番号の一意的な組み合わせをノードの量として算出したところ、19,550件であった。全数として約20,000ノードと推定される。

(1) コンテンツの流通状況

流通コンテンツのうち、およそ半数のコンテンツが著作物と推定される。

著作物と推定されるもの	8,550	音楽	1,016		
		映像	4,002	映画	28
				映画(海外)	79
				映像	524
				映像(海外)	69
				アニメ	3,123
				PV	179
		プログラム	311	アプリ	77
				ゲーム	234
		書籍	2,281	書籍	580
				コミック	1,701
		その他	940		
アダルト	6,764				
同人	3,723				
不明ファイル	485				
危険ファイル	18				
情報	460				
計	20,000				

※「著作物と推定されるもの」とは本調査で権利の所在が推定できるもの

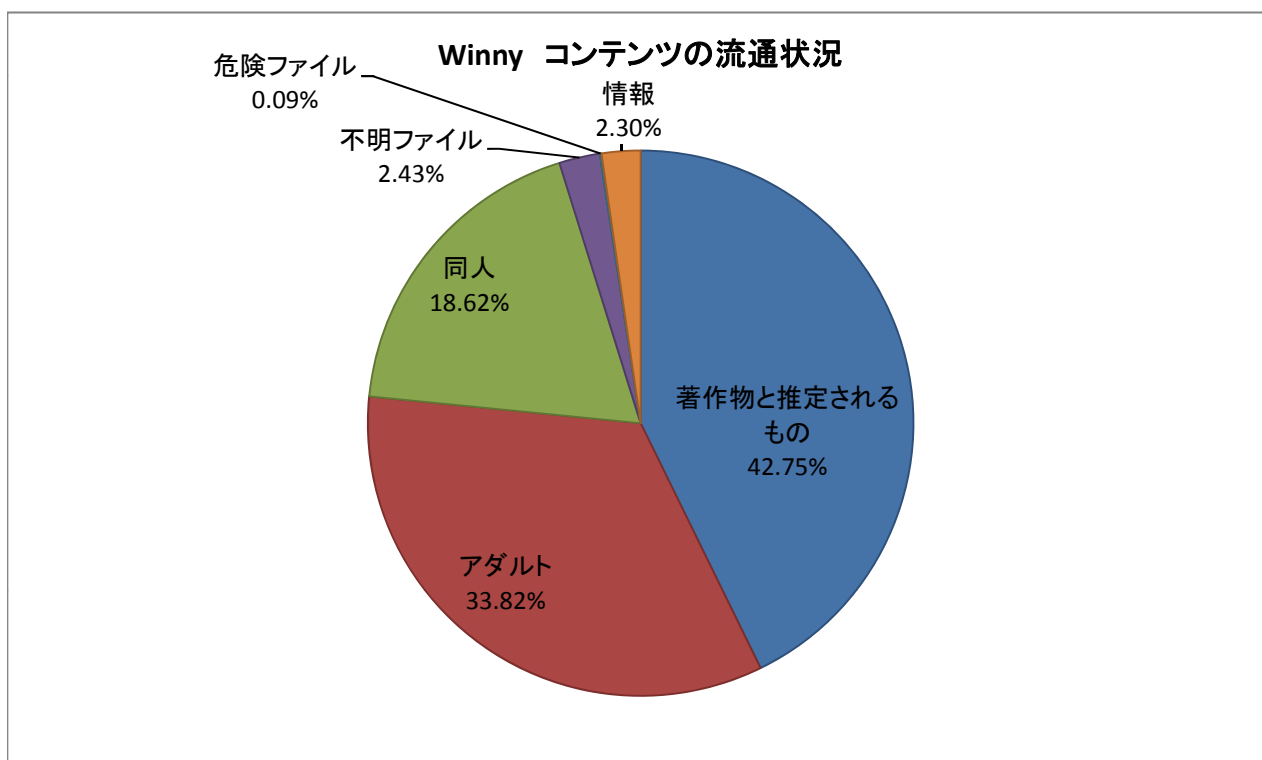
※「アダルト」、「同人」とは本調査で権利の所在が判別できないため、権利の対象に関する調査は見送ったもの

※「不明ファイル」とはタイトルからコンテンツの内容が判別できないもの

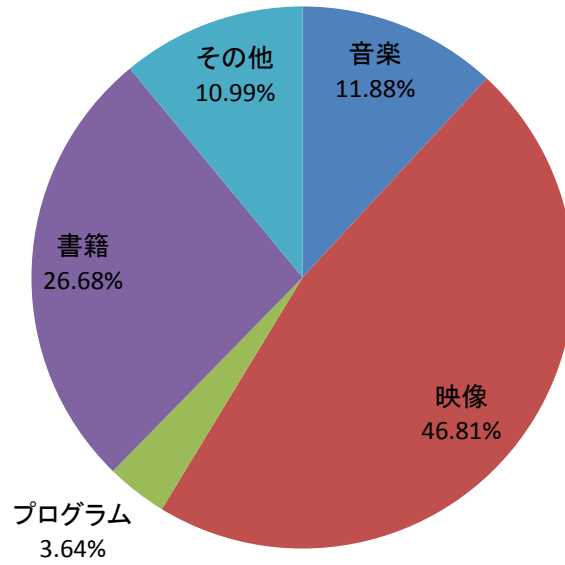
※「危険ファイル」とはタイトル、拡張子からウイルスなどと推定されるもの

※「情報」とはウイルス感染などで流出した個人・組織等の情報だと推定されるもの

※「その他」とはコンテンツの分類が音楽、映像関連、プログラム、書籍関連に含まれないもの



Winny 著作物と想定されるコンテンツの内訳



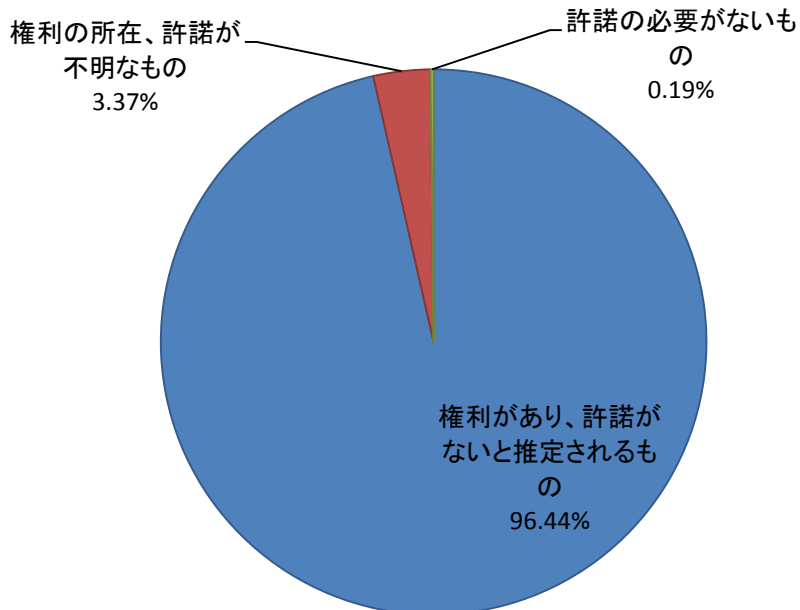
(2) 権利の対象性

① 全体

著作物と推定されるもののうち、96.44%に権利があり、かつ許諾がないものと推定される。

権利があり、許諾がないと推定されるもの	8,246
権利の所在、許諾が不明なもの	288
許諾の必要がないもの	16

Winny 権利の対象性(全体)



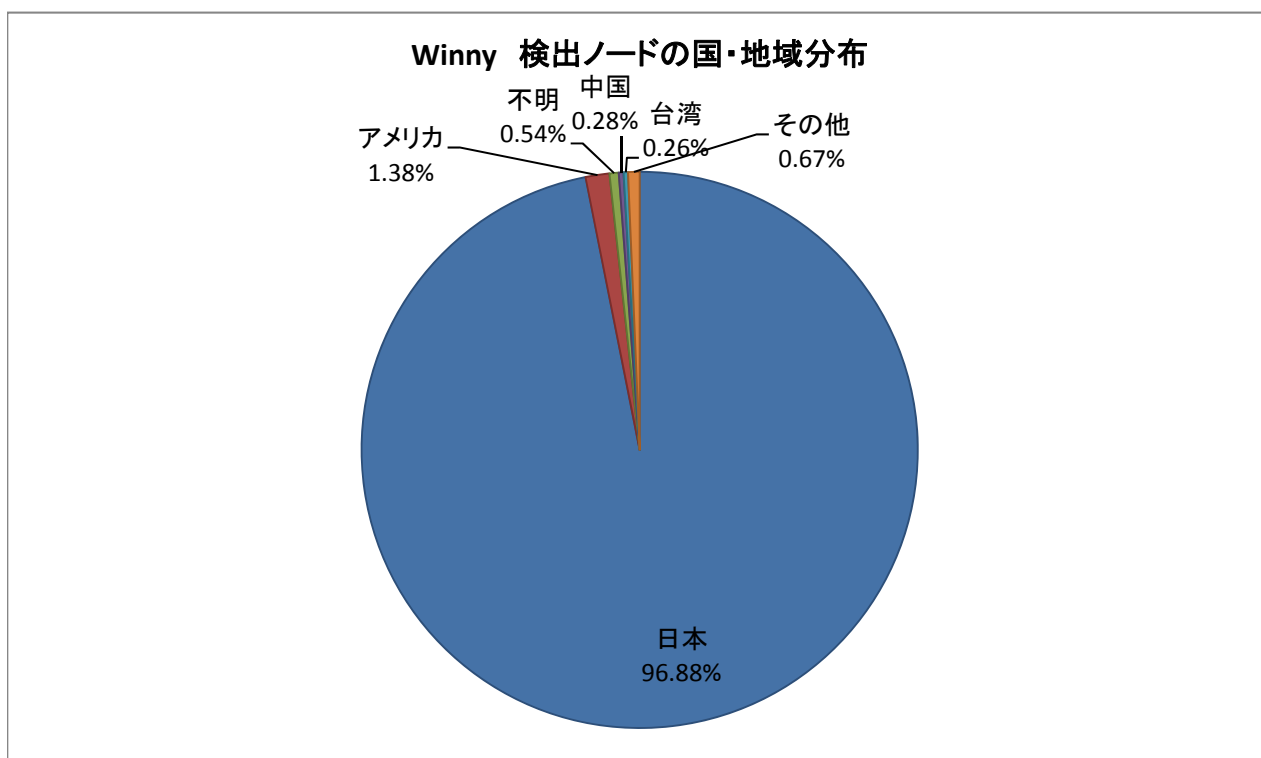
②ジャンルごと

権利の対象性について、コンテンツジャンルごとに集計した。音楽、映像、プログラム、書籍の著作物については90%以上、その他の著作物については75%以上が権利の対象であり、無許諾で送信されていると推定されるものであった。

	権利あり	権利の所在、許諾不明	許諾の必要なし
音楽	97.24% (988)	1.97% (20)	0.79% (8)
映像	99.63% (3,987)	0.37% (15)	0.00% (0)
プログラム	91.64% (285)	6.75% (21)	1.61% (5)
書籍	99.82% (2,277)	0.09% (2)	0.09% (2)
その他	75.43% (709)	24.47% (230)	0.11% (1)

(3) 検出ノードの国・地域

一日で検出した19,550件のノードについて国・地域を調べた結果、約96.88%が日本国内IPの利用であった。



2.Share

本調査でIPアドレスとポート番号の一意的な組み合わせをノードの量として算出したところ、58,683件が検出された。全数として約59,000ノードと推定される。

(1)コンテンツの流通状況

流通コンテンツのうち、およそ半数のコンテンツが著作物と推定される。

著作物と推定されるもの	9,849	音楽	686		
		映像	6,327	映画	41
				映画(海外)	201
				映像	568
				映像(海外)	196
				アニメ	4,936
				PV	385
		プログラム	272	アプリ	132
				ゲーム	140
		書籍	2,173	書籍	258
				コミック	1,915
		その他	391		
アダルト	5,339				
同人	3,932				
不明ファイル	849				
危険ファイル	0				
情報	31				
計	20,000				

※「著作物と推定されるもの」とは本調査で権利の所在が推定できるもの

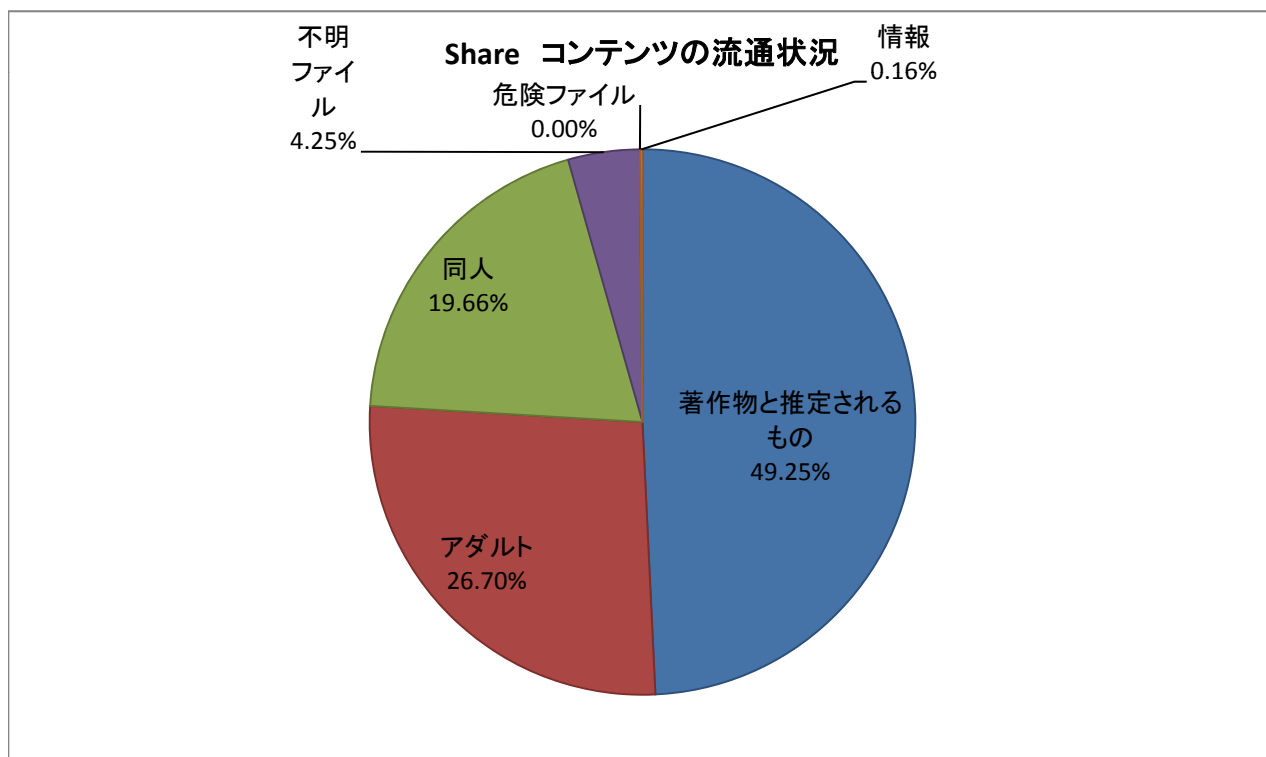
※「アダルト」、「同人」とは本調査で権利の所在が判別できないため、権利の対象に関する調査は見送ったもの

※「不明ファイル」とはタイトルからコンテンツの内容が判別できないもの

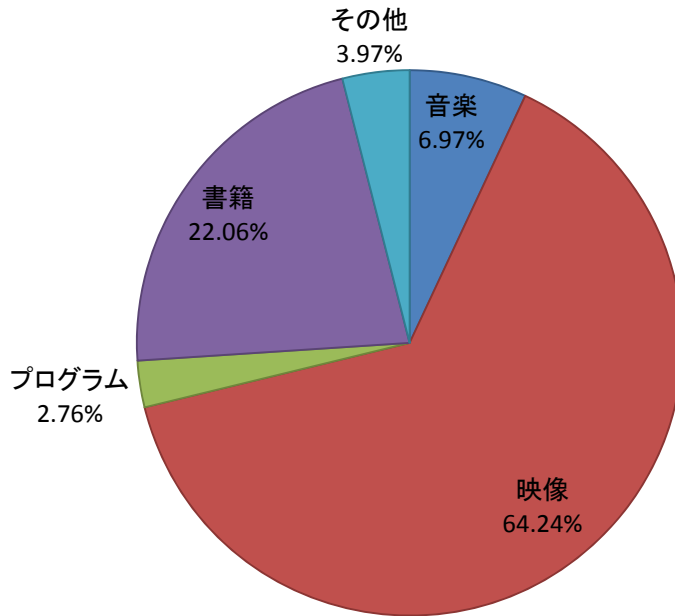
※「危険ファイル」とはタイトル、拡張子からウイルスなどと推定されるもの

※「情報」とはウイルス感染などで流出した個人・組織等の情報だと推定されるもの

※「その他」とはコンテンツの分類が音楽、映像関連、プログラム、書籍関連に含まれないもの



Share 著作物と推定されるコンテンツの内訳



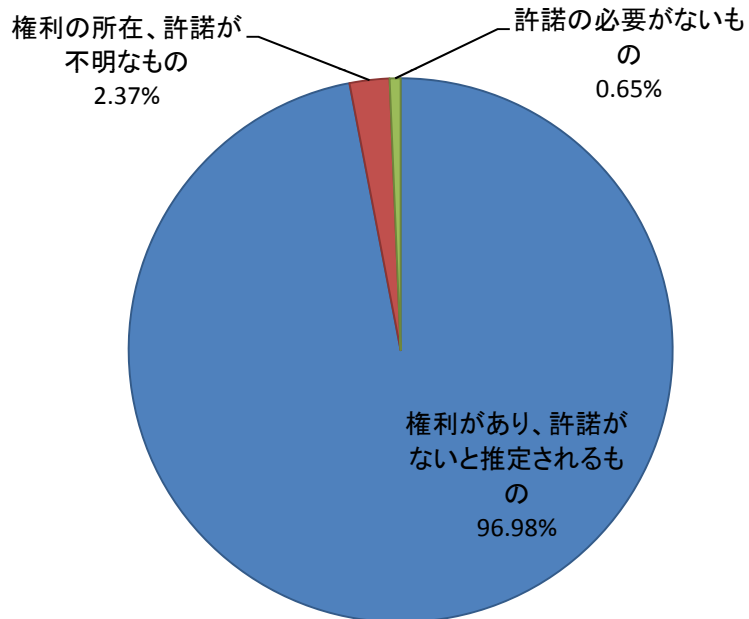
(2) 権利の対象性

① 全体

著作物と推定されるもののうち、96.98%に権利があり、かつ許諾がないものと推定される。

権利があり、許諾がないと推定されるもの	9,552
権利の所在、許諾が不明なもの	233
許諾の必要がないもの	64

Share 権利の対象性



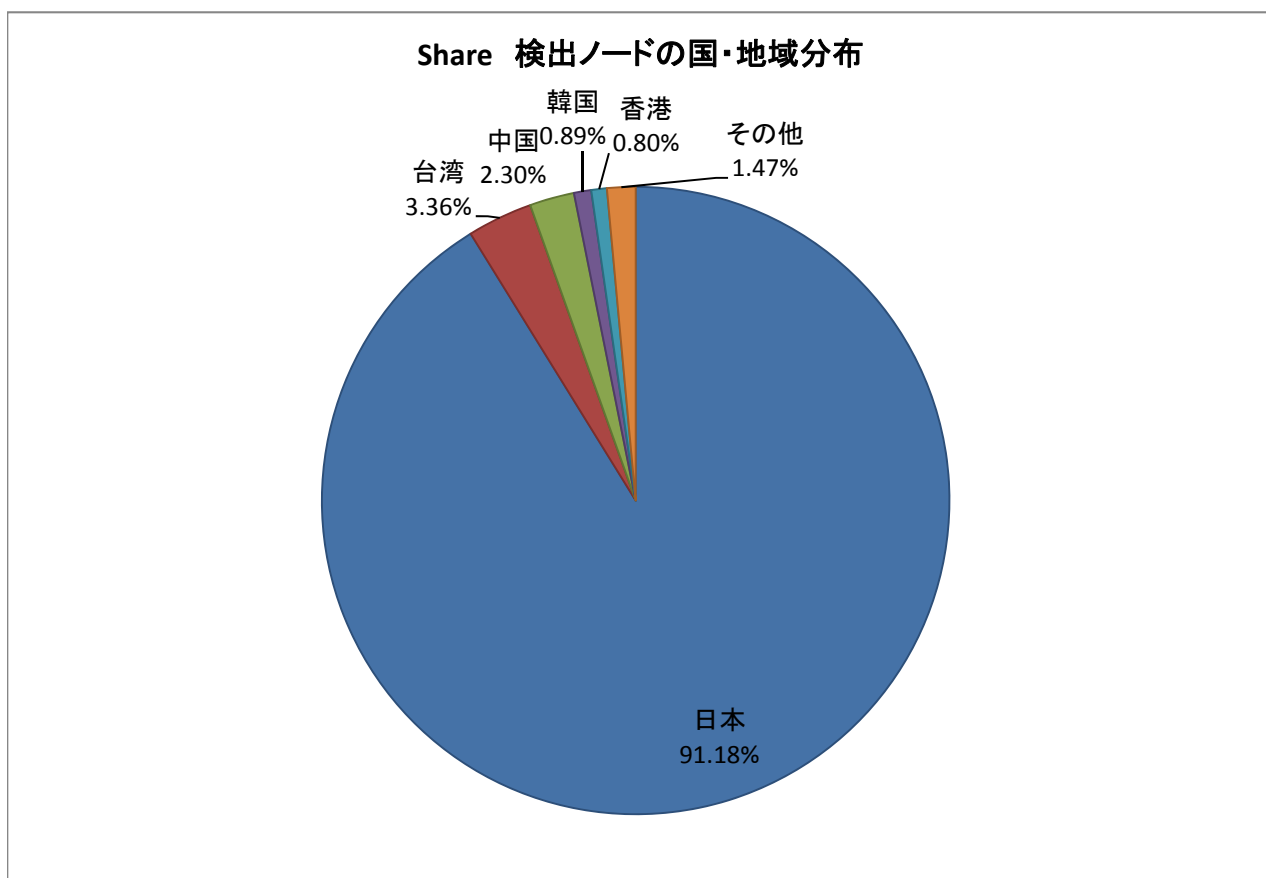
②ジャンルごと

権利の対象性について、コンテンツジャンルごとに集計した。音楽、映像、書籍の著作物については90%以上、プログラムの著作物については80%以上、その他の著作物については70%以上が権利の対象であり、無許諾で送信されていると推定されるものであった。

	権利あり	権利の所在、許諾不明	許諾の必要なし
音楽	92.42% (634)	2.62% (18)	4.96% (34)
映像	98.67% (6,243)	1.33% (84)	0.00% (0)
プログラム	81.99% (223)	7.35% (20)	10.66% (29)
書籍	99.86% (2,170)	0.09% (2)	0.04% (1)
その他	72.12% (282)	27.88% (109)	0.00% (0) (0)

(3) 検出ノードの国・地域

一日で検出した58,683件のノードについて国・地域を調べた結果、約91.18%が日本国内IPの利用であった。



3.PerfectDark

本調査でIPアドレスとポート番号の一意的な組み合わせをノードの量として算出したところ、34,158件が検出された。全数として約34,000ノードと推定される。

(1)コンテンツの流通状況

流通コンテンツのうち、およそ6割のコンテンツが著作物と推定される。

著作物と推定されるもの	12,140	音楽	1,410		
		映像	6,274	映画	68
				映画(海外)	153
				映像	1,562
				映像(海外)	582
				アニメ	3,743
				PV	166
		プログラム	719	アプリ	289
				ゲーム	430
		書籍	2,730	書籍	644
				コミック	2,086
		その他	1,007		
アダルト	5,103				
同人	2,379				
不明ファイル	353				
危険ファイル	0				
情報	25				
計	20,000				

※「著作物と推定されるもの」とは本調査で権利の所在が推定できるもの

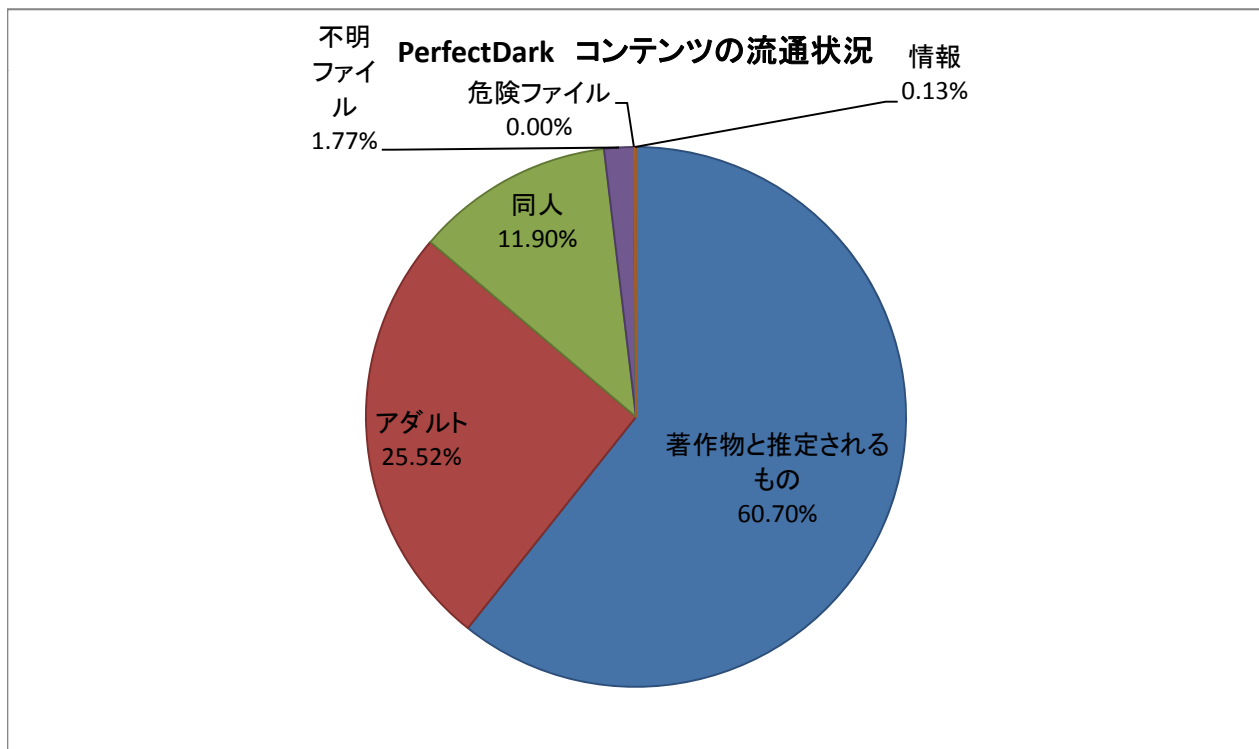
※「アダルト」、「同人」とは本調査で権利の所在が判別できないため、権利の対象についての調査は見送ったもの

※「不明ファイル」とはタイトルからコンテンツの内容が判別できないもの

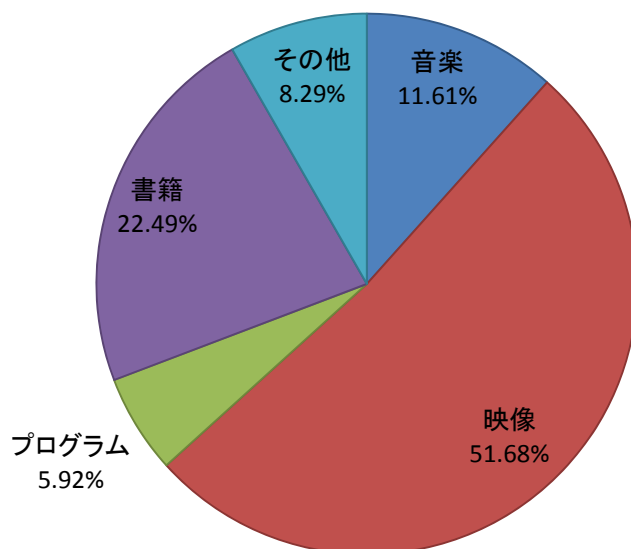
※「危険ファイル」とはタイトル、拡張子からウイルスなどと推定されるもの

※「情報」とはウイルス感染などで流出した個人・組織等の情報だと推定されるもの

※「その他」とはコンテンツの分類が音楽、映像関連、プログラム、書籍関連に含まれないもの



PerfectDark 著作物と推定されるコンテンツの内訳



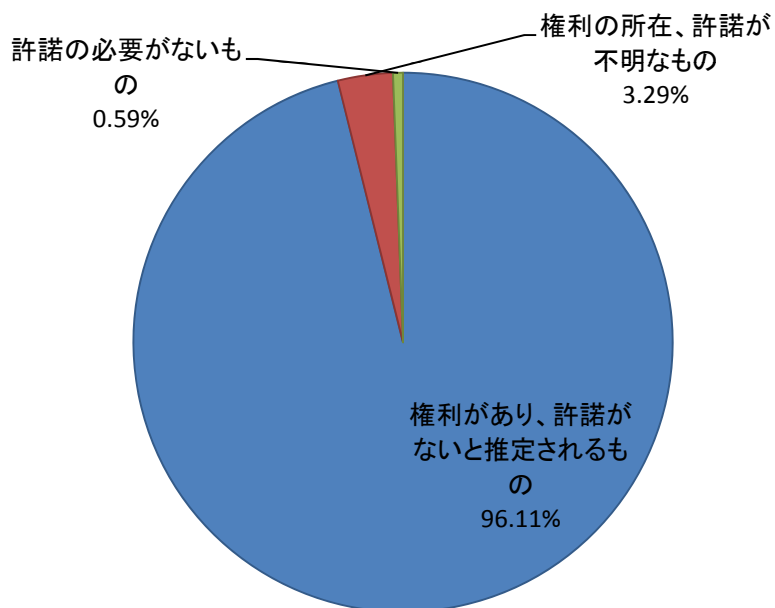
(2) 権利の対象性

①全体

著作物と推定されるもののうち、94.02%に権利があり、かつ許諾がないものと推定される。

権利があり、許諾がないと推定されるもの	11,668
権利の所在、許諾が不明なもの	400
許諾の必要がないもの	72

PerfectDark 権利の対象性(全体)



②ジャンルごと

権利の対象性について、コンテンツジャンルごとに集計した。音楽、映像、プログラム、書籍の著作物については90%以上、その他の著作物については85%以上が権利の対象であり、無許諾で送信されていると推定されるものであった。

	権利あり	権利の所在、許諾不明	許諾の必要なし
音楽	97.38% (1,373)	1.70% (24)	0.92% (13)
映像	97.05% (6,089)	2.95% (185)	0.00% (0)
プログラム	86.23% (620)	6.12% (44)	7.65% (55)
書籍	99.85% (2,726)	0.00% (0)	0.15% (4)
その他	85.40% (860)	14.60% (147)	0.00% (0)

(3) 検出ノードの国・地域

一日で検出した34,158件のノードについて国・地域を調べた結果、約95.16%が日本国内IPの利用であった。

