

## P2Pファイル共有の実態調査

インターネットへのアクセス環境の向上により、近年著しく、P2Pアプリケーションが普及してきています。中でも、P2Pファイル共有アプリケーションの普及は、通信事業者、著作権ビジネスにかかわる者たちをはじめ、さまざまな領域への影響が避けられません。しかしながら、ファイル転送時にはサーバを介さないP2Pアプリケーションの特性から、大規模な情報収集は困難とされてきました。本稿では、P2Pファイル共有アプリケーションの測定手法と測定結果を紹介し、P2Pファイル共有の規模、共有されているファイルの実態を示します。

おおい けいたい<sup>1)</sup> かめい せじ<sup>2)</sup>  
大井 恵太 / 亀井 聡  
もり たつや<sup>1,2)</sup>  
森 達哉

<sup>1)</sup> NTT情報流通プラットフォーム研究所

<sup>2)</sup> NTTサービスインテグレーション基盤研究所

### P2Pファイル共有とは

P2Pファイル共有の起源は、P2Pの代名詞ともなったNapster<sup>\*1</sup>にさかのぼります。P2Pファイル共有アプリケーションの動作している端末はピアと呼ばれ、P2Pファイル共有ネットワークに参加します。P2Pファイル共有アプリケーションは、ネットワークを介してピアのディスク上にあるファイルを検索する機能と、検索によって各ユーザが発見したファイルを直接、あるいは他のピアを介して自分のピアへと転送する機能を持っています。

### 実態調査の目的

近年、このようなP2Pファイル共有アプリケーションが普及することにより、インターネットのトラヒックが爆発的に増加しています。すでに、平均的なユーザに比べてデータ転送量が多いユーザの契約の解除を会員規約にうたう、といった対応を行う通信事業者も現れるなどの影響が出てきています。こういった通信事業者に対する影響の評価や将来予測を行うためには、P2Pファイル共有によるトラヒックやファイルの共有状況といった実態調査

が欠かせません<sup>(1)</sup>。

また共有されているファイルの内容に関しても、いわゆる違法ファイルの共有が多くを占めているという、著作権ビジネス上の問題もあります。これに対しては、現行法の厳密な適用や法改正により、違法とされるファイルの共有を防ごうとする試みもあり、2001年以降、実際に逮捕者も発生しています。一方で、現行の著作権法に対する改革論もあり、関心を集めています。こういった議論の際に、法改正の有効性や本当にP2Pの普及が著作権ビジネスを疎外しているかといった検証や判断のためにも、実態調査に基づく現状把握が必要となっています。

### さまざまなP2Pファイル共有アプリケーション

実際に広く使われているP2Pファイル共有アプリケーションについて紹介します。

前出のNapsterでは、ファイル検索用の中央サーバを必要としていました。このため、検索トラヒックはWebと同様のパターンを示します。しかし、トラヒックの大部分を占めるファイルの転送トラヒックは、各ピア間で中央

サーバを介さずに直接発生するため、Webとは大きく異なるトラヒックを示します。Napster社はP2Pサービスを停止しましたが、Napsterと同様のプロトコルを利用した互換サーバ/互換クライアントは現在も用いられています。

特に、互換クライアントの機能を持つWinMX<sup>\*2</sup>は日本国内で広く普及しており、国内ネットワークにおいて顕在化しているP2Pトラヒックの中では最大のものでされています。

Napsterの後に登場したGnutella<sup>\*3</sup>は、中央にサーバを必要としないアーキテクチャを特徴とし、ファイル転送

\*1 Napster: Napster社によって開発された初期のファイル共有アプリケーション。学生を中心に爆発的に普及し、最盛期のユーザ数は8000万人にも上ったとされています。最近P2Pではない音楽配信会社として復活しました。

\*2 WinMX: Frontcode Technologies社によって開発されたファイル共有アプリケーション。

\*3 Gnutella: Winampの開発元として知られるNullsoft社のプログラムによって開発されましたが、即座に親会社であるAOLによって公開停止となりました。このオリジナルバージョンを基に互換アプリケーションが開発されました。仕様がオープンなためにさまざまな研究の対象となっており、Gnutella2等の次世代プロトコルも提案、実装されています。また同じプログラミングによってwasteというP2Pコラボレーションソフトも公開されましたが、同様に親会社によって再配布停止処分を受けました。

だけでなく、検索も各ピア間で自律分散的に通信することで実現しています。中央サーバを持たないことから、第三者がサービスの停止や規制を行うことも難しくなっています。

その後さまざまなP2Pファイル共有アプリケーションが登場しており、現在世界で最大規模のP2Pネットワークを構成しているファイル共有アプリケーションはKaZaA<sup>\*4</sup>だとされています。しかし、ファイル名が多言語に対応していないことなどから、日本国内ではWinMXほどには普及していません。

国内では、ほかに匿名性の高いFreenetを基に作成されたとするWinny<sup>\*5</sup>と呼ばれるアプリケーションも存在します。このWinnyはトラフィック面でも急激に伸びており、通信事業者の網内部といった測定場所によっては、WinMXのトラフィックをも上回っている、という測定報告もあります。

### アプリケーションレイヤでの測定

P2Pアプリケーションでは図1のように、P2Pアプリケーション動作ピア間でIPネットワークとは独立したP2Pネットワークを物理ネットワークと独立したレイヤの論理網としてオーバーレイで構築します。このため、IPレイヤの測定だけではユーザ数の規模やファイルの流通状況といった情報を得ることは困難であり、全体像把握のためには、アプリケーションレイヤでの測定を合わせて実施することが必要になります。

またプロトコル仕様が公開されていないアプリケーションについては、ネットワークレイヤの測定では利用状況を推定することさえも困難です。しかも、アプリケーションによって収集可能な情報は異なるため、アプリケーション

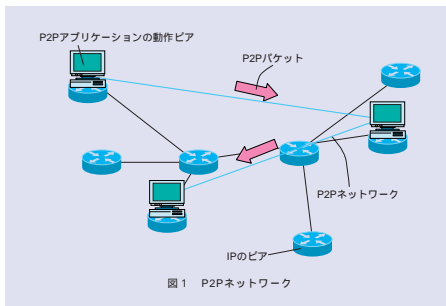


図1 P2Pネットワーク

ごとに独自の測定手法を開発する必要があります。

以下では、Gnutellaを例に論理網トラフィック測定手法および条件を説明し、各アプリケーションの測定結果を紹介します。

#### 測定手法

Gnutellaは、自律動作するピアによって、P2Pネットワークを構築し、検索や生存確認といったネットワークを維持するための制御パケットはすべてこのネットワーク上でやり取りされます。

制御パケットには次の5種類があります。

- ・PING：他のピアを発見するためのパケット。
- ・PONG：PINGに対する応答パケット。アドレス、利用可能容量を含む。
- ・Query：検索文字列を含んだ検索要求パケット。ピア間をあらかじめ指定した回数だけ転送される。
- ・QueryHit：Queryに対して要求に合致したファイルを持っていたピアの応答パケット。ファイルサイズとURLを含む。

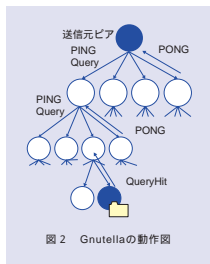


図2 Gnutellaの動作図

・PUSH：ファイル送信者側がファイアウォール内にいる場合のファイル送り込み要求パケット。

これらのパケットは、図2のように送受信されます。各ステップでの動作を次に示します。

別のピアにPINGを送信。

\*4 KaZaA：オーストラリアShaman Networks社によって配布されるファイル共有ソフトウェア。FastTrackと呼ばれるP2PネットワークはP2P業界最大のシェアを誇ります。昨年末にオランダ最高級で合法判決を勝ち取りました。

\*5 Winny：団産のファイル共有アプリケーション。国立大学助手が作成したとされています。

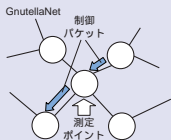


図3 Gnutella測定イメージ

PINGを受信したピアは送信元ピアにPONGを送信し、さらに別のピアに対して送信元ピアの識別子を格納したPINGを送信。ピアを経由するたびに転送回数を減算。

PONGは経由してきたピアを逆順に経由し、PING送信元ピアに到着。

検索の際は、Queryを送信し、該当ファイルを所有するピアはQueryHitを元ピアに送信。

#### 測定条件

測定には改造を施したピアを用意し、制御パケットの収集を行うことで実施しました(図3)。ピア間で直接行われるファイル転送については測定の対象とはしていません。このため、以下の測定結果では、流通しているファイルのうちで、検索対象となったファイルの情報に限定されていることに留意する必要があります。

また同時に論理網トラフィック測定技術を用いて、動作中のピアのネットワークレイヤでのトラフィックを測定することにより、P2P論理ネットワーク上の情報も併せて収集し、P2Pネットワーク規模の情報測定を行いました。本技術に関してはP2Pアプリケーションに汎用的に適用できる技術となっています。

表 各アプリケーションでの測定結果

種別	固有IPアドレス	固有ファイル数	合計ファイル容量	平均ファイル容量
Gnutella	30 052	3 883 752	1.3 PB	330 MB
WinMX	350	165 933	25 TB	150 MB

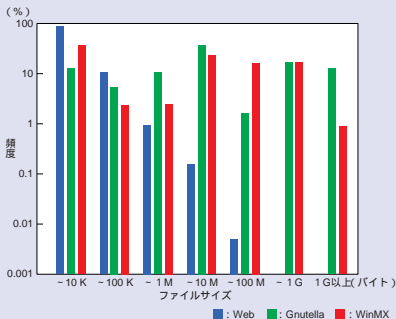
P: 10<sup>15</sup> T: 10<sup>12</sup> M: 10<sup>6</sup>

図4 ファイルサイズの分布

#### 測定結果

##### (1) ネットワーク規模

ネットワーク規模を示すデータとして、Gnutella, WinMXについて取得したデータにおける固有IPアドレス、固有ファイル数、合計・平均ファイル容量を表に示します。

Gnutellaについては、昨年上期に週末の68時間測定した結果、WinMXはGnutellaとの比較のためにサンブル測定した結果です。

Gnutellaでは、3万程度のIPアドレスが収集できており、ネットワーク規模の大きさを推定することができます。ただし、より長期間にわたって測定を続け、より多くのIPアドレスが収集できたとしても、IPアドレス数 ユーザ数ではないことに注意が必要です。この場合は、IPアドレスの有効期間も

考慮してはなりません。

##### (2) ファイルサイズ

Gnutella, WinMX, およびWebについて、それぞれのネットワークで流通するファイルサイズの比較を行いました。

ファイルサイズのヒストグラムを両対数グラフで描いたものを図4に示します。Gnutella, WinMXは、Webに比べてファイルサイズが非常に大きいことが分かります。

共有されているファイルの拡張子分布を図5に示します。音声ファイルのmp3が大きな割合を占め、ファイルサイズが大きいことが多いavi, mpgといった動画ファイルが続いています。

##### (3) 参照回数の分布

Gnutellaにおける参照回数の分布を図6に示します。図6は、x軸にファ

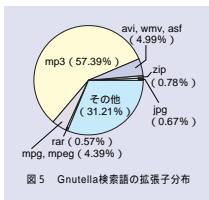


図5 Gnutella検索語の拡張子分布

イルの参照回数 ( $n$ ),  $y$ 軸に $n$ 回参照されたファイル数とした両対数グラフです。

図6でみて取れる直線性は、Lotokaの法則（またはZipfの第二法則）と呼ばれるものです。一般に、個々のユーザが何かを選んだときに発生する記録では、このようなLotokaの法則にのった形状（Lotoka分布）が生じます。

Lotoka分布の履歴に対しては、総参照回数に比べて総コンテンツ数が少ない場合に、コンテンツの参照回数が少ない領域（左上の領域）において、直線性を保たず、下方に曲がること、などが知られています<sup>(2)</sup>。

さらに、Lotoka分布の履歴では、協調フィルタリングシステムが有効に働くことが知られており、商品推薦サービスなどに利用されています<sup>(3)</sup>。このことから、Lotoka分布であるP2Pネットワークの履歴についても、マーケティング分野などへの活用の可能性を有しています。

## 今後の予定

NTT情報流通プラットフォーム研究所では、今回観察されたLotoka型分布を示す履歴の性質を利用し、より大規模な履歴の分析や利用を可能とする分析方式を考えています。

またP2Pファイル共有アプリケー

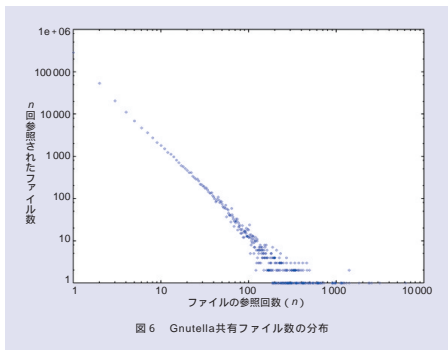


図6 Gnutella共有ファイル数の分布

ションのユーザ規模は増加傾向にあり、多方面への影響を評価、予測するうえでも継続的な調査が必要となっています。

P2Pトラヒックを識別する方式は、現在アプリケーションごとの特徴をパターンファイル化する方式が主に用いられています (Ellacoya, P-Cube等)。しかし、これらの方式は新たなアプリケーションの登場やバージョンアップごとに新たなパターンファイルをつくる必要があり、誤動作の恐れも高くなります。また通信内容（ペイロード）を監視する必要があるために、大規模化への対応が困難となっています。NTTサービスインテグレーション基盤研究所では本稿で活用した論理網トラヒック測定技術と本技術により収集したピアの情報を用いてP2Pトラヒックを識別するトラヒック分離方式を開発しています<sup>(4)</sup>。これによって、ネットワークへのP2Pトラヒックの影響を見積もることや、分離したトラヒックに対して独自の制御を行うことが容易となります。

## 参考文献

- 森・亀井・大井：“P2Pトラヒックの測定と特性評価”，2003信学ソダ，SB-3-1，2003。
- 村中・松田・会田・本橋・佐藤：“有限アドレス空間に対するinternetアクセスパターンの特性分析”，信学技報，IN2001-56，2001。
- 本橋・佐藤・金井：“ポータルサービスにおける広告・マーケティング技術”，NTT技術ジャーナル，Vol.14，No.1，pp.85-87，2002。
- 亀井・森・大井：“P2Pファイル共有の実態と課題”，信学技報，CO2003-40，2003。



(左から) 大井 恵太/ 亀井 聡/  
森 達哉

P2Pは、ISPによる規制の導入や不正利用等、否定的文脈で語られることが多いですが、技術そのものは光時代のキラアプリケーションとして有力です。今後もネットワーク、コンテンツビジネス両方の面から、実態調査は欠かすことができません。

## 問い合わせ先

NTT情報流通プラットフォーム研究所  
TEL 0422-59-6150  
FAX 0422-59-2699  
E-mail ooi.keita@lab.ntt.co.jp