2007－CSEC－37 (9)
2007／5／25

クローリング手法を用いたP2Pネットワークの観測

寺田真敏[†1] 鵜飼裕司[†2] 金居良治[†2]
畑田充弘[†3] 松木隆宏[†4] 宮川雄一[†5]

[†1)] (株)日立製作所 システム開発研究所
[†2)] eEye Digital Security
[†3)] NTT コミュニケーションズ株式会社
[†4)] 株式会社ラック
[†5)] 株式会社クロスワープ

概要：P2P(Peer to Peer)ファイル交換ソフトウェア利用が広がる中，その利用実態に関する調査報告は数少ない．本調査研究の目的は，P2Pファイル交換ソフトウェア環境の利用実態に関する具体的な調査方法やその方法に基づき調査した結果を示すことで，調査手法の有効性を示すと共に，P2Pファイル交換ソフトウェア環境に関する調査活動を支援することにある．本稿では，P2Pファイル交換ソフトウェア環境の調査研究事例として，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クローリング手法を用いて収集したデータを元にWinny／Share稼動ノード数，ファイル数，ファイル保持ノードの特定に関する観測結果について述べる．
キーワード：P2P，クローリング，Winny，Share

P2P network observation using crawling method.

Masato Terada[†1] Yuji Ukai[†2] Ryoji Kanai[†2]
Mitsuhiro Hatada[†3] Takahiro Matsuki[†4] Yuichi Miyagawa[†5]

Abstract: P2P(Peer to Peer) file exchange software spreads out, the requirements of investigation reports about P2P network are increasing such as node counts, file counts and security condition. In this paper, we describe crawling method for P2P network observation to elucidate node counts, file counts and so on. Our proposal is to collect information of various node and key from P2P network such as Winny and Share via node by node. Also we show the validity of our approach by implemented crawling system.
Key words: P2P, crawling, Winny, Share

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，その利用実態に関する調査報告は数少ない．本調査研究の目的は，P2Pファイル交換ソフトウェア環境の利用実態に関する具体的な調査方法を示すこと，その方法に基づき調査した結果を示すことで，調査手法の有効性を示すと共に，P2Pファイル交換ソフトウェア環境の利用実態に関する調査活動を支援することにある．報告者らは，これまでWinnybot[1]と呼ぶツールを用いて収集したデータを元にWinny稼動ノード数，ファイルのプロパティが格納されているWinnyのキー流通量について調査を実施した[2]．本稿では，Winnybotで採用しているクローリング手法を用いた稼動ノード数・ファイル流通量に関する観測とファイル保持ノードの特定に関する観測手法を述べると共に，クローリング手法で得られたWinny／Shareネットワークの観測データの比較結果を報告する．

2 関連研究

(1) P2Pファイル交換ソフトウェア環境の稼動ノード数・ファイル流通量に関する観測

Winnyネットワークの稼動ノード数については，文献 3) (調査時期：2006年8月)において，“平日で39万から多い日では41万，土日になると43万から44万以上のノード数”であると報告している．また，Shareネットワークについては，文献 4) (調査時期：2006年12月)において，“1日あたり10万から15万ノード”であると報告している．

[†1)] System Development Lab. Hitachi Ltd.
890 Kashimada, Saiwai-ku, Kawasaki, Kanagawa, 212-8567 Japan
[†2)] eEye Digital Security
1 Columbia, Aliso Viejo, California 92656, United States
[†3)] NTT Communicatons Corporation
21F Tokyo Opera City Tower
3-20-2 Nishi-Shinjuku, Shinjuku-ku, Tokyo, 163-1421 Japan
[†4)] Little eArth Corporation Co., Ltd
I-5-2 Higashi-Shinbashi, Minato,Tokyo, 105-7111 Japan.
[†5)] CROSSWARP Inc.
2-27-10 Higashi, Shubuya, Tokyo, 150-0011 Japan

(2) 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観測

ファイル保持ノードの特定に関しては，安全で効果的な著作権侵害監視の手法のひとつとして，ユーザがキャッシュ，アップロードに関わらずファイルを保持しているかどうかを特定する方法とその検証結果を報告している[5]．この報告では，Winny ネットワークを網羅的に観測した上で特定のハッシュ値の検知回数が多いノードを抽出する事で，実ファイルを保持しているノードを絞り込む手法は有効であることを指摘している．

3 クローリング手法による P2P ネットワークの観測

本章では，調査にあたり利用したクローリング手法について述べる．

3.1 用語定義

クローリング手法を説明するにあたり，本稿で使用する用語を説明する．

- ノード
  IP アドレスを用いた Winny／Share ノードの識別子であり，ノード数算出に使用する．
- プライマリキー
  “IP アドレス＋ポート番号＋ファイルのハッシュ値”から構成したキーの識別子であり，Winny／Share ノードを加味したファイル識別子として，ファイル保持ノードの特定に使用する．
- ハッシュキー
  ファイルのハッシュ値を用いた識別子であり，本稿では重複のないファイル数，すなわち，一意なオリジナルファイル数の算出に使用する．
- 仮想キー
  仮想キーは Winny で定義された用語であり，キーの拡散や検索クエリによって取得したキーを Winny ノード内に持っているが，対応するキャッシュファイル（キャッシュブロック保有率が 0%）を持たないキーのことである．

3.2 クローリング手法

P2P モデルで構成されたファイル交換システムには，Napster[6]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 1)．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，ファイル保持ノードの特定，すなわち，ノードの IP アドレスを特定することができる．

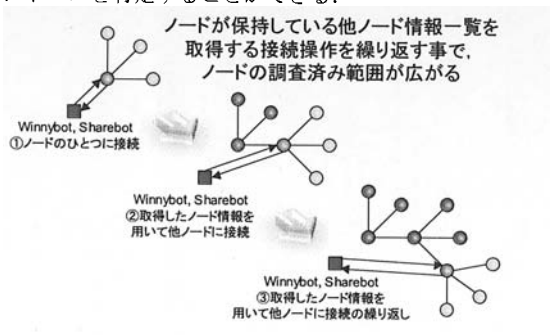

図 1：クローリング手法の概要

3.3 Winny を対象としたクローリング

Winnybot はクローリング手法を実装した Winny ネットワーク用観測ツールであり，図 2に示すモジュールから構成されている．観測エンジンは，Winny ネットワーク上のいずれかのノードに接続した後，拡散クエリ送信要求(コマンド 0x0A)を送信し，クエリ送信(コマンド 0x0D)を受信する操作を繰り返すことで，Winny 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ノード／キー情報を網羅的に収集する．また，観測エンジンが収集したキー情報を定期的にデータベースに格納する．
Winnybot で観測可能な項目を表 1，表 2に示す．

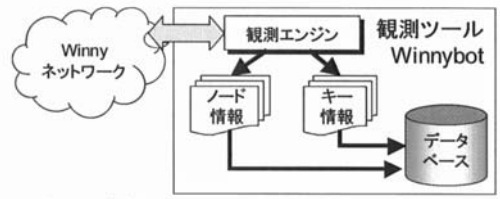

図 2：観測ツール Winnybot のモジュール構成

表 1：キー情報として観測可能な項目

項目
キーの IP アドレス
キーのポート番号
BBS スレッド管理ノードの IP アドレス
BBS スレッド管理ノードのポート番号
ファイルサイズ
キー更新日時
被参照ブロック数
キー消滅判定タイマー
ファイルのハッシュ値
トリップ情報

表 2：ノード情報として観測可能な項目

項目
IP アドレス
ポート番号
回線速度 (K バイト/秒)
クラスターワードの md5 値

3.4 Share を対象としたクローリング

Sharebot[7][8]はクローリング手法を実装した Share ネットワーク用観測ツールであり，図 3に示すモジュールから構成されている．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．Sharebot で観測可能な項目を表 3に示す．

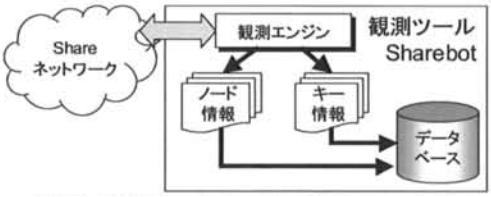

図 3：観測ツール Sharebot のモジュール構成

表 3：キー情報として観測可能な項目

項目
キーの IP アドレス
キーのポート番号
キーの更新日時
ファイルのハッシュ値
ファイルの名称

4 観測結果

本章では，前述の観測環境を用いて実施した稼動ノード数，ファイル数に関する観測と，ファイル保持ノードの特定に関する観測について述べる．なお，本調査は，社団法人コンピュータソフトウェア著作権協会から提供された観測データを元に実施した．

4.1 Winny ネットワークの調査

(1) 観測期間

2006 年 10 月 6 日(金)～10 月 15 日(日)の 10 日間

(2) 観測装置

- 観測装置台数：9 台
- 観測データ数：平均 7,100 万件／日，計約 7 億件のキー情報を収集(図 4).

(3) 流通量に関する観測結果

(a) 推定方法

Winny ノードの稼動数ならびにファイルの流通量算出にあたっては，単位時間に観測装置台数の増加に伴って発生する一意な観測件数の収束状況を用いて算出した[2]．ここで単位時間とは観測件数の集約期間である．単位時間 24 時間とした Winny ノードの稼動数算出を例にとると，24 時間の間に観測したデータ全体を対象に，観測装置毎に観測した一意なノード数を足し合わせた総計(以降，単純総計)と観測装置全体で観測した一意なノード数の累積(以降，累積総計)からノードの一意な観測件数の収束状況を得て，その収束点を Winny ノードの稼動数とする．すなわち，単純総計に対して累積総数が収束した件数を Winny ノードの稼動数とする(図 5).

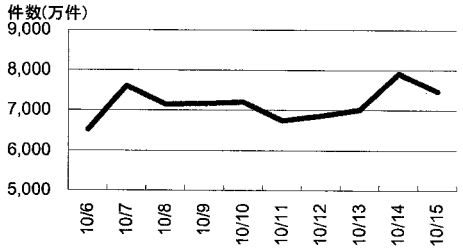

図 4：観測期間中のキー観測総数[Winny]

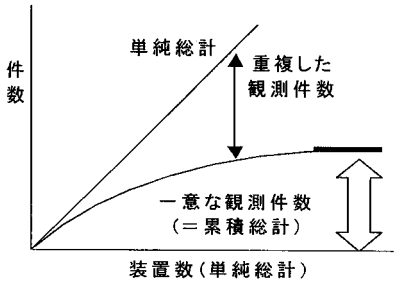

図 5：単純総計と累積総計とを用いた収束判定

(b) Winny ノードの稼動数

表 1の“キーの IP アドレス”を観測項目とし，24 時間を単位時間とした場合，累積総計の収束状況から Winny ノード稼動数は約 35 万ノードとなる(図 6).

(c) 一意なオリジナルファイル数

表 1の“ファイルのハッシュ値”を観測項目とし，24 時間を単位時間とした場合，累積総計の収束状況から Winny ネットワーク上に存在する一意なオリジナルファイル数は約 450 万ファイルとなる(図 7).

(d) 拡張子別のファイル数

“ファイルのハッシュ値”に対応するファイル名からファイル拡張子を抽出した結果，一意なオリジナルファイル件数のうち，拡張子 JPG (24%)，ZIP (20%)，MP3(18%)，AVI(18%)が約 80%を占める．

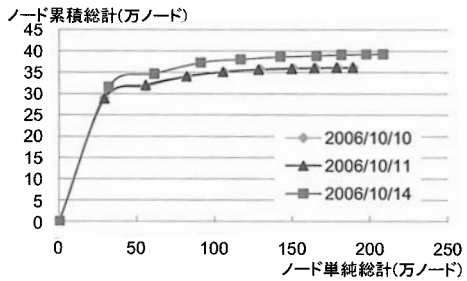

図 6：ノードの累積総計(24 時間)[Winny]

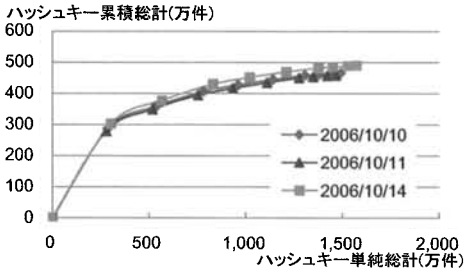

図 7：ハッシュキーの累積総計(24 時間)[Winny]

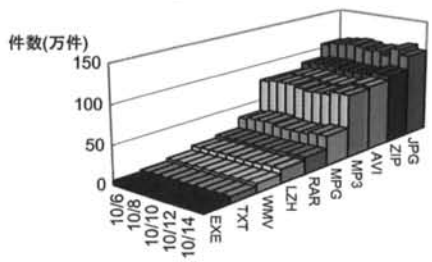

図 8：拡張子別ファイル件数[Winny]

(4) ファイル保持ノードの特定に関する観測

(a) 仮想キーの出現頻度

ファイル保持ノードの特定に関する観測では，参照ノードとして用意した Winny ノードを対象に調査を行なった．調査期間中に，参照ノードに関係する仮想キーは 1,047 件[*1]観測され，そのうち 97%が単位時間 24 時間の出現頻度は 1 回であった(図 9)．隣接時間帯とは，4 時ならびに 5 時という隣り合った時間帯に観測されたことを意味し，この観測件数も含めると 99%が出現頻度 1 回である．

(b) 特定ファイルハッシュ値のキー出現頻度

参照ノードが関係する特定ファイルハッシュ値のキーの出現頻度を図 10に示す．

4.2 Share ネットワークの調査

(1) 観測期間

2007 年 2 月 21 日(水)～4 月 9 日(月)の 48 日間

(2) 観測装置

- 観測装置台数：10 台
- 観測データ数：平均 560 万件／日，計約 3 億件のキー情報を収集（図 11）．

(3) 参照ノードの設定

ファイル保持ノードの特定に関する観測では，参照ノードとして用意した Share ノードを対象に調査を行なった．

- 低速な通信速度(50KByte/s)設定での稼動期間
  2007 年 2 月 21 日から 4 月 3 日
- 高速な通信速度(500KByte/s)設定での稼動期間
  2007 年 4 月 4 日から 4 月 9 日

*1) 同一時間帯（一時間以内）に同一の仮想キーが複数観測された場合には，一件とカウントしている．

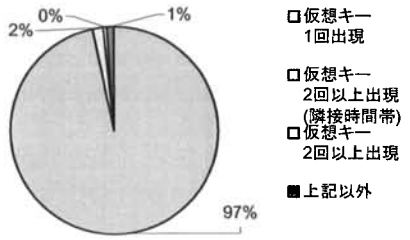

図 9：参照ノードが関係する仮想キーの出現頻度

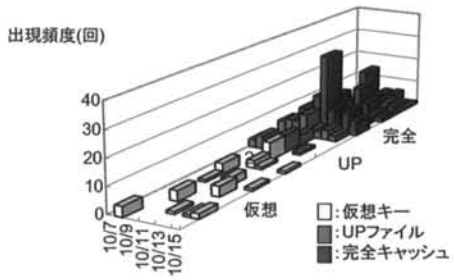

図 10：参照ノードが関係するキー出現頻度[Winny]

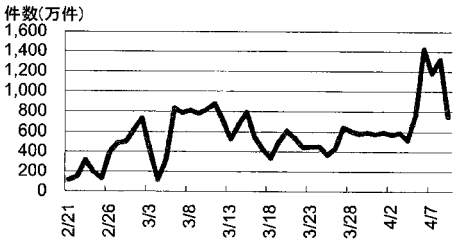

図 11：観測期間中のキー観測総数[Share]

(3) 流通量に関する観測

(a) 推定方法

Winny ネットワークの調査で用いた推定方法を用いる．

(b) Share ノードの稼動数

表 3の“キーの IP アドレス”を観測項目とし，24 時間を単位時間とした場合，Share ノードの稼動数は約 15 万ノードとなる(図 12)．

(c) 一意なオリジナルファイル数

表 3の“ファイルのハッシュ値”を観測項目とし，24 時間を単位時間とした場合，Share ネットワーク上に存在する一意なオリジナルファイル数については約 40 万ファイルとなる(図 13)．

(d) 拡張子別のファイル数

“ファイルのハッシュ値”に対応するファイル名からファイル拡張子を抽出した結果，一意なオリジナルファイルのうち，拡張子 ZIP(23%)，AVI(19%)，JPG(14%)が約 60%占める(図 14)．

(4) ファイル保持ノードの特定に関する観測

ファイル保持ノードの特定に関する観測では，

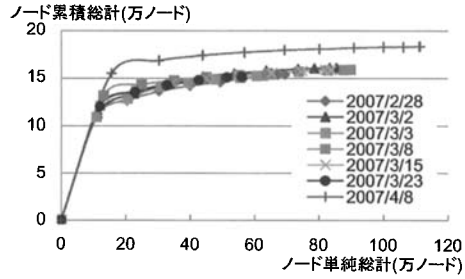

図 12：ノードの累積総計(24 時間)[Share]

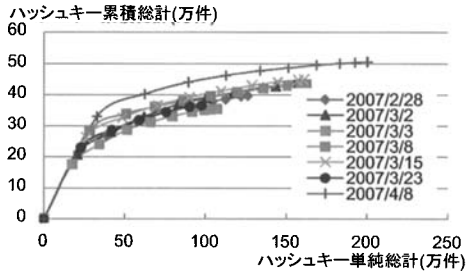

図 13：ハッシュキーの累積総計(24 時間)[Share]

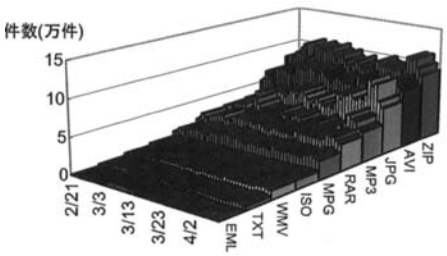

図 14：拡張子別のファイル数[Share]

表 4：キャッシュタイプ毎の出現頻度[Share]

出現頻度	UP	完全	部分	拡散完全	拡散部分	非保持
1				11	332	29
2		2		5	104	11
3				3	28	3
4					13	
5					7	
6	1				1	2
7					1	
8						
9						1
≧10	6	7				1
計	7	9	0	19	486	47

分類		説明
UP		アップロードファイルのキャッシュ
完全		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		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)	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
	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がキーを観測した)

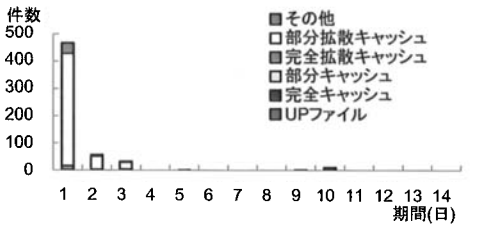

図 15：参照ノードに関係するキーの出現期間[Share]

参照ノードを対象に調査を行なった．調査期間中に観測された参照ノードに関係する一意なプライマリキーは 568 件であり，その内訳を表 4に示す．また，参照ノードと関係するキーの観測件数 2,245 件を対象に最古観測日と最新観測日との間隔を図 15に示す．最古観測日は，観測期間中に参照ノードと関係するプライマリキーのうち最も古い日付であり，最新観測日は最も新しい日付である．この最新観測日と最古観測日の幅を出現期間とした[*2]．図 15の出現期間 1 日以下は約 82%であり，観測されたキー(総数約 3 億件)の中から約 200 万件(ユニークプライマリキー約 100 万件に相当)を対象とした場合には，出現期間 1 日以下は約 89%となった．

4.3 考察

調査結果に基づき，Winny／Share ネットワークの比較を行なう．

(1) 流通量に関する観測結果(表 5)

表 5：稼動ノード，ファイル流通量に関する結果

項目	Winny	Share
ノード数	約 35 万ノード	約 15 万ノード
オリジナルファイル数	約 450 万ファイル	約 40 万ファイル
拡張子別のファイル数	拡張子 JPG, ZIP, MP3, AVI が約 80%を占める．	拡張子 ZIP, AVI, JPG が約 60%を占める．拡張子 ISO がトップ 10 に含まれる．

(2) ファイル保持ノードの特定に関する観測結果

(a) Winny ネットワーク

図 9の結果と仮想キーの生存期間[*3]とを考慮すると，単位時間 24 時間に 1 回しか観測できなかったプライマリキーは，キャッシュブロックが保有されたプライマリキーの可能性は低く，キー転送のみに関与している仮想キーであると推定できる．一方，キャッシュブロックが保有されているプライマリキーは単位時間 24 時間に複数回出現する傾向が見られる．複数回出現という特徴を持つプライマリキーを観測することで，プライマリキーからファイル保持ノードの IP アドレスを特定することができると考えられる．

(b) Share ネットワーク

*2) 本調査における最大幅は観測期間である 48 日となる．

*3) 表 1のキー消滅判定タイマーフィールドに格納された値であり，初期値は 1,500 秒．

低速な通信速度設定の参照ノードでは，アップロードファイル，完全キャッシュのキー情報のみを観測した．高速な通信速度設定の場合には，アップロードファイル，完全キャッシュ，拡散キャッシュのキー情報に加え，ファイルを保持していないキー情報を観測した(図 16，表 4)．ファイルを保持していないキー情報の観測件数は，Share ノードの設定により変動すると思われるが，キー情報はノードのファイル保持動作に関わっていると思われる．

観測されたキー(総数約 3 億件)の中から約 200 万件を抽出し，プライマリキーの観測時刻とキー情報に含まれるキーの更新日時(表 3)との差分の最小値をプライマリキー毎に算出した結果を図 17に示す．図 17から，キーの更新日時から約 11 時後に本 Sharebot 環境で観測される頻度が高いことがわかる．また，参照ノードと関係するキーの観測件数 2,245 件を対象とした，初観測時刻と参照ノードにおけるファイル生成時刻の間隔，キー更新日時から観測時刻までの間隔を図 18に示す．図 18から，参照ノードでは，キー情報が作成された約 9 時間後にファイルが作成され，さらに約 2 時間後に本 Sharebot 環境で観測されるという傾向が見られる．このことから，ファイルを保持していないキー情報は，キーは作成されたが，キャッシュファイル作成に至らなかったと類推される．

Winny／Share いずれの場合もプライマリキーの出現頻度が多く，さらに出現期間が長い場合，そのノードは，特定ファイルハッシュ値に対応する実ファイルを保持している可能性が高い．このことから，クローリング手法を用いてプライマリキーを広域観測することで特定のファイルを持った Winny／Share ユーザの IP アドレスを特定することは可能であると考える．

5　おわりに

本稿では，Winnybot／Sharebot を用いて収集したデータを元に Winny／Share 稼動ノード数，ファイル数を推定すると共に，Winny／Share においてファイル保持ノードの IP アドレスが特定可能であることを示した．

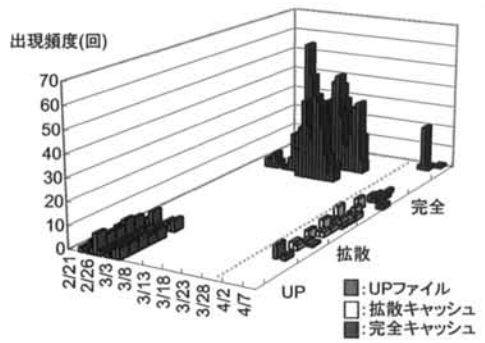

図 16：参照ノードが関係するキーの観測状況[Share]

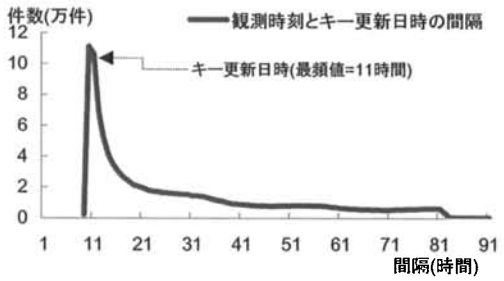

図 17：観測時刻とキー更新日時の間隔[Share]

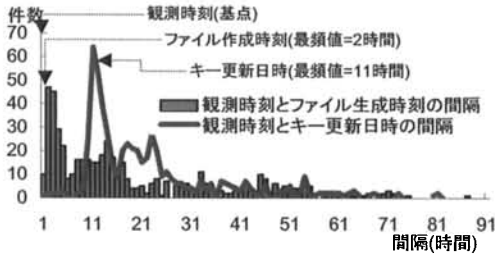

図 18：観測時刻とファイル生成時刻の間隔[Share]

今後の課題は，P2P ファイル交換ソフトウェア環境を対象とした定常的な観測手法の確立とノード数や流通量推定に関する手法の改善などが挙げられる．

謝辞

本研究の一部は社団法人コンピュータソフトウェア著作権協会の支援を受け実施した．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] 鵜飼裕司，“Inside Winny　～ Winny の解析とそのセキュリティ脅威分析”，
https://sec.scs.co.jp/eeye/shiryo.html
[2] 寺田真敏，鵜飼裕司，金居良治 他，“P2P ファイル交換ソフトウェア環境を対象とした観測に関する一考察”，SCIS2007 (2007)
[3] NetAgent，“Winny ノード数の推移”，
http://www.onepointwall.jp/winny/winny-node.html
[4] NetAgent，“Share とは？”，
http://forensic.netagent.co.jp/share_what.html
[5] (株)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手法についてのレポート”，http://www.crosswarp.com/info/070118.html
[6] Napster, http://www.napster.com/
[7] 鵜飼裕司, “P2P ソフト Share の暗号を解析，ネットワーク可視化システムを開発”，
http://itpro.nikkeibp.co.jp/article/Watcher/20070122/259207/
[8] 鵜飼裕司, “Share ネットワーク可視化システム「Sharebot」の仕組みと運用方針”，
http://itpro.nikkeibp.co.jp/article/Watcher/20070124/259460/