P2P ファイル交換ソフトウェア環境 Winnyp を対象とした観測
Observation of P2P file exchange environment Winnyp

安藤 槙悟† 寺田 真敏† 土居 範久†
Shingo Ando Masato Terada Norihisa Doi
中央大学大学院 理工学研究科†

1. はじめに

今日，P2P ファイル交換ソフトウェアを用いた，配布の認められていないファイルの交換による著作権侵害，Winny を介したマルウェアによる情報漏洩が社会的な問題となっている．このような問題の背景として，P2P ネットワークが持つ匿名性と水平型の分散ネットワークという特徴が，P2P ネットワーク全体の把握を困難にしていることが挙げられる．

本稿の目的は，ノード数，ファイル流通状況など，P2P ネットワークの実態を明らかにしていくことで，P2P ネットワークに存在する匿名性の問題，さらに著作権侵害，情報漏洩などの問題解決に寄与していくことにある[1].

本稿では，調査報告の少ない Winnyp に関する利用実態に関する調査をおこない，ファイル名、拡張子、ファイルサイズに注目した調査結果をもとに，Winny2 と Winnyp ノード群から構成されるネットワークの流通状況の比較結果を報告する．

2. Winny2 と Winnyp の概要

P2P ファイル交換ソフトウェアとは，個々のノード同士の自律的な通信を利用して不特定多数のノード同士とファイルを交換するためのソフトウェアである．Winny は高速性と匿名性の両立を目指した P2P ファイル交換ソフトウェアで，Winny2 は Winny に BBS 機能を追加したバージョンである．Winnyp は，Winny2 プロトコルに変更を加えた P2P ファイル交換ソフトウェアで，Winny2 と互換性があり，Winnyp ノード同士が接続するためのプロトコルと，Winny2 ノードと接続するためのプロトコルを持ち合わせている．しかし，Winnyp ノードの中には，Winnyp ノード同士でのみ通信するノードが存在する．そのため，Winny2 と Winnyp によって構成される Winny ネットワークは，Winny2 ノード群と Winnyp ノード群によって構成されている．また，Winny2 ノードとも通信する Winnyp ノードは，Winny2 ノード群と Winnyp ノード群とのファイル交換の橋渡しを果たすことになっている．

3. クローリング調査による Winny ノード群の観測

3.1 クローリング調査

クローリング調査とは，ノードが保持している他ノード情報の一覧を取得した後，取得した他ノード情報を用いて同様の動作を繰り返していくことで，P2P ファイル交換ソフトウェアが稼働するノードを網羅的に調査する方法である．この調査を通して，IP アドレス，ポート番号などのノード情報と，ファイルの所在やファイル情報を格納したキー情報の 2 つを収集することができる．

したがって，ある観測時間内に集めたノード情報を重複を省きながら累積していくことで，ある観測期間内に稼働していたノード数を推定できる．

同様に，ファイル本体から計算された，ファイルの識別情報として使用されているキー情報のハッシュ値を重複を省きながら累積していくことで，ある観測期間内に流通していた一意なファイル数を推定できる．

3.2 Winnyp ノード群の観測

2 章で述べたように，Winnyp は 2 つのプロトコルを持ち合わせており，接続先ノードによって使い分け，通信をおこなう．相手から接続を要求された場合，Winny2 プロトコルであれば Winny2 のプロトコルを使用し，Winnyp であれば Winnyp プロトコルを使用する．

本調査では，Winny2 ノード群の観測については，Winny2 プロトコルを用いたクローリング調査を実施し，Winnyp ノード群の観測については，Winnyp プロトコルを用いたクローリング調査を実施した．Winnyp ノード群の具体的な観測は，Winnyp プロトコルを使用した通信を開始する．接続が完了すれば，接続先ノードは Winnyp であると判断し，Winnyp プロトコルを用いたクローリング調査を実施する．もし，接続先ノードとの接続が完了しない場合には，接続先ノードは Winnyp 以外であると判断し終了する．Winny2 ノード群の観測も同様に実施した．

4. Winny ネットワークの観測結果

Winny2 と Winnyp それぞれのノード群を対象にクローリング調査をおこなった際の観測環境は表 1 の通りである．

表 1 クローリング調査の観測環境

	Winny2	Winnyp
観測期間	2008 年 9 月 1 日(月)の 24 時間	
観測装置台数	10 台	9 台
観測キー数	約 4 億 545 万件	約 3 億 9643 万件

4.1 ノードの稼働数

ノード情報のうち，IP アドレスとポート番号の組み合わせを一意のノード識別情報とし，Winny2 と Winnyp ノード群のそれぞれにおいて，重複を省きながらノード識別情報件数を累積した結果， Winny2 で 212,555 ノード，Winnyp で 11,064 ノードに相当するノード識別情報件数を得た．また，Winny2 と Winnyp 同士のノード識別情報の重複から， Winnyp ノードのうち，Winny2 と通信可能であったノードは 8,631 件であった．

4.2 流通ファイル数

キー情報のうち，ファイルのハッシュ値を一意のファイル識別情報とし，Winny2 と Winnyp ノード群のそれぞれ

において，重複を省きながらファイル識別情報件数を累積した結果，Winny2 ノード群に流通しているファイル数は約 518 万件，Winnyp ノード群に流通しているファイル数は約 242 万件であった．また，ファイル識別情報の重複から，双方のネットワークで観測されたファイルは約 188 万件であった(図 1).

図 1　Winny2/Winnyp クローリング調査による流通ファイルの重複度

4.3 流通ファイルの拡張子分布

キー情報に含まれるファイル名から，ファイル拡張子を抽出し，Winny2 と Winnyp ノード群のそれぞれに存在する拡張子の分布を調査した結果，Winny ネットワーク全域で最も多く観測されたのは拡張子 jpg(30%)であった(図 2).

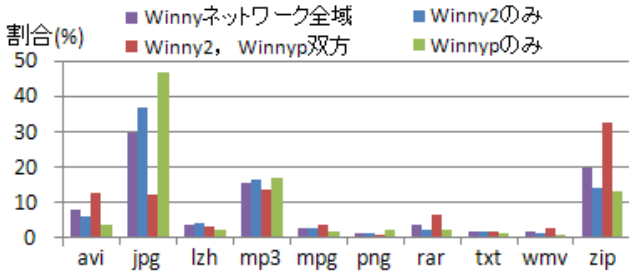

図 2　流通ファイルの拡張子分布

4.4 流通ファイルのファイルサイズ分布

キー情報に含まれるファイルサイズから，Winny2 と Winnyp ノード群で流通するファイルサイズ分布を調査した結果，Winny2，Winnyp 双方に流通しているファイルは，他の領域に比べ，8Mbyte 以上のファイルが多く流通している (図 3).

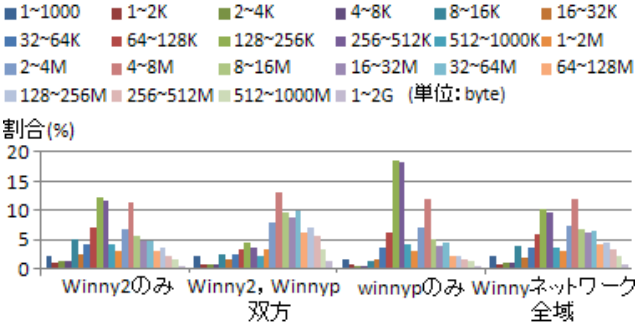

図 3　流通ファイルのファイルサイズ分布

4.5 ファイル名に注目した調査

キー情報に含まれるファイル名から，先頭にファイルの分類情報に相当するクラスタワードの記載されているファイルの割合を調査した結果，Winny2，Winnyp 双方に流通しているファイルの割合(70%)が最も多かった(図 4). また，avi/jpg/mp3/zip 拡張子を持つファイルに絞って調査した結果，各々の領域において拡張子 zip が最も多く流通しており，最も少なかったのは Winnyp ノード群のみに流通している拡張子 jpg のファイルであった(図 5).

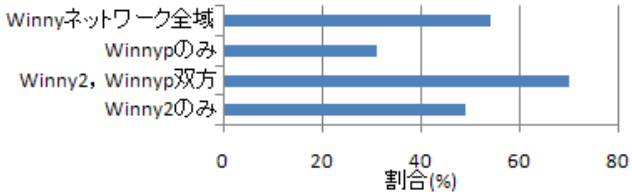

図 4　クラスタワードが記載されているファイルの割合

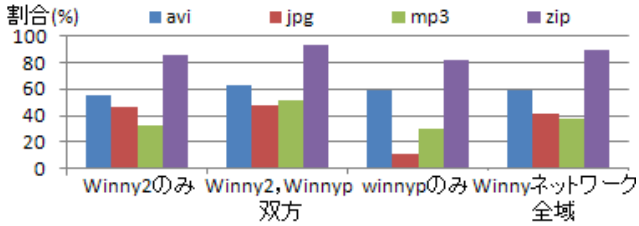

図 5　拡張子に注目した，クラスタワードが記載されているファイルの割合

4.6 情報漏洩ファイル

マルウェアによって特徴的なファイル名が付加されることを利用して情報漏洩ファイルを特定し，情報漏洩ファイルの割合とその拡張子分布を調査した結果，Winny2 ノード群のみの領域において，情報漏洩ファイルが最も多く流通していた(図 6)

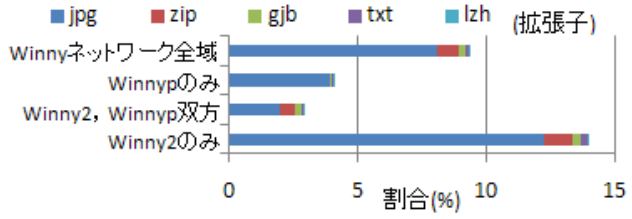

図 6　拡張子別の情報漏洩ファイルの割合

5. まとめ

本稿では，Winny2 と Winnyp ノード群を対象とした，クローリング調査を通して，Winny2 と Winnyp のノード数，流通ファイル数を推定すると共に，Winny2 と Winnyp ノード群が構成するネットワークにおけるファイル流通状況の調査結果を報告した．

今後は，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合わせて，P2P ファイル交換ソフトウェア環境が生み出す社会的問題の解決を推進する予定である．

参考文献

[1]寺田 真敏, 鵜飼 祐司, 金居 良治, 畑山 充弘, 松木 隆宏, 宮川 雄一, “クローリング手法を用いた P2P ネットワークの観測”, 情報処理 CSEC 研究報告, Vol.2007, No.48 (2007/5).