

**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

2012年5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. 第1回調査結果	3
1. Winny	3
(1) コンテンツの流通状況	3
(2) 権利の対象性	5
2. Share	6
(1) コンテンツの流通状況	6
(2) 権利の対象性	8
3. PerfectDark	9
(1) コンテンツの流通状況	9
(2) 権利の対象性	11
III. 第2回調査結果	12
1. Winny	12
(1) コンテンツの流通状況	12
(2) 権利の対象性	14
2. Share	15
(1) コンテンツの流通状況	15
(2) 権利の対象性	17
3. PerfectDark	18
(1) コンテンツの流通状況	18
(2) 権利の対象性	20

I. 調査概要

1. 調査実施日

第1回：2011年11月25日17:00～11月26日17:00（24時間）

第2回：2012年3月2日17:00～3月3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合法ファイル抽出	④のデータのファイル名に「合法」のキーワードがあるデータを抽出
⑦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) 権利の対象性の算出方法

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
- ・ アダルト
- ・ 同人
- ・ 不明ファイル
- ・ 危険ファイル
- ・ 合法ファイル

Ⅱ. 第1回調査結果

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、全数として約40,000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	9,497	音楽	1,501			
		映像	4,012	映画		39
				映画(海外)		88
				映像		614
				映像(海外)		102
				アニメ		3,039
				PV		130
		プログラム	531	アプリ		138
				ゲーム		393
		書籍	2,441	書籍		650
コミック				1,791		
その他	1,012					
アダルト	6,380					
同人	3,313					
不明ファイル	429					
危険ファイル	9					
情報	372					
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

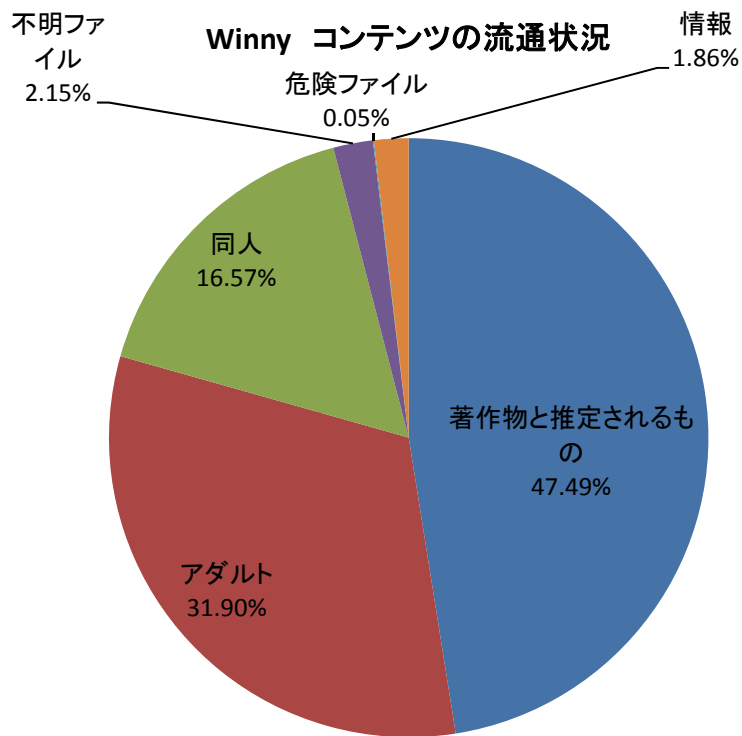
※「アダルト」、「同人」とは本調査で権利の所在が判別できないため、権利の対象に関する調査は見送ったもの

※「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

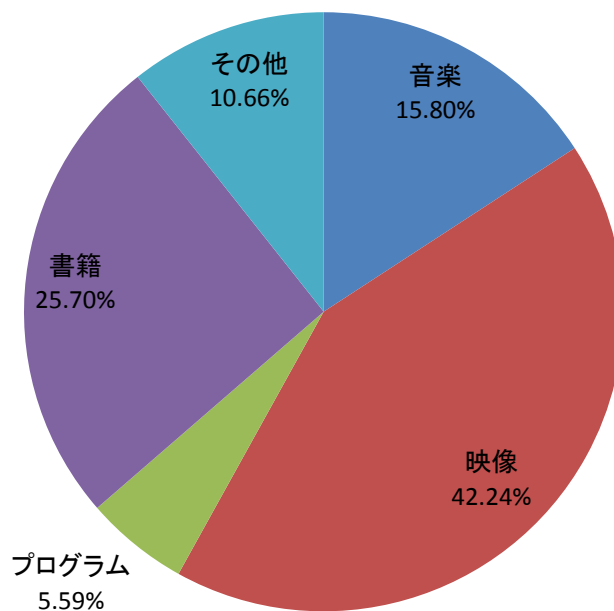
※「危険ファイル」とはタイトル、拡張子からウイルスなどと推定されるもの

※「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

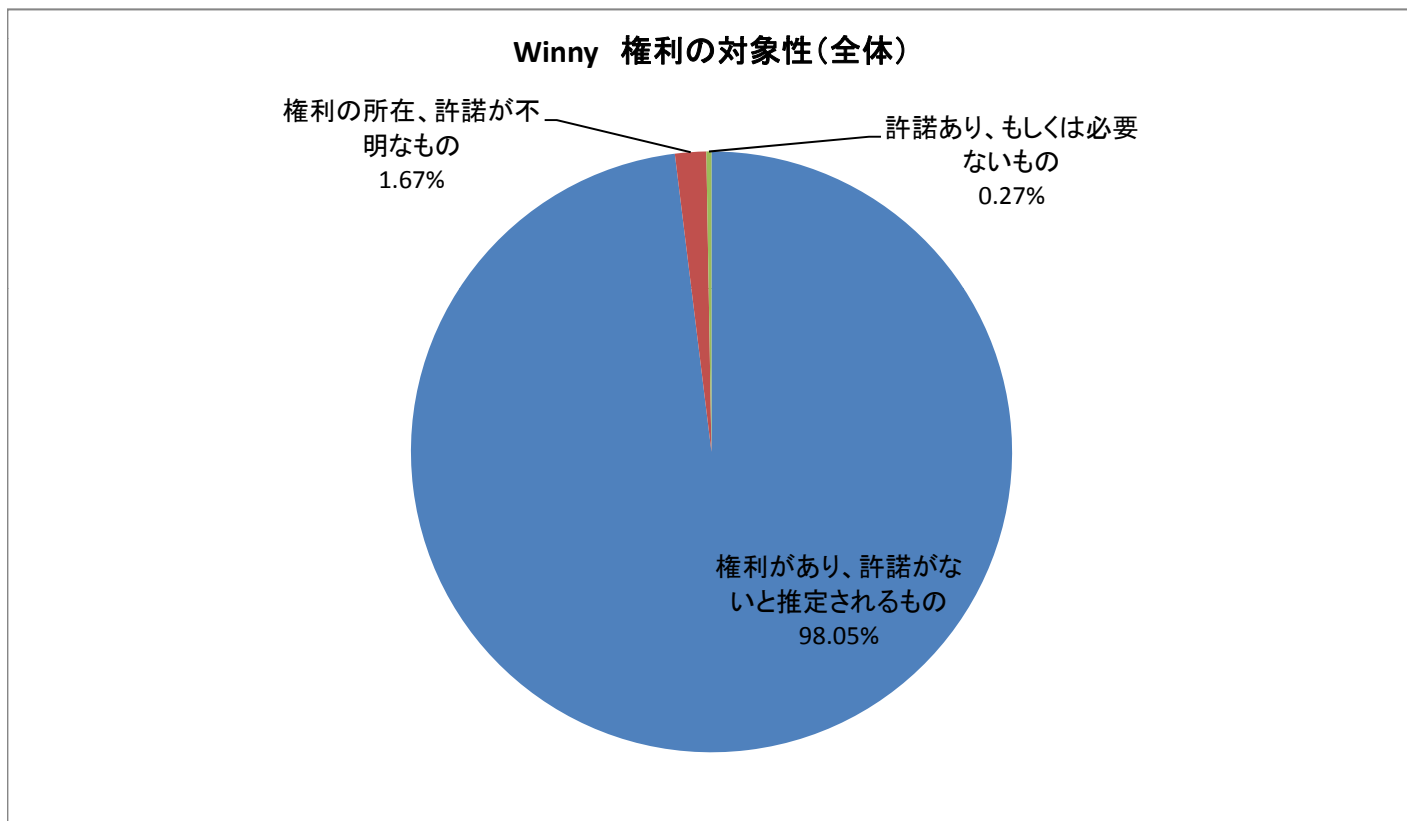


(2) 権利の対象性

①全体

著作物と推定されるもののうち、98.05%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	9,312
権利の所在、許諾が不明なもの	159
許諾あり、もしくは必要ないもの	26



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95%以上、その他の著作物については87%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	許諾あり
音楽	98.60% (1,480)	1.20% (18)	0.20% (3)
映像	99.72% (4,001)	0.27% (11)	0.00% (0)
プログラム	96.23% (511)	1.13% (6)	2.64% (14)
書籍	99.91% (2,439)	0.04% (1)	0.04% (1)
その他	87.05% (881)	12.15% (123)	0.79% (8)

2.Share

本調査でIPアドレスとポート番号の一意的組み合わせをノードの量として算出したところ、全数として約100,000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	9,313	音楽	703		
		映像	4,897	映画	83
				映画(海外)	232
				映像	426
				映像(海外)	134
				アニメ	3,949
				PV	73
		プログラム	368	アプリ	130
				ゲーム	238
		書籍	2,769	書籍	326
				コミック	2,443
		その他	576		
アダルト	6,670				
同人	3,336				
不明ファイル	519				
危険ファイル	0				
情報	162				
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

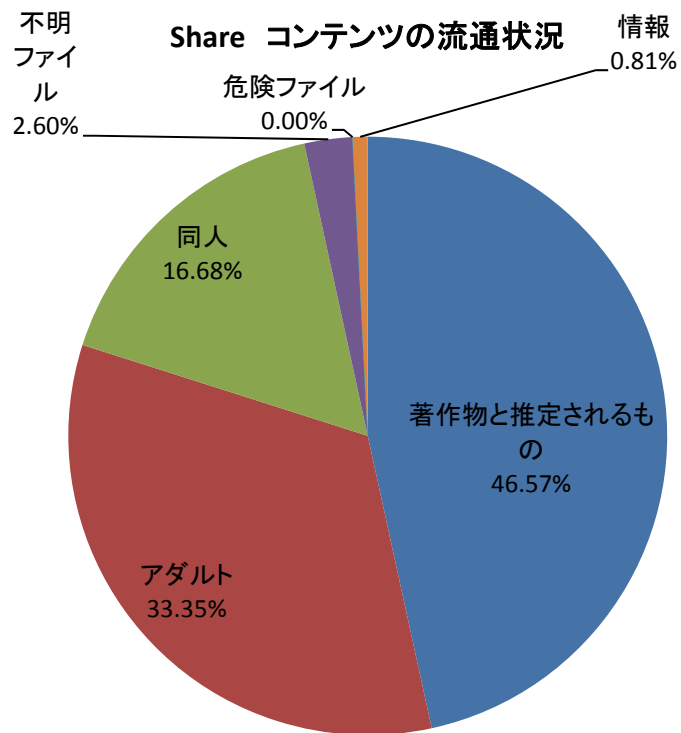
※「アダルト」、「同人」とは本調査で権利の所在が判別できないため、権利の対象に関する調査は見送ったもの

※「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

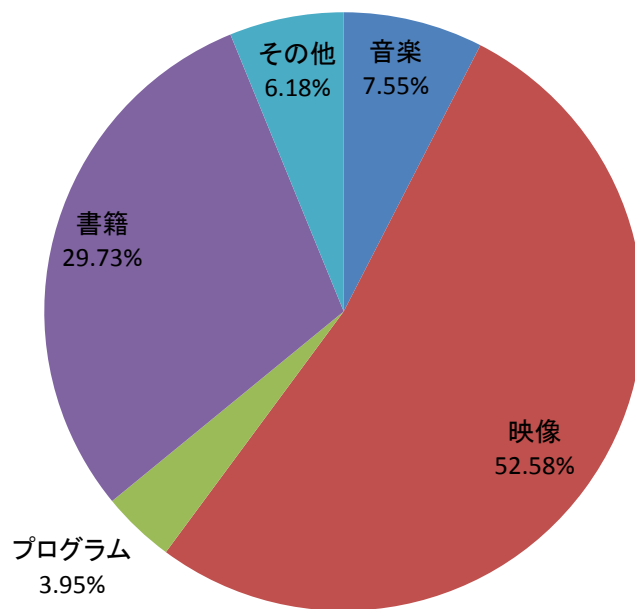
※「危険ファイル」とはタイトル、拡張子からウイルスなどと推定されるもの

※「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 著作権と推定されるコンテンツの内訳

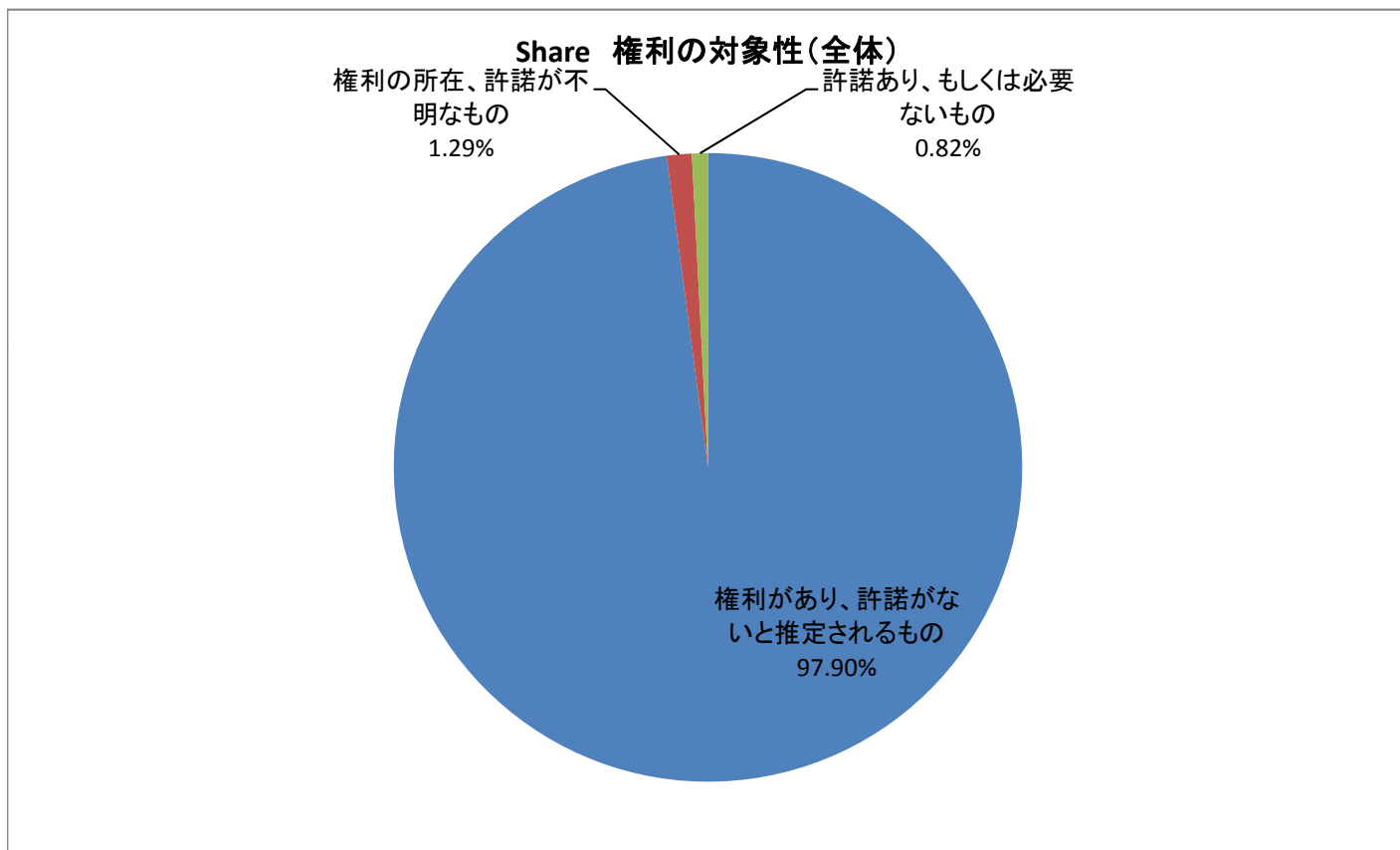


(2) 権利の対象性

① 全体

著作物と推定されるもののうち、97.90%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	9,117
権利の所在、許諾が不明なもの	120
許諾あり、もしくは必要ないもの	76



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97%以上、プログラム、その他の著作物については80%程度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	許諾あり
音楽	97.30% (684)	2.70% (19)	0.00% (0)
映像	99.80% (4,887)	0.20% (10)	0.00% (0)
プログラム	77.99% (287)	1.90% (7)	0.00% (0)
書籍	99.89% (2,766)	0.11% (3)	0.00% (0)
その他	85.59% (493)	14.06% (81)	0.35% (2)

3.PerfectDark

本調査でIPアドレスとポート番号の一意的組み合わせをノードの量として算出したところ、全数として約54,000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、約65%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	12,917	音楽	1,605			
		映像	6,743	映画		69
				映画(海外)		159
				映像		1,612
				映像(海外)		482
				アニメ		4,187
				PV		234
		プログラム	887	アプリ		313
				ゲーム		574
		書籍	2,480	書籍		578
コミック				1,902		
その他	1,202					
アダルト	4,420					
同人	2,233					
不明ファイル	406					
危険ファイル	0					
情報	24					
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

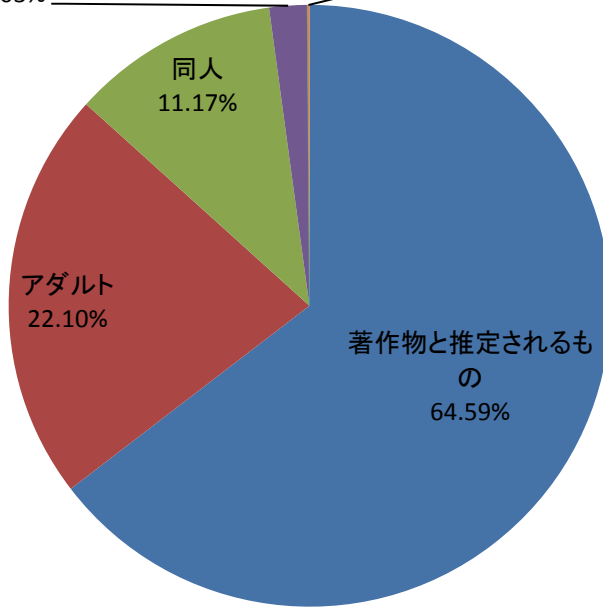
※「危険ファイル」とはタイトル、拡張子からウイルスなどと推定されるもの

※「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

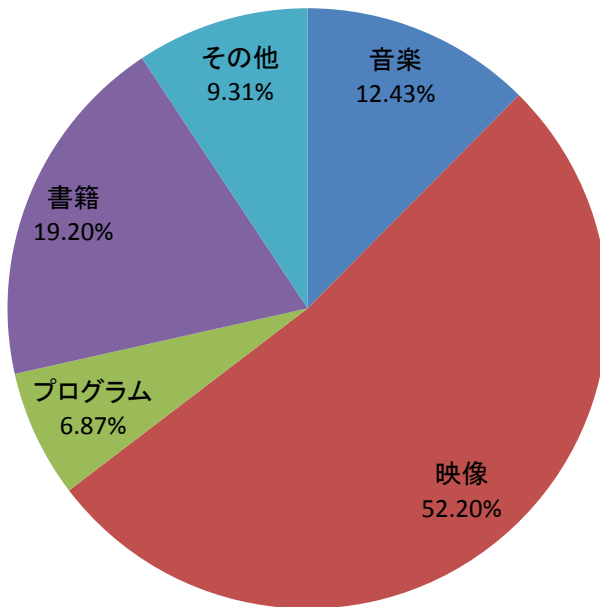
※「その他」とはコンテンツの分類が音楽、映像関連、プログラム、書籍関連に含まれないもの

PerfectDark コンテンツ流通状況

不明ファイル 2.03%
危険ファイル 0.00%
情報 0.12%



PerfectDark 著作権と推定されるコンテンツの内訳

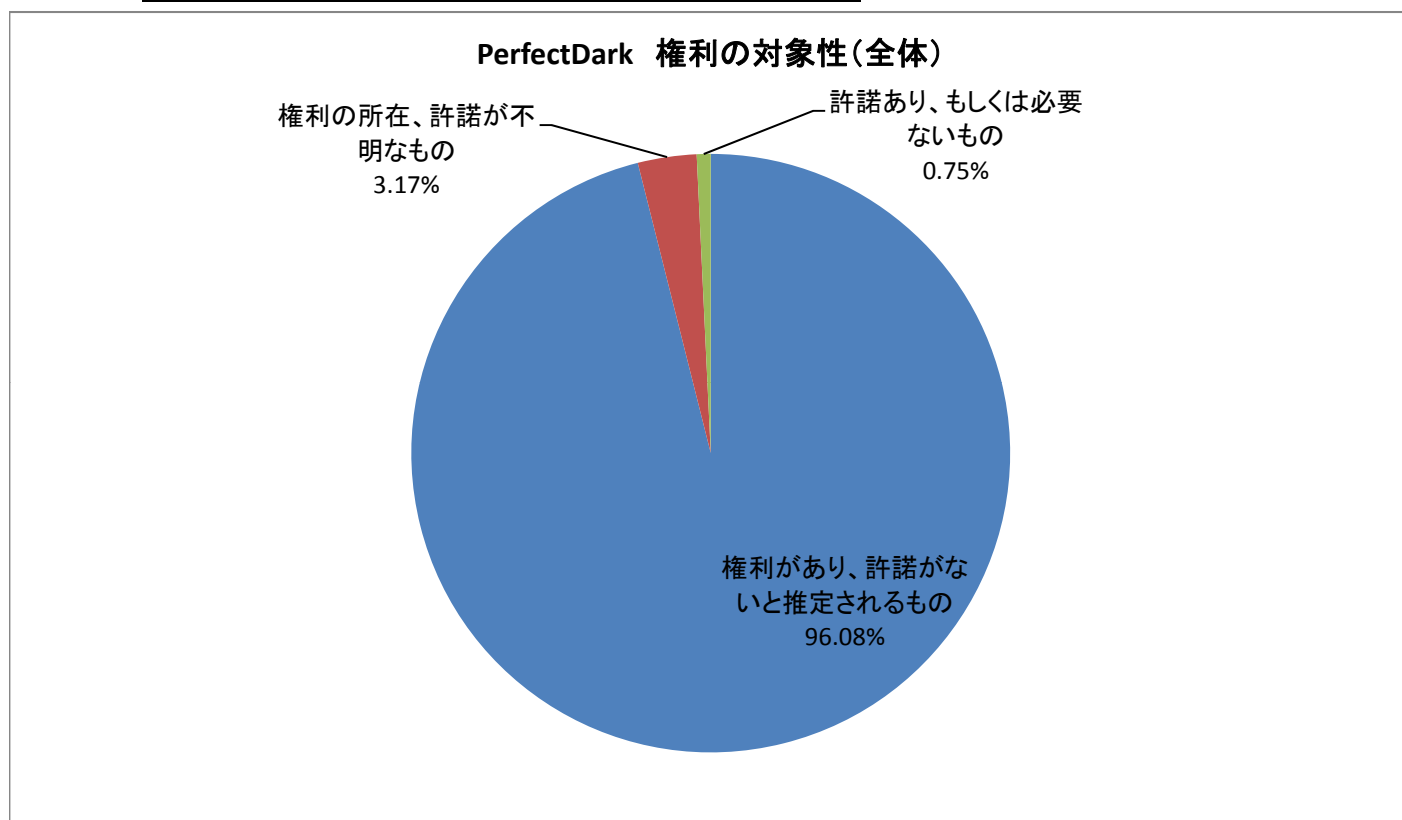


(2) 権利の対象性

①全体

著作物と推定されるもののうち、96.08%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	12,411
権利の所在、許諾が不明なもの	409
許諾あり、もしくは必要ないもの	97



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、その他の著作物については88%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	許諾あり
音楽	94.45% (1,516)	5.11% (82)	0.44% (7)
映像	97.42% (6,569)	2.57% (173)	0.01% (1)
プログラム	88.04% (781)	2.71% (24)	9.24% (82)
書籍	99.56% (2,469)	0.40% (10)	0.04% (1)
その他	89.52% (1,076)	9.98% (120)	0.50% (6)

Ⅲ. 第2回調査結果

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、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	9,288	音楽	1,468			
		映像	4,119	映画		46
				映画(海外)		94
				映像		622
				映像(海外)		94
				アニメ		3,102
				PV		161
		プログラム	378	アプリ		107
				ゲーム		271
		書籍	2,390	書籍		699
コミック				1,691		
その他	933					
アダルト	6,383					
同人	3,365					
不明ファイル	376					
危険ファイル	10					
情報	578					
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

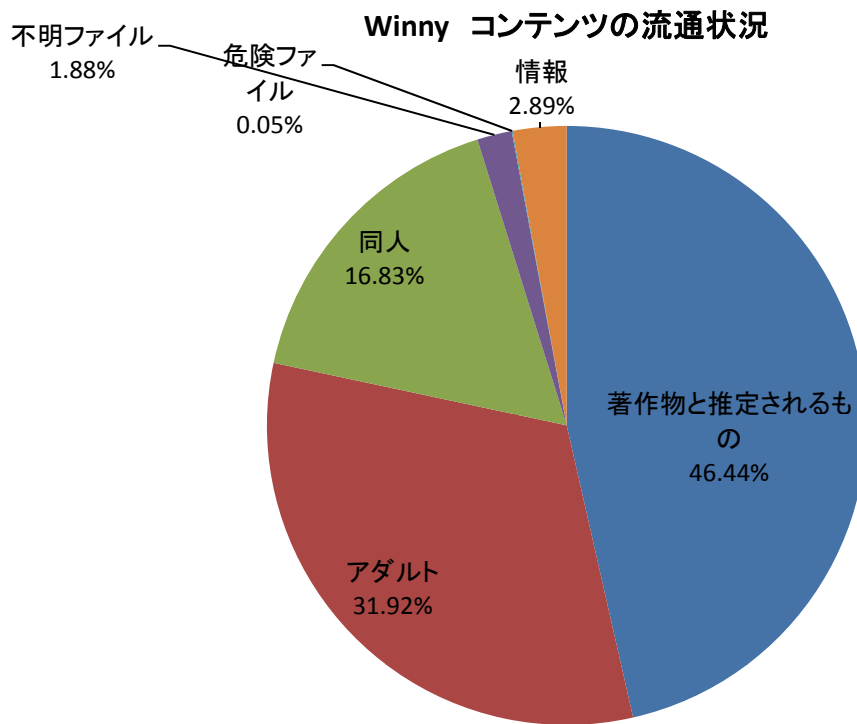
※「アダルト」、「同人」とは本調査で権利の所在が判別できないため、権利の対象に関する調査は見送ったもの

※「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

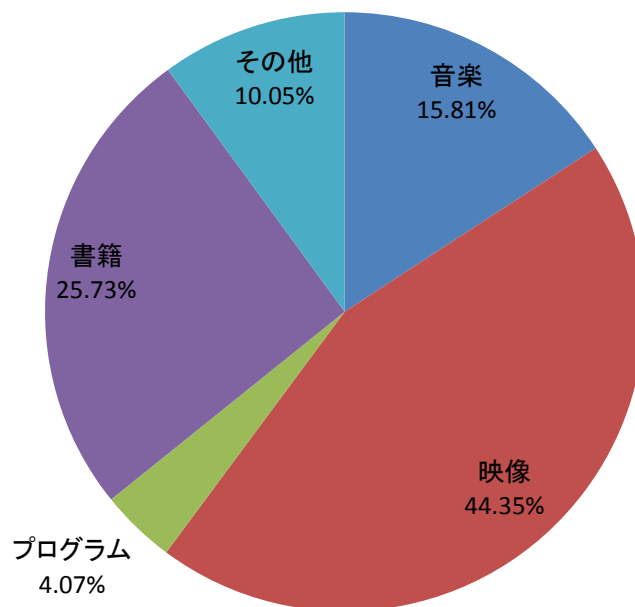
※「危険ファイル」とはタイトル、拡張子からウイルスなどと推定されるもの

※「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 著作権と推定されるコンテンツの内訳

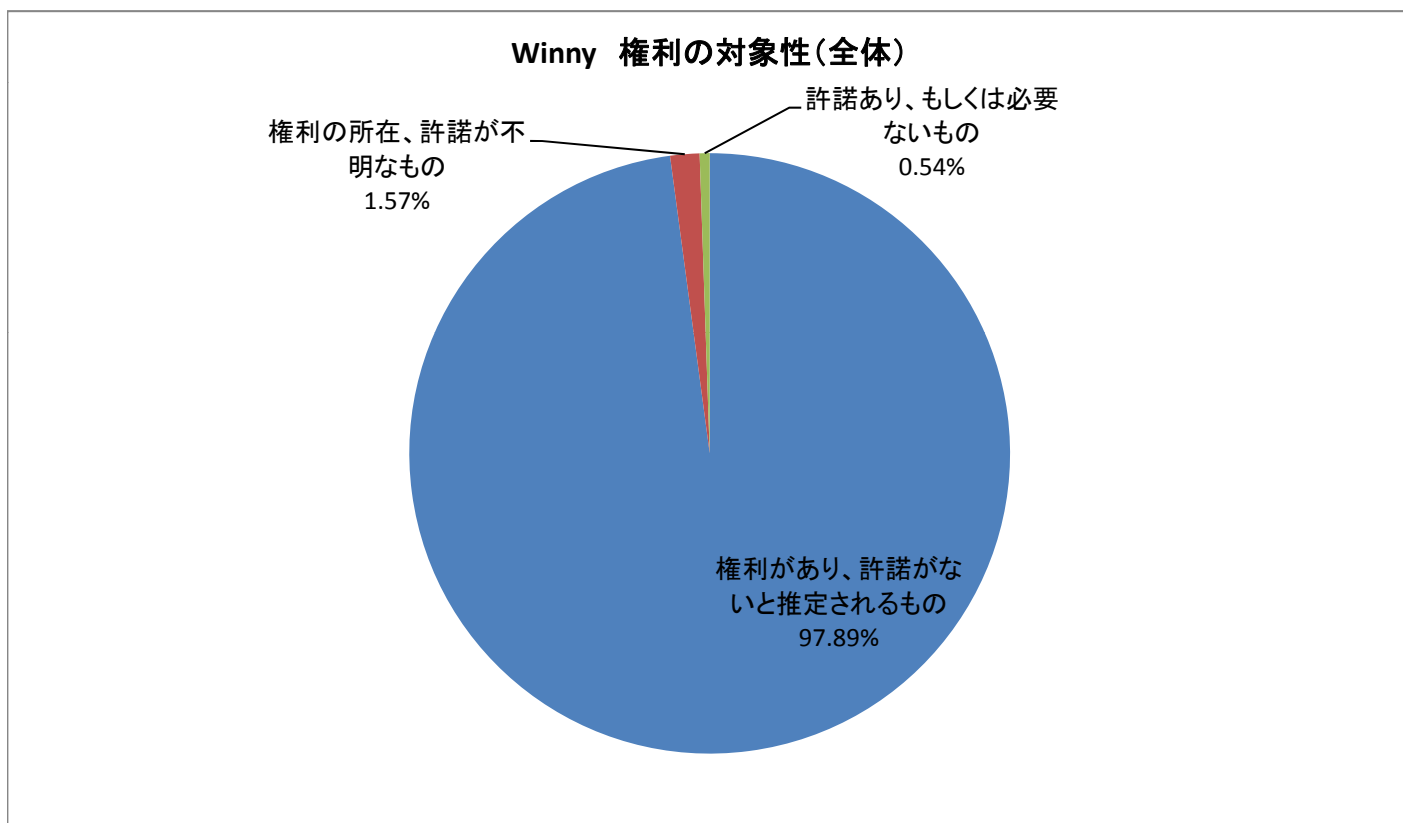


(2) 権利の対象性

① 全体

著作物と推定されるもののうち、97.89%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	9,092
権利の所在、許諾が不明なもの	146
許諾あり、もしくは必要ないもの	50



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	許諾あり
音楽	98.37%(1,444)	0.82%(12)	0.82%(12)
映像	99.66%(4,105)	0.29%(12)	0.05%(2)
プログラム	93.39%(353)	4.76%(18)	1.85%(7)
書籍	99.87%(2,387)	0.00%(0)	0.13%(3)
その他	86.07%(803)	11.15%(104)	2.79%(26)

2.Share

本調査でIPアドレスとポート番号の一意的組み合わせをノードの量として算出したところ、全数として約92,000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	10,748	音楽	843		
		映像	6,398	映画	136
				映画(海外)	234
				映像	878
				映像(海外)	434
				アニメ	4,423
				PV	293
		プログラム	308	アプリ	190
				ゲーム	118
		書籍	2,785	書籍	407
				コミック	2,378
		その他	414		
アダルト	5,563				
同人	2,781				
不明ファイル	792				
危険ファイル	7				
情報	109				
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

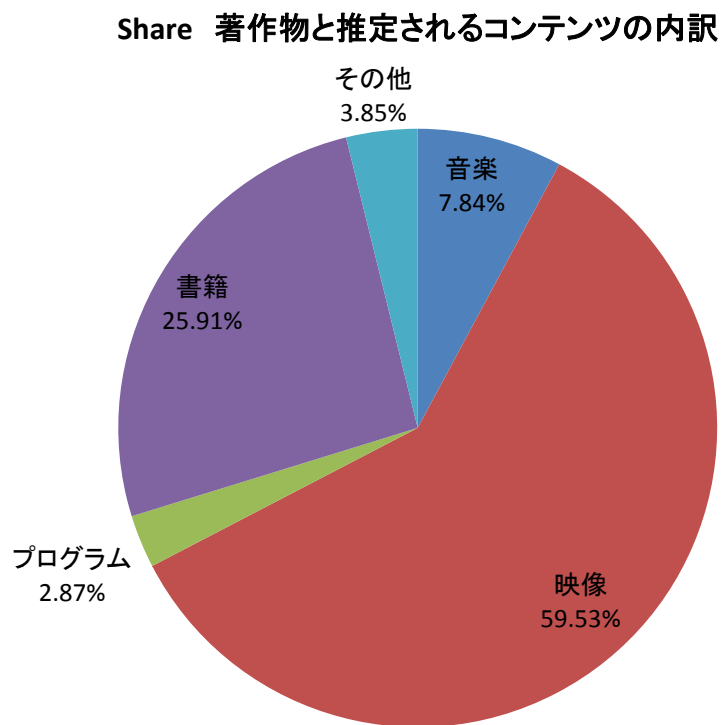
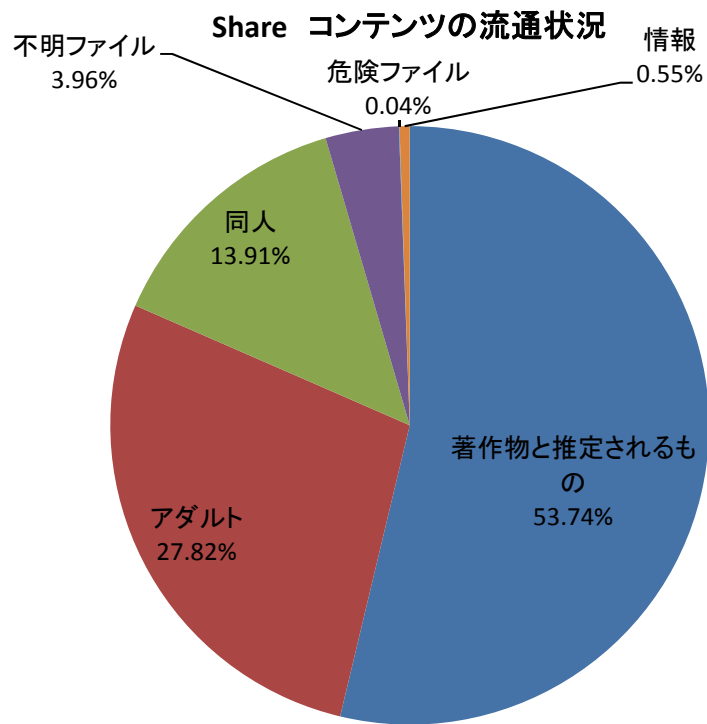
※「アダルト」、「同人」とは本調査で権利の所在が判別できないため、権利の対象に関する調査は見送ったもの

※「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

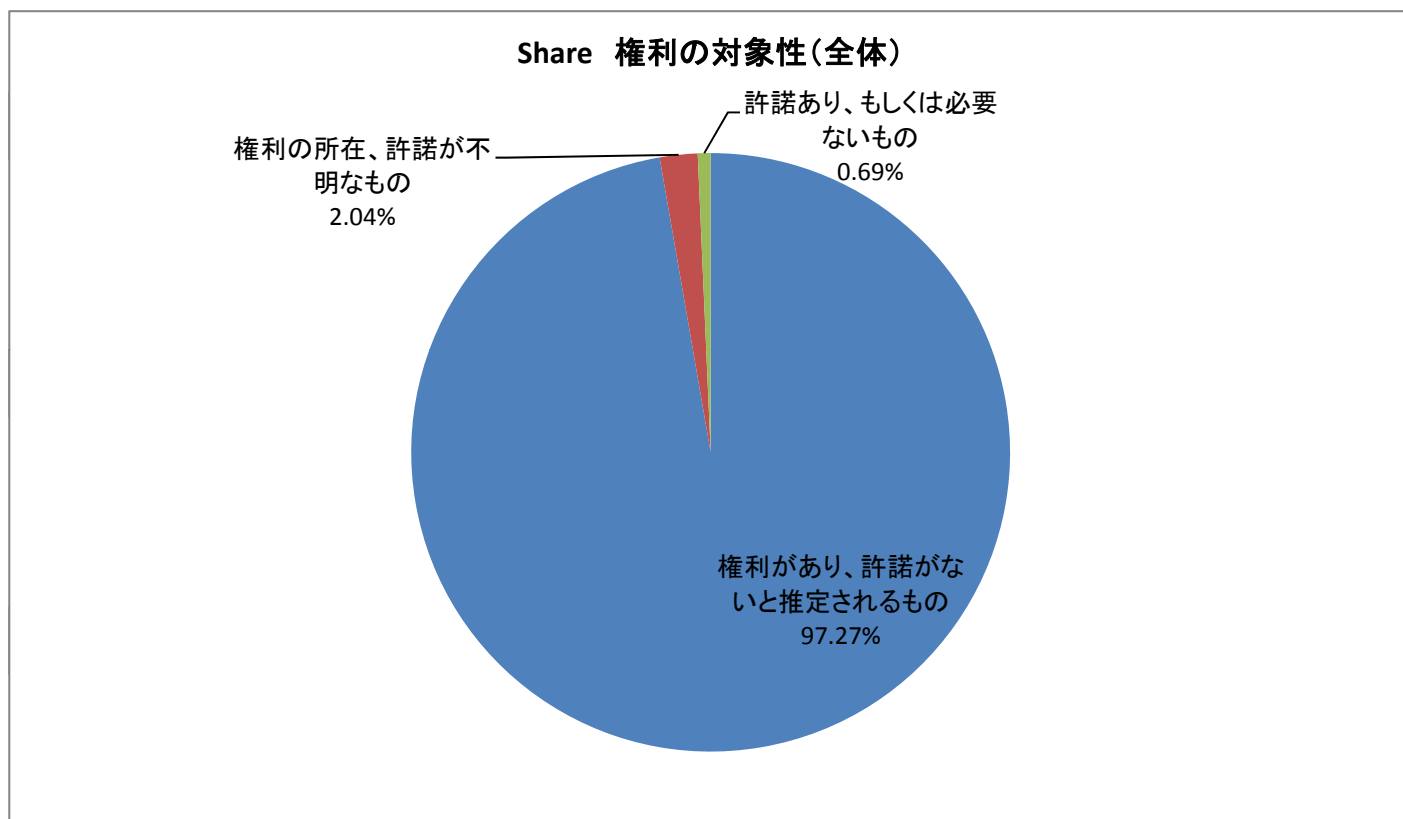


(2) 権利の対象性

①全体

著作物と推定されるもののうち、97.27%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	10,455
権利の所在、許諾が不明なもの	219
許諾あり、もしくは必要ないもの	74



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98%以上、プログラム、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	許諾あり
音楽	99.29%(837)	0.24%(2)	0.47%(4)
映像	98.87%(6,326)	1.13%(72)	0.00%(0)
プログラム	72.08%(222)	8.44%(26)	19.48%(60)
書籍	99.78%(2,779)	0.14%(4)	0.07%(2)
その他	70.29%(291)	27.78%(219)	1.93%(8)

3.PerfectDark

本調査でIPアドレスとポート番号の一意的組み合わせをノードの量として算出したところ、全数として約49,000ノードと推定される。

流通コンテンツのうち、約64%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	12,856	音楽	1,583		
	映像	6,911	映画		81
			映画(海外)		171
			映像		1,743
			映像(海外)		523
			アニメ		4,223
			PV		170
	プログラム	768	アプリ		295
			ゲーム		473
	書籍	2,480	書籍		536
コミック				1,944	
その他	1114				
アダルト	4,528				
同人	2,208				
不明ファイル	373				
危険ファイル	0				
情報	35				
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

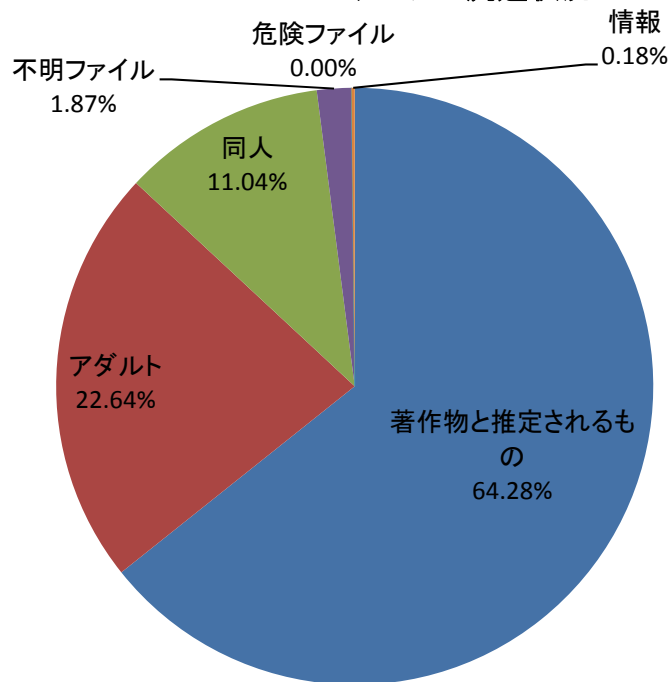
※「不明ファイル」とはタイトルからコンテンツの内容が判別できないもの

※「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

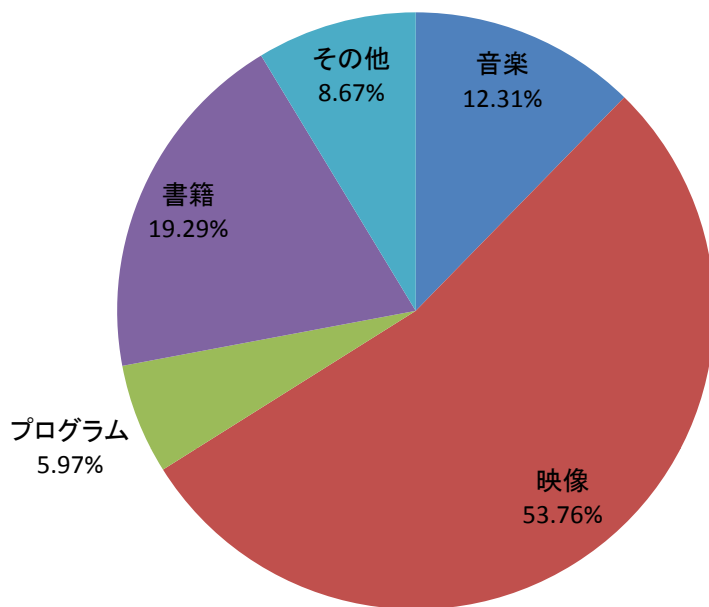
※「情報」とはウイルス感染などで流出した個人・組織等の情報だと推定されるもの

※「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

PerfectDark コンテンツの流通状況



PerfectDark 著作物と推定されるコンテンツの内訳

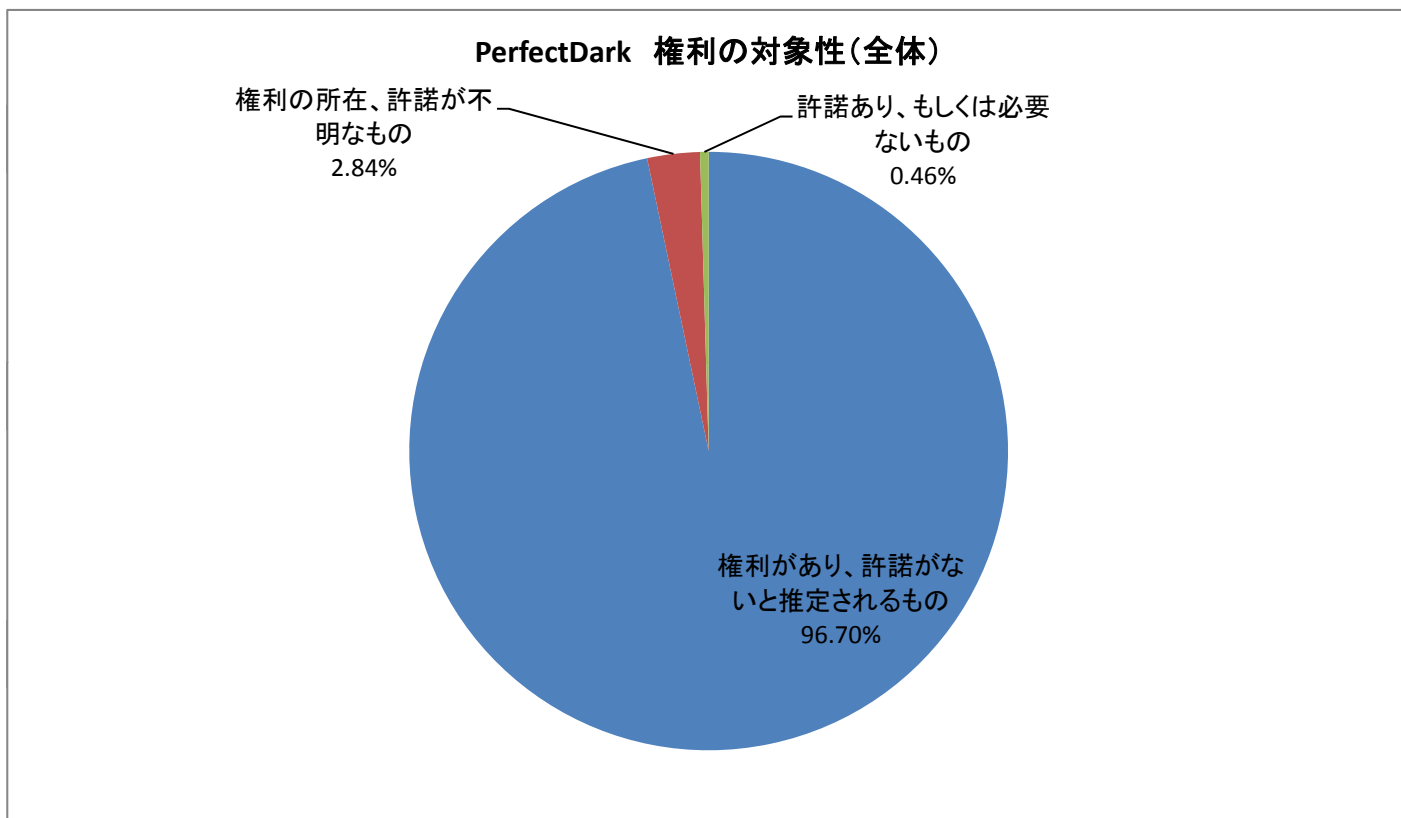


(2) 権利の対象性

①全体

著作物と推定されるもののうち、96.70%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	12,432
権利の所在、許諾が不明なもの	365
許諾あり、もしくは必要ないもの	59



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88%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	許諾あり
音楽	97.85%(1,549)	1.14%(18)	1.01%(16)
映像	97.18%(6,716)	2.82%(195)	0.00%(0)
プログラム	91.67%(704)	5.34%(41)	2.99%(23)
書籍	99.88%(2,477)	0.04%(1)	0.08%(2)
その他	88.51%(986)	9.87%(110)	1.62%(18)