

# P2P ファイル共有ネットワークにおける コンテンツの定量的特性の測定と評価

渡部 友也†

大坐 島 智‡

川島 幸之助‡

† 東京農工大学 工学部

‡ 東京農工大学大学院 共生科学技術研究院

## 1 はじめに

近年、ADSL や FTTH などの広帯域接続環境の普及により、インターネットにおけるブロードバンド化が急速に進んできている。インターネットのブロードバンド化が進むにつれ、従来の通信方式であったサーバー/クライアント型のネットワーク形態とは異なる Peer-to-Peer (以降 P2P とする) と呼ばれるネットワーク形態を構築するアプリケーションが普及してきた。その中でも、特に P2P ファイル共有アプリケーションが目立っている。P2P ファイル共有アプリケーションでは動画などのマルチメディアファイルが共有されることが多く、ネットワーク資源の浪費という問題がある。ネットワーク上の流通ファイルの調査の方法として、クロウリングという方法がある。クロウリングとは対象の P2P ネットワークに参加し、通信を行うことによって情報を収集する方法である。本稿では、この方法を用いてデータを収集するクローラを用いて収集したデータを解析、評価した。

## 2 測定・評価

P2P ファイル共有アプリケーションがネットワークに与える影響及び流通ファイルの特性を調べるため、クローラを用いて収集したデータを解析・評価した。測定では、Winny には東京農工大学川島研究室開発のクローラを、Share には eEye Digital Security 社の Sharebot[2] を用いた。データは 2007 年 12 月 17 日に 8 時間収集したものをを用いた。

### 2.1 ピアの共有ファイル数の解析

各ピアが共有しているファイル数とピア数の関係、各ピアが共有しているファイル数と全体のファイル数との関係を調べた。共有ファイル 0 のピアは対象外としている。図 1 では、Winny, Share 共に、共有ファイル数 10 個以下のピアは全ピア数の 8 割を占めていた。また、Winny では 10 個以下のファイルを共有するピア

が所有するファイル数は、ピア全体で共有されているファイル数の 24.6 パーセント、Share では 43.6 パーセントであった。このことから、共有ファイル数の多いピアを重点的に制御すると効率のいいファイル流通制御を実現できるといえる。また、Winny では共有ファイル数 1000 個近くのピアが少数存在するが、Share では 200 個以上共有しているピアはほぼ 0 であった。

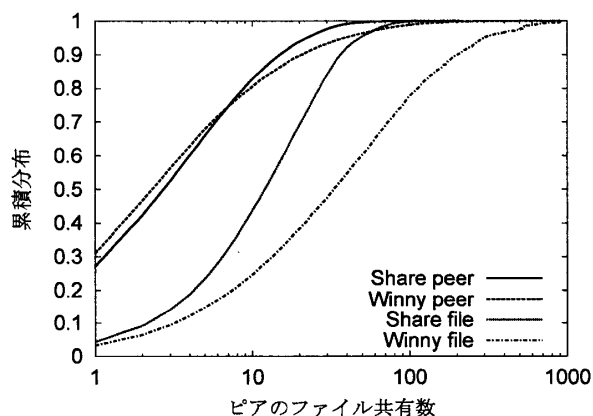


図 1: ピアの共有ファイル数の累積分布 (ピア、ファイルベース)

### 2.2 Share 流通データファイルサイズの推測

P2P ファイル共有アプリケーションがネットワークに与える影響の中で、流通ファイルのサイズは大きな要因となる。しかし、今回使用した Sharebot ではファイルサイズを取得することができない。そこで、Winny で収集したデータから Share 流通データのファイルサイズを推測するために、Winny と Share で同じファイル名を持つものを検出し、検出されたファイルは Winny 流通ファイルと同じファイルサイズであるとして、Share 流通ファイルサイズの推測を行った。表 1 で示している Share (推測) のファイル数とは、Share 流通ファイルの中で、Winny 流通ファイルと同じファイル名のデータのことである。Share (推測) の平均ファイルサイズと中央値は、Winny のそれらよりも大きいという結果が得られた。しかし、Winny では扱えるファイルサイ

ズは 2GB である。Share では 32GB まで扱えるという使用を考えると、Share では 2GB 以上のファイルが流通していることになる。しかし、今回の推定方式ではそのファイルは考慮しないので、平均サイズや中央値はさらに大きいものであると考えられる。

表 1: 流通データのファイルサイズ

	ファイル数	平均サイズ	中央値
Winny	4000000	165.98MB	28.5MB
Share	800000	-	-
Share (推測)	42651	202.76MB	86.5MB

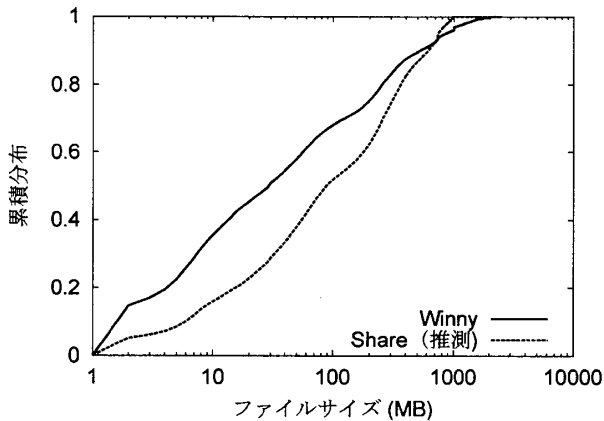


図 2: ファイルサイズの累積分布

図 2 の Winny では、10MB 以下のファイルが 35.5 パーセントでありファイルサイズの小さいファイルが多いことが示されている。小さいファイルは jpg, mp3, txt などのファイルであった。また、Share の (推測) では 500MB 以上の大きいファイルサイズが多いことが図 2 から読み取れる。このことから、Share 流通ファイルは Winny よりも大きいものが多いといえる。

### 2.3 流通ファイルの拡張子分析

流通ファイルの拡張子の分布を調べた。図 3, 4 より、Winny では zip が、Share では avi がもっとも多かった。また、Winny では mp3, jpg といった小さいファイルサイズのものも多く、Share では rar, ISO といった大きいファイルサイズのもの多かった。これらの傾向から、Share 流通ファイルの方が平均サイズは大きいと考えられる。また、Share で 9 パーセントある ISO ファイルは Winny では 1 パーセント以下であった。こ

れは、ISO ファイルは CD/DVD をイメージ化したものなので、Winny で扱えるファイルサイズを超えることが多く、rar ファイルに分割されているためである。Share も rar ファイルが多いが、Winny とは違い分割されてないファイルが多かった。rar は zip に比べて高圧縮率なので、大容量ファイルは rar で圧縮されているファイルが多いと考えられる。

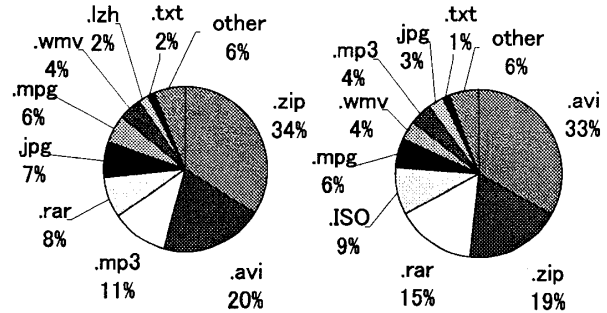


図 3: Winny の拡張子分布 図 4: Share の拡張子分布

## 3 おわりに

本稿では、クローラを用いてデータを収集・解析した。それにより、Share は Winny よりもファイルサイズの大きいデータの通信が多いという結果が示され、Share がネットワークに与える影響は大きいものであると考えられる。よって、今後 Share のファイル流通制御方法の研究の必要性が高くなっているといえる。今後の課題としては、継続的に P2P ファイル共有アプリケーションの特性を測定し、動向を把握することにより、効果的なファイル流通制御方式、管理方式の開発が考えられる。

### 謝辞

本研究の一部は科研費基盤 C(No. 18500047), および、電気通信普及財団の研究助成の支援を受けており、ここに記して感謝する。

### 参考文献

- [1] ネットエージェント, <http://www.netagent.co.jp/>
- [2] Sharebot, <http://scs.co.jp/eeye/sharebot.html>
- [3] 亀井聡, “P2P 技術がネットワークインフラに及ぼす影響と課題”, 日本ソフトウェア科学会, pp.14-18, 2005.