2008－CSEC－42（18）
2008 / 7 / 24

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

寺田真敏[†1]　宮川雄一[†2]　松岡正明[†3]
松木隆宏[†3]　鬼頭哲郎[†1]　仲小路博史[†1]

[†1)](株)日立製作所　システム開発研究所
〒212-8567 神奈川県川崎市幸区鹿島田 890
[†2)](株)クロスワープ
〒150-0011 東京都渋谷区東 2-27-10
[†3)](株)ラック
〒105-7111 東京都港区東新橋 1-5-2 汐留シティセンター11 階

概要：P2P ファイル交換ソフトウェア環境において，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などの問題が深刻になっている．本稿では，P2P ファイル交換ソフトウェア環境において，マルウェア混入有無，著作権上の適切性などのファイル属性情報を登録するデータベースについて検討を行った．また，データベースに登録した P2P ファイル交換ソフトウェア環境で流通するコンテンツを利用し，コンテンツの流通実態を調査したので報告する．

キーワード：P2P，情報流出，マルウェア，著作権

Fesibility study of Database system for information sharing of P2P File Exchange Environment

Masato Terada[†1]　Yuichi Miyagawa[†2]　Masaaki Matsuoka[†3]
Takahiro Matsuki[†3]　Tetsuro Kito[†1]　Hirofumi Nakakoji[†1]

[†1)] System Development Lab. Hitachi Ltd.
890 Kashimada, Saiwai-ku, Kawasaki, Kanagawa, 212-8567 Japan
[†2)] CROSSWARP Inc.
2-27-10 Higashi, Shubuya-ku, Tokyo, 150-0011 Japan
[†3)] Little eArth Corporation Co., Ltd
1-5-2 Higashi-Shinbashi, Minato-ku,Tokyo, 105-7111 Japan

Abstract: Recently, there are many problems regarding the P2P file exchange environment on the Internet. The need to reconsider the current P2P file exchange environment for information leak and copyright is obvious. In this paper, we propose the database system (P2PDB) which includes the attribute of file, such as malware injection and copyright violation, to improve the present P2P file exchange environment. And, we show the investigation report for the trend of contents in the P2P file exchange environment by the database system (P2PDB).

Key words: P2P, Information Leak, Malware, Copyright

1　はじめに

P2P ファイル交換ソフトウェア環境において，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などの問題が深刻になっている．本研究では，これら課題を解決するため，P2P ファイル交換ソフトウェア環境において，意図しないファイルの流出を防ぎ，持ち込まれたくないファイルの流入を防ぐ情報流通対策アーキテクチャとそれをベースにした情報流通対策システムを提案している[1].

本稿では，情報流通対策アーキテクチャの機能部品のひとつである，マルウェア混入有無，著作権上の適切性などのファイル属性情報を登録するデータベースについて検討を行った．また，データベースに登録した P2P ファイル交換ソフトウェア環境で流通するコンテンツを利用し，コンテンツの流通実態を調査したので報告する．

2　関連研究

本章では，ファイル属性情報を登録するデータベースの関連研究として，流通ファイルの制御／把握の研究について整理する．

(1) 流通ファイルの制御

ファイルの流通を制御する方式として，文献 2)

は有害コンテンツの拡散を抑制するフィルタを共有する方式，文献 3)ではコンテンツにパーミッション情報を付与し，利用者のポリシーに従って配信を行う方式，文献 4)ではコンテンツと共に流通する著作権情報に記載された利用条件の違反，あるいはコンテンツの改ざんを検出した場合に，コンテンツの転送を制限する方式を提案している．

実フィールドでの流通ファイルの制御としては，著作権上適切ではないと思われるファイルのハッシュ値を Web サイトで掲載するという方法が取られている[5]．

(2) 流通ファイルの把握

文献 6) (調査時期：2006 年 6 月)では，約 18,000 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推計値は 176 万人で，音楽，映像などのファイル交換に利用されていると報告している．同じく文献 7) (調査時期：2007 年 9 月)では，クローリング手法を用いて収集したデータを元に Winny 稼動ノード数約 30 万台，流通ファイル数約 550 万，Share 稼動ノード約 21～22 万台，流通ファイル数約 65 万～70 万と推定している．また，約 2 万名を対象としたアンケート調査結果では，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 9.6%(2006 年 6 月の調査では 3.5%)となり，利用者増加を報告している．

文献 8)(調査時期：2003 年 4 月)では，WinMX，Gnutella，Winny を対象としたコンテンツ分析を行なっており，Winny では，全ファイル数が 23 万，ファイルサイズが約 63MB，avi や mpg などの動画像，zip や rar といった圧縮ファイルが流通していることを報告している．

以上のように，マルウェア混入有無，著作権上の適切性などのファイル属性情報を登録したデータベースの構築についての報告はこれまでにない．

3 情報流通対策アーキテクチャと情報流通対策システム

本章では，情報流通対策向けデータベース(以降，P2PDB)が前提とする情報流通対策アーキテクチャ／情報流通対策システムについて概説する．

情報流通対策アーキテクチャは，トラフィック／稼動ノード数／ファイル流通量の把握，意図しないファイル流出の防止，著作権上適切ではないファイル交換の抑止を統合的に推進するための枠組みであり，5 つの機能部品から構成される(図 1，表 1)．

情報流通対策システムは，情報流通対策アーキテクチャをベースとし，P2P ファイル交換ソフトウェアの利用を妨げることなく，インターネット利用者が安全に，安心して P2P ファイル交換ソフトウェアを利用できる環境を提供することを目的としたシステムである．また，情報流通対策システム自身は，(1) 意図しないファイルアップロードの防止機能，(2) 暴露ウイルス感染通知機能，(3) 著作権上適切ではないファイルのダウンロードの抑止機能，(4) 不正活動ホストの広報機能を有している．

本稿で提案する P2PDB は，情報流通対策アーキテクチャにおいて，ファイル属性情報を登録したデータベースとして位置付けられる．さらに，情報流通対策システムにおいては，(3) 著作権上適切ではないファイルのダウンロードの抑止機能を構成する部品であり，利用者端末側で稼動するダウンロードしたファイルの隔離と連携し，ファイル流入の抑止を実現する．なお，情報流通対策システムでの具体的な連携手法については，項番4.2において述べる．

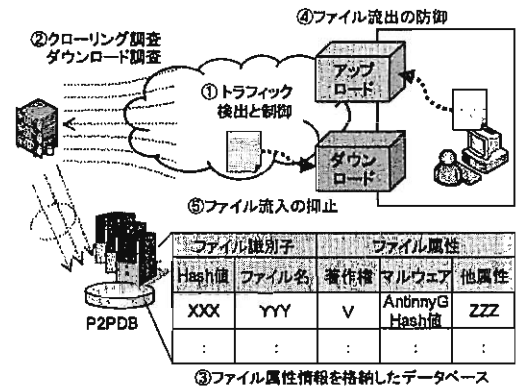


図 1：情報流通対策アーキテクチャ

表 1：情報流通対策アーキテクチャの機能部品

機能名	概要
①トラフィック検出と制御	ネットワーク側で，意図しないファイル流出や著作権上適切ではないファイル交換を検出し，必要に応じて遮断を行う．
②クローリング調査／ダウンロード調査	クローリング調査は P2P ファイル交換ソフトウェアが稼動するノードを網羅的に調査する方法である．ダウンロード調査は，P2P ファイル交換ソフトウェア環境で流通するコンテンツを収集し，マルウェア混入有無，著作権上の適切性を判定する．
③ファイル属性情報を登録したデータベース (P2PDB)	ダウンロード調査の結果として，ファイルを一意に識別する情報を，マルウェア混入有無，著作権上の適切性などのファイル属性情報と共に登録する．流出したファイルや適切ではないファイルのダウンロードを遮断するための基礎データとして利用する．
④ファイル流出の防止	端末側で意図しないファイル流出を検出し，必要に応じて遮断やインシデント発生の広報を行う．
⑤ファイル流入の抑止	端末側で著作権上適切ではないファイルダウンロードを検出し，必要に応じて遮断やインシデント発生の広報を行う．

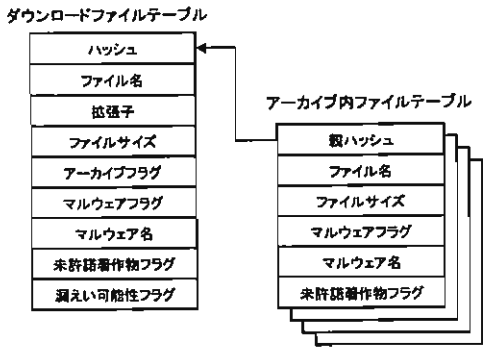
4 情報流通対策向けデータベース

本章では，情報流通対策アーキテクチャにおいて，ファイル属性情報を登録したデータベースとしての役割を果たす情報流通対策向けデータベース(P2PDB)の機能と開発中のプロトタイプシステムの仕様について述べる．提案する P2PDB の役割は，著作権上適切ではないファイルだけではなく，マルウェアの混入したファイルなどのダウンロードすべきではないファイルの流入を抑止することにある．このため，求められる機能としては，流通しているファイルの属性情報の収集と，利用者端末側へのファイルの属性情報の提供とが挙げられる．

4.1 流通しているファイルの属性情報の収集

(1) 機能概要

流通しているファイルの属性情報の収集は，著作権上適切ではないファイルやマルウェアの混入したファイルを識別するための情報を収集することであり，情報流通対策アーキテクチャの機能部品のひとつであるダウンロード調査と連携し実現する．具体的には，ダウンロード調査において，マルウェア混入有無や著作権上の適切性などの確認を行ない，その結果をファイルを一意に識別する情報と共に，マルウェア混入有無，著作権上の適切性などのファイル属性情報として P2PDB に登録する．



属性情報	概要
アーカイブフラグ	ダウンロードファイルがアーカイブファイルか否かを示す．
マルウェアフラグ	ダウンロードファイルにマルウェアを含むか否かを示す．
マルウェア名	ダウンロードファイル自身がマルウェアの場合に記載する．
未許諾著作物フラグ	ダウンロードファイルが著作物に該当するか否かを示す．
漏えい可能性フラグ	ダウンロードファイルが情報漏えいのファイルの可能性があるか否かを示す．

図 2：ファイル属性情報テーブル関連図

(2) プロトタイプシステムの仕様

P2PDB プロトタイプシステムでは，保持すべきファイル属性情報として，マルウェア／著作物の視点に加えて，複数ファイルを包含するアーカイブ型ファイルへの対応と，ファイル名に基づき情報漏えいファイルに該当する可能性を判定した結果とを登録可能な構成としている．図 2に P2PDB プロトタイプシステムのファイル属性情報テーブル関連図を示す．ダウンロードファイルテーブルでは，ダウンロードにより取得した P2P ファイル交換ソフトウェア環境上で流通している実ファイル情報を登録する．また，アーカイブ内ファイルテーブルでは，ダウンロードにより取得した実ファイルがアーカイブ型ファイルであった場合，アーカイブ内に含まれているファイル情報を登録する．

4.2 利用者端末側へのファイルの属性情報の提供

(1) 機能概要

利用者端末側へのファイルの属性情報の提供は，利用者端末側で稼動する情報流通対策システムの著作権上適切ではないファイルのダウンロードの抑止機能と連携し，著作権上適切ではないファイルの流入抑止を実現することにある．

ファイルのダウンロード操作が行われたときの動作概要を図 3に示す．利用者端末上の情報流通対策システムから P2PDB に問合せを行い，該当ファイルの属性情報が著作権上適切ではないファイルか否かを確認する．P2PDB から適切ではないとの結果が返信された場合には該当ファイルを隔離する．また，「その他の流通ファイル」については P2P ファイル交換ソフトウェアのダウンロードフォルダに流通可能なファイルとして登録する．

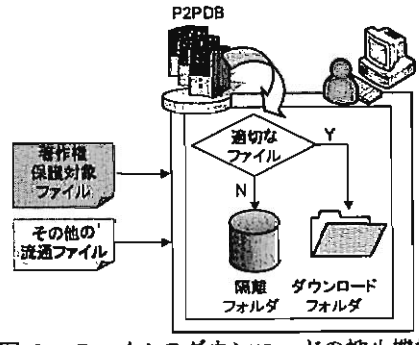


図 3：ファイルのダウンロードの抑止機能

商品名称等に関する表示

TRENDMICRO はトレンドマイクロ株式会社の登録商標です．本稿に記載されている会社名，製品名は，それぞれの会社の商標もしくは登録商標です．

```
HTTP リクエスト
  http://<server>/getfileinformation?q=<sha1_hash>

HTTP レスポンス
  <header>.....</header>
  <file>
  <hash>.....</hash>
  <mal_flg>0</mal_flg>
  <rights_flg>0</rights_flg>
  <leak_flg>1</leak_flg>
  </file>
```

図 4：Web API による問合せインタフェース例

(2) プロトタイプシステムの仕様

P2PDB プロトタイプシステムでは，Web API による問合せインタフェースを提供する．図 4はプロトタイプシステムで提供するインタフェース例である．情報流通対策システムが指定したファイルを一意に識別する情報(ファイルの SHA1 ハッシュ値)をキーに，該当するファイルの属性情報として，マルウェア，著作物，漏えい情報のそれぞれ該当するフラグを返答する．

5 情報流通対策向けデータベースを用いた調査

本章では，P2PDB プロトタイプシステムに登録したコンテンツを利用して調査した，ファイルの属性情報の対象となっているマルウェアと著作権上適切ではないファイルの流通状況について述べる．

5.1 調査方法

平成 20 年 1 月～2 月に，ダウンロード調査を通して得られた Winny で流通しているファイルを対象に 2 回の調査を実施した．調査対象としたファイル数は，1 回目：10,964 ファイル，2 回目：10,805 ファイルである．調査方法は，ダウンロード調査と詳細調査の 2 段階から構成した．

(1) ダウンロード調査

ダウンロード調査の流れは次の通りであり，マルウェアの混入有無や著作権上の適切性などの確認までを行なう．

(a) ファイルダウンロード

クローリング調査より得られた実ファイルの保持可能性が高いファイル所在情報に基づき，専用ツールを用いてダウンロードを実施する[a].

(b) マルウェア判定

ダウンロードファイルに対して，TRENDMICRO 社製ウイルス対策ソフトを用いてマルウェアの混入有無判定を行う．マルウェアを検出した場合にはマルウェアフラグを設定し P2PDB に記録する．

(c) 著作物判定

ダウンロードファイルに対して，マニュアル作業により著作権上適切ではないファイルか否かの判定を行う．未許諾著作物の場合には，未許諾著作物フラグを設定し P2PDB に記録する．なお，判定にあたっては，「プロバイダ責任制限法に基づく信頼性確認団体」である，社団法人コンピュータソフトウェア著作権協会に協力を依頼した．

(2) 詳細調査

詳細調査では，P2PDB に登録されたファイルを対象に，マルウェアの傾向ならびに著作権上適切ではないコンテンツの傾向を調査した．

5.2 調査結果

(1) 調査概要

第 1 回調査ではマルウェアを含むと判定されたファイルが 506 件であり，その中からマルウェアが 1,309 個検出された．第 2 回調査ではマルウェアが混入するファイル 550 件で 1,309 個のマルウェアが検出された．また，いずれの調査においても，約 6 割強が著作物と推測されるファイルであった．

表 2：第 1 回調査結果の概要

項目		件数	比率
マルウェアを含むファイル (検出したマルウェア数)		506 (1,309)	4.6%
	アイコン偽装を含む	468	
	ファイル名とアイコン偽装を含む	170	
著作物を含むと考えられるファイル		7,170	65.4%
調査対象ファイル総計		10,964	−

表 3：第 2 回調査結果の概要

項目		件数	比率
マルウェアを含むファイル (検出したマルウェア数)		550 (1,280)	5.0%
	アイコン偽装を含む	506	
	ファイル名とアイコン偽装を含む	195	
著作物を含むと考えられるファイル		6,786	62.8%
調査対象ファイル総計		10,805	−

(2) マルウェアの流通状況

本調査を通して得られた流通状況は次の通りである．

- マルウェアを含むと判定されたファイル 1,056 件のうちアイコン偽装を含むファイルは 943 件
- アイコン偽装を含むファイル 943 件のうちフォルダのアイコン偽装が 727 件(検出したマルウェアの 68.8%)(図 5)
- 検出したマルウェア 2,589 件のうち約 2 割弱が二重拡張子あり(461 件)，あるいは拡張子前に多量のスペースあり(496 件)のファイル名

a)仮想キーと呼ばれる「他ノードが保持するファイルを自ノードが保持しているように記述する情報」に基づきファイルのダウンロードを行なうと，Winny ノード間でファイルの中継動作が発生する．中継動作はファイルの複製を伴うため，結果として著作物やマルウェアの複製を助長してしまうことになる．情報流通対策アーキテクチャの実現にあたっては，クローリング調査により，実ファイル保持可能性が高いファイル所在情報を抽出した後，専用ツールでダウンロードする調査方法を採用することで，ファイルの中継動作を発生しないような運用方式を採用している．

- Winny 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(図 6)．なお，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．
- 検出したマルウェア 2,589 件のうち約 7 割が情報漏えいを目的とした Antinny ウイルスならびにその亜種である(図 7)．

(3) 著作権上適切ではないファイルの流通状況

本調査を通して得られた流通状況は次の通りである．

- 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(図 8)．
- 著作物としては，映像系ファイル(64%)が多く流通している(図 9)．また，映像系ファイルは，ほぼ全てが権利があり，かつ許諾が無いと推測されるファイルである(図 10)．

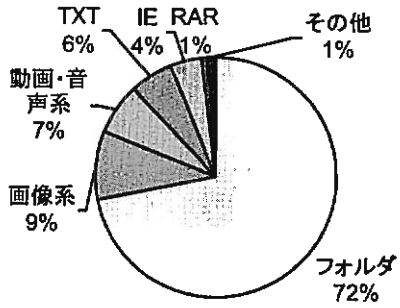


図 5：アイコン偽造の内訳

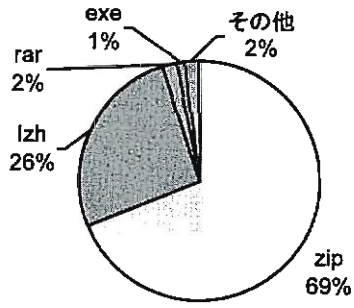


図 6：マルウェアを含むコンテンツの拡張子内訳

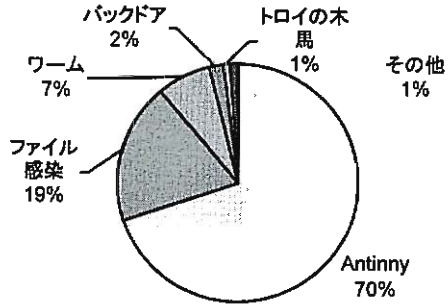


図 7：既知マルウェアの内訳

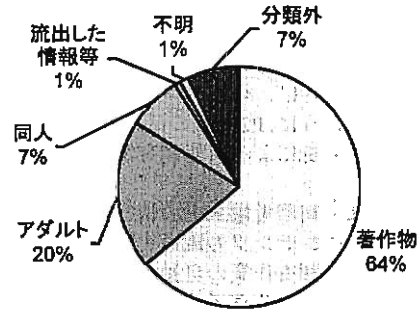


図 8：コンテンツの内訳

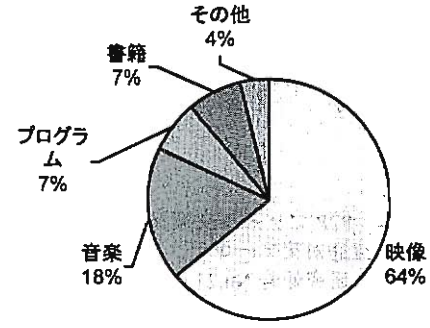


図 9：著作物の内訳

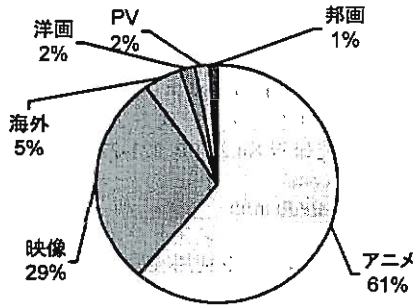


図 10：映像系ファイルの内訳

6 おわりに

本稿では，P2P ファイル交換ソフトウェア環境において，マルウェア混入有無，著作権上の適切性などのファイル属性情報を登録するデータベース P2PDB について報告した．さらに，開発中のデータベースに登録した P2P ファイル交換ソフトウェア環境で流通するコンテンツを利用し，コンテンツの流通実態調査として，マルウェアの流通状況と著作権上適切ではないファイルの流通状況について報告した．

提案する P2PDB の役割は，著作権上適切ではないファイルやマルウェアの混入したファイルなどのダウンロードすべきではないファイルの流入を抑止することにある．さらに，今回のコンテンツの流通実態調査で示したように，P2PDB の活用方法としては，登録した情報の分析による実態調査が有効であることを示した．

今後の課題は，利用者端末上の情報流通対策システムとの連携，コンテンツの流通実態調査を継続的に実施するための調査作業の自動化，P2PDB の継続運用を実現するための管理手法などがが挙げられる．

謝辞

本研究は総務省から受託した「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」の成果の一部です．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) 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アーキテクチャの検討，情報処理学会 CSEC 研究報告 No.21 pp.243-248(2008 年 3 月)
2) 伊吹和也 他：フィルタ共有による P2P ネットワーク上の有害コンテンツ拡散抑制，情報処理学会 DSM 研究報告 No.72 pp.7-12 (2007 年 7 月)
3) 今本吉治 他：セキュア P2P のためのユーザ主導型コンテンツ交換方式，情報処理学会 DPS 研究報告 No.22 pp.7-12(2004 年 3 月)
4) 齋藤武比古 他：モバイルエージェントを活用した P2P 環境における著作権管理方法の提案、情報処理学会 DPS 研究報告 No.26 PP. 221-226 (2006 年 3 月)
5) http://hashdb.com/
http://www.sharedb.info/
6) 社団法人コンピュータソフトウェア著作権協会：「ファイル交換ソフト利用実態調査」結果発表、(2006 年 7 月 25 日),
http://www2.accsjp.or.jp/topics/release1.html
7) 社団法人コンピュータソフトウェア著作権協会：第 6 回「ファイル交換ソフト利用実態調査」、(2007 年 12 月 21 日)、
http://www2.accsjp.or.jp/topics/release_fs.html
8) 大井恵太 他：P2P ファイル共有におけるコンテンツ分析，情報処理学会 DPS 研究報告 No.87 pp.17-24(2003 年 8 月)