

## P2P ファイル共有ネットワークのグラフ的特徴の解析および評価

平野誠<sup>†</sup> 吉田雅裕<sup>†</sup> 大坐島智<sup>‡</sup> 川島幸之助<sup>†</sup>

東京農工大学<sup>†</sup> 電気通信大学<sup>‡</sup>

### 1. はじめに

近年、ブロードバンド回線の普及により、P2P ファイル共有アプリケーションが多くの人たちによって使用されるようになった。P2P ファイル共有ソフトは、大きなサイズのファイル転送やファイルの検索機能を提供するために膨大なトラフィックを生み出す。そのためネットワーク帯域を圧迫し問題となっている。しかし反面 P2P 技術は従来のクライアント/サーバモデルと比較し、サーバが存在しないため負荷分散や導入コストの削減などが期待でき、技術的な利用価値は高い。

P2P ネットワークに対してそのトポロジデータを収集し、ネットワークが実際にどのような特徴を持っているか解析することは、よりよいネットワークトポロジの考案に役立つ。また P2P ネットワークのシミュレーションにも利用できる。

そこで本研究では、世界的に有名であり最も大きな P2P ファイル共有ネットワークとされる Gnutella に注目した。Gnutella はウルトラピアと呼ばれる主にクエリ転送を受け持つノード群と、リーフピアと呼ばれるクエリ転送には貢献しないノード群の 2 層を持つピア P2P 型のファイル共有ネットワークである。今回 Gnutella ネットワークにおけるノードの情報および、ノードが持つリンクの情報を収集するためにクロールリング手法を用いたアクティブ測定を行った。クロールリング手法とは、P2P ファイル共有ネットワークに参加しているノードに接続して、そのノードや隣接ノードの情報を受け取り、これを繰り返すことによってネットワークを網羅的に測定する方法である。本研究では精度の高いトポロジデータを収集するために、複数の端末を並列動作させてクロウラの高速化を図った。結果として Gnutella ネットワークのトポロジデータを並列クロウラを用いて収集し、解析を行うことで Gnutella ネットワークのグラフ特性を明らかにした。

### 2. トポロジデータの取得

本研究では、P2P ファイル共有ネットワークのトポロジ情報を収集するためにクロールリング手法を用いたアクティブ測定を行った。Gnutella プロトコル ver0.6 以降においては、ネットワークに参加している任意のウルトラピアに対してハンドシェイクを行う際に Crawl と呼ばれる特定のヘッダを明記することで、そのピアが隣接しているウルトラピアおよびリーフピアの情報を得ることができる。これを繰り返し行うことによって Gnutella ネットワークのトポロジ情報の収集を行った。

クロールリングを行う際にはネットワークをより早く測定することが重要となる。理想的には瞬間的にスナップショットを取得できることが望ましいが、現実的には物理的な時間がかかる。つまりクロールリングを開始した時点のネットワークの状態は、クロールリングを行っている最中にもノードの参加や離脱により変化してしまう。そのため時間が経過するとともに収集したデータの精度が落ちてしまうことが問題となる。本研究ではより高速に測定を行うために 8 台の端末を用いた並列動作するクロウラを作成した。

### 3. 測定

本実験では 2009 年 10 月 27 日 19 時 35 分から Gnutella クローラを用いて Gnutella ネットワークのトポロジデータの収集を行った。クロールリングの経過時間と発見ノード数の関係を図 1 に示す。上の方の曲線がリーフピアの発見数、下の方の曲線がウルトラピアの発見数である。

グラフに注目すると経過時間とともに発見ノード数が緩やかに収束していることがわかる。しかしながらクロールリングを行っている最中にも、新たに Gnutella ネットワークに参加してくるノードが存在するため、発見数が完全に一定の値に収束するということがない。そのためある程度時間が経過したら十分に発見ノード数が収束したと考えクロールリングを打ち切る必要がある。今回は測定結果より 180 秒経過の時点で十分に発見ノード数が収束していると考え、それ以前に発見され

A Topological Analysis and Evaluation for Peer-to-Peer File Sharing Network

<sup>†</sup>Makoto HIRANO Masahiro YOSHIDA Konosuke KAWASHIMA, Tokyo University of Agriculture and Technology  
<sup>‡</sup>Satoshi OHZAHATA, The University of Electro-Communications

たノードおよびリンクを用いて解析を行った。

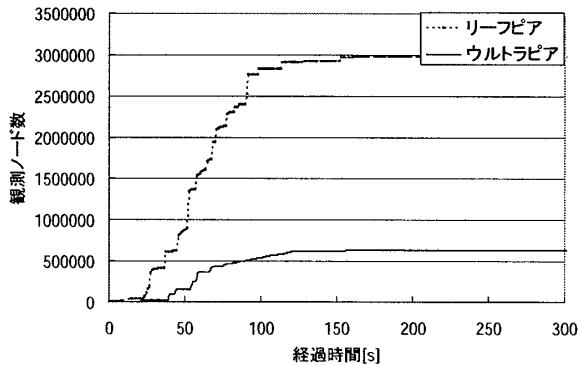


図 1: 発見ノード数と経過時間の関係

#### 4. 解析

測定により得られたデータの解析を行った。本実験ではウルトラピア間における次数分布、クラスタ係数の算出、ウルトラピア間における平均距離の解析を行った。次数分布の解析結果を図 2 に示す。参考文献[1]と比較して過去の分布と似た分布をしていることがわかる。それに対して観測ノード数は本実験の方が 20 万程度多い。近年、ノードの増加は激しいが次数分布を保ったままである。しかし Gnutella はクエリの転送にブロードキャストを採用している。解析結果より任意のピアにおいて 7 hop 程度でほぼ全ての他のノードへ到達できることから、検索機能は果たせていると考えられる。しかしノードの増加に伴ってトラフィックは膨大になっているということが予想できる。

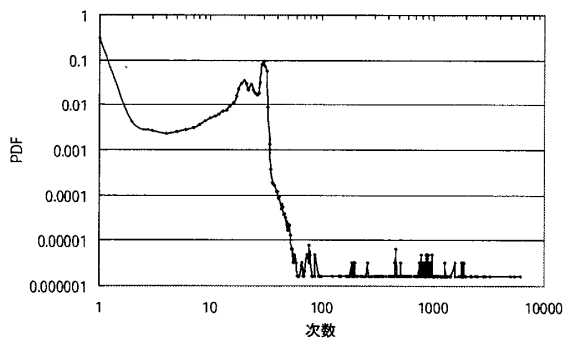


図 2: ウルトラピア次数分布

図 2 よりウルトラピア同士における次数の分布は 30 に集中していることがわかる。これは Gnutella クライアントの実装によって接続が制限されているためだと考えられる。またまれに次数の非常に大きいノードが存在しているがこれは研究目的など、改良されたノードであると考えられる。

また観測した Gnutella グラフと同じノード数、リンク数を持つランダムグラフをシミュレーションにより作成し、比較を行うことで Gnutella のグラフ特性を調べた。クエリ転送に関して次数の高いノード、すなわち研究などの目的によって改良されたノードがクエリ転送に貢献しているかどうかという点は不明である。そこで次数分布の解析結果より、観測ノード数が一気に減少する次数 35 を閾値としてそれ以上の次数のノードを含めた場合と含めなかった場合の両方を図 3 に示した。グラフに着目すると、ランダムグラフと実際の Gnutella トポロジはほぼ同様の平均距離を持つということがわかる。またクラスタ係数を算出したところ Gnutella は約 0.009、ランダムグラフは約 0.00005 となり Gnutella の方が 100 倍程度のオーダーでクラスタ係数が大きい。これは Gnutella がスモールワールド性を持つことを示す。

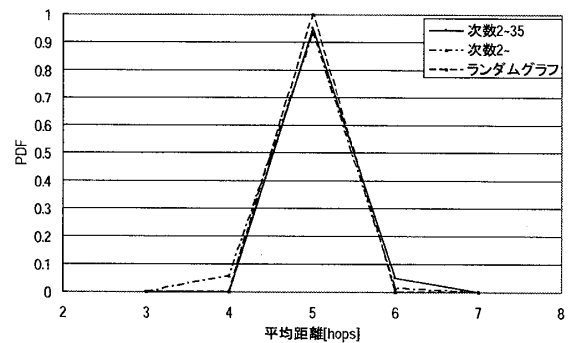


図 3: 平均距離の比較(ウルトラピア)

#### 5. まとめ

解析結果から、Gnutella は次数分布が実装に依存し次数 30 に集中していることからスケールフリー性を持たないグラフである。またランダムグラフとの平均距離、クラスタ係数の比較よりスモールワールド性を持つグラフと結論付けた。

#### 謝辞

本研究の一部は科研費基盤 B(No. 20300025) の研究助成によるものである。ここに記して感謝する。

#### 参考文献

[1] Chao Xie, Sijie Guo, Rejaie, R., Yi Pan, "Examining Graph Properties of Unstructured Peer-to-Peer Overlay Topology," IEEE Global Internet Symposium 2007, pp.13 - 18, 2009.

[2] LimeWire, <http://www.limewire.com/>