

## Peer-to-Peer ネットワークにおける データアクセス頻度を考慮した確率的な複製配置手法について

木戸 裕樹<sup>†</sup>      前田 和彦<sup>†</sup>      原 隆浩<sup>†</sup>      西尾 章治郎<sup>†</sup>

<sup>†</sup> 大阪大学大学院情報科学研究科マルチメディア工学専攻  
〒565-0871 大阪府吹田市山田丘 1-5  
E-mail: {kido.yuki,k.maeda,hara,nishio}@ist.osaka-u.ac.jp

近年、計算機の高性能化とネットワークのブロードバンド化に伴い、Peer-to-Peer(P2P) ネットワークを用いたデータ共有に関する研究が注目されている。本稿では、ピュア P2P 型ネットワークを用いたデータ共有サービスにおいて、各ピアの負荷分散と検索効率の向上を目的とした複製配置手法を提案する。提案手法では、検索クエリが要求データを保持するピアを発見した経路上のピアにおいて、各データに対するアクセス頻度の順位に応じて、複製を作成するか否かを決定する。さらに本稿では、シミュレーション実験により、提案手法の有効性を検証する。

## Replication Strategy Depending on Access Frequency in Peer-to-Peer Networks

Yuki KIDO<sup>†</sup>      Kazuhiko MAEDA<sup>†</sup>      Takahiro HARA<sup>†</sup>      Shojiro NISHIO<sup>†</sup>

<sup>†</sup>Dept. of Multimedia Eng., Graduate School of Information Science and Technology, Osaka University

Recently, there has been increasing interest in research of data sharing with peer-to-peer networks. In this paper, assuming a data sharing service, we propose a new replication strategy that achieves not only load balancing but also improvement of search performance. The proposed strategy creates replicas at each peer on the path along which a query is successfully forwarded, depending on the peer's rank of access frequency to the data. Moreover, we verify the effectiveness of the proposed strategy by simulation experiments.

### 1 はじめに

近年、計算機の高性能化とネットワークのブロードバンド化に伴い、Peer-to-Peer(P2P) モデルを用いたネットワークサービスに注目が集まっている。Web サービスなどの、既存ネットワークサービスとして最も主流になっているクライアント・サーバモデルとは異なり、P2P モデルでは、サーバやクライアントといった明確な区別はなく、端末(ピア)同士が対等な接続形態を取り、互いにサービスを提供しあう。このような P2P モデルは、クライアント・サーバモデルに比べ、サービス提供における負荷が分散されるため、スケーラビリティや耐障害性の面で優れているという特徴をもつ。

P2P モデルを用いたネットワークサービスでは、サービスの提供者が分散するため、各ピアが要求するサービスの提供者を検索する機構が必要となる。P2P モデルを用いたネットワークサービスの形態は、サービス提供者の検索機構によって、クライアント・サーバシステムを融合させたハイブリッド P2P 型

と、完全な分散環境であるピュア P2P 型に分類される。

ハイブリッド P2P 型は、サービス提供が可能なピアを検索するためにクライアント・サーバモデルを用いる。ハイブリッド P2P 型システムを適用したアプリケーションとしては、Napster[6] などがある。この P2P システムでは、検索サービスを提供するサーバが、ネットワーク上の全ピアの識別子(例: IP アドレス)やそれらのピアが提供可能なサービスを、インデックス情報として一括に管理する。そのため、各ピアがサービスを検索する場合、サーバに問合せを行うことで、サービス提供可能なピアを容易に見つけることができる。しかし、限られたサーバのみに検索の機能が集中しているため、サーバが故障もしくは停止した場合、システム全体が停止してしまう。さらに、完全な分散環境ではないため、クライアント数に対するスケーラビリティを得ることができず、クライアント・サーバモデルの問題点を完全に解決することができない。

ピア P2P 型のネットワークサービスでは、ハイブリッド P2P 型のようにインデックス情報を管理するサーバは存在せず、各ピアが自律分散的に動作し、サービス検索を行う。ピア P2P 型のシステムを適用したアプリケーションの代表例として、Freenet[3]、Gnutella[4] などがある。この P2P システムでは、各サービスを提供可能なピアに関する情報のネットワーク全体の一部分を、各ピアが管理する。また、各ピアはサービス要求(クエリ)を伝搬するために、物理ネットワークとは独立した論理的な検索ネットワークを構成する。サービスの要求は、検索クエリを論理ネットワーク上で隣接する他ピアへ送信し、要求サービスを提供可能なピアを発見する(クエリがヒットする)までこれを繰り返すことによって実現される。従って、実際の物理ネットワークのトポロジによっては、クエリの転送による検索時間が長くなってしまいう上、クエリを転送するホップ数に制限をかけた場合に、検索の成功を保證することができない。しかし、システム全体が完全な分散システムとなるため、クライアント数に対するスケラビリティ、高い耐障害性を実現することができる。

本研究では、P2P ネットワークのサービス利用者が急速に増加していることを考慮して、数万規模のユーザが参加するような大規模なデータ共有サービスを想定する。そのため、クライアント数に対するスケラビリティの高いピア P2P 型のデータ共有を想定する。

ピア P2P 型ネットワークを用いたデータ共有では、各ピアがデータ検索におけるクエリの転送や、データ提供の役割を担う。したがって、検索における負荷分散、検索効率およびデータの可用性向上のために、ネットワーク上の複数のピアにデータの複製を配置することが一般的である。Cohen らはピア P2P 型ネットワークにおいて、各データに対するアクセス頻度が与えられた場合、ネットワーク全体に配置すべき各データの複製数について議論している [2]。その中で、ピアが保持すべき各データの複製数の比を、ピアの各データに対するアクセス頻度の平方根の比と等しくすることで、ネットワークトラフィックや各ピアの負荷分散、検索効率において最も良い性能が得られることを証明している。また Cohen らが証明した理想的な数の複製を配置するための複製配置手法に関する研究も行われている [5]。この研究では、クエリがヒットした際に、クエリを発行したピアから要求データを持つピアまでにクエリが経由した全てのピアに複製を配置する“パス複製法”(図 1)を用いることで、理想数に近い数の複製を配置できることを確認している。しかし、3 章で

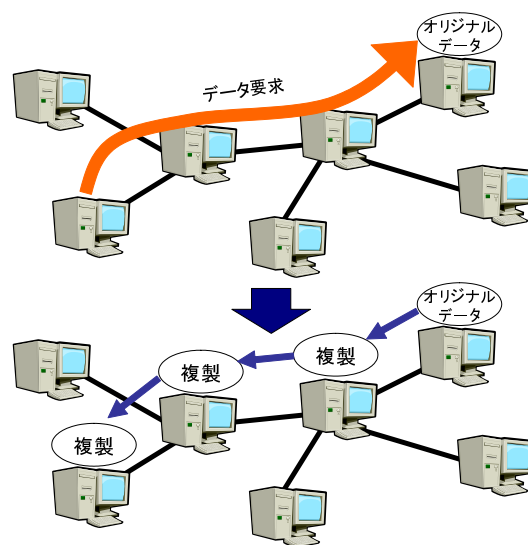


図 1: パス複製法を用いた複製配置

示すように、パス複製法によって作成される複製数は理想値に十分近いとは言えず、改善の余地がある。

そこで、本稿では、P2P ネットワークにおける検索効率の向上と負荷分散を目的とした複製配置手法を提案する。提案手法は、パス複製法と同様に検索クエリが経由したピアを対象として複製の配置を行う。このとき、提案手法では、複製の配置対象となる各ピアにおいて、データに対するアクセス頻度の順位を考慮して、複製を作成するか否かを決定することで、パス複製法よりもさらに理想値に近い数の複製をネットワーク上に配置する。

以下では、2 章で関連研究について述べ、3 章で提案手法を着想するに至った準備実験について説明する。4 章で提案手法について説明し、提案手法の性能評価を行う。最後に 5 章で本稿のまとめを行う。

## 2 関連研究

本章では、ピア P2P 型ネットワークを用いたデータ共有に関する代表的な研究として、検索ネットワークのトポロジを作成するためのプロトコルや、検索手法、複製の配置手法を説明し、本研究との関連性について議論する。

### 2.1 構造型トポロジを用いた P2P システム

構造型トポロジを用いた P2P システムは、DHT (Distributed Hash Table) と呼ばれる方法を用いて、ネットワークのトポロジやデータを検索するために必要なキー配置を厳密に定義するシステムである。構造型トポロジを用いた P2P システムとしては、Content-Addressable Network(CAN)[7]、

Pastry[8], Chord[9] などがある。この P2P システムは、ピアの識別子にハッシュ関数を適用してハッシュ値を算出し、それらのあるアドレス空間の一点に対応付ける。また、ピアが保持する各データのキー情報も同様にデータ名などをもとにハッシュ値を求め、アドレス空間内で最も近いピアにそのデータを格納する。データを検索する際には、検索対象のデータのハッシュ値を算出することで、そのハッシュ値に最も近いピアにクエリを転送することが可能になり、要求するデータのキー情報を得ることができる。このように、ネットワークのトポロジやクエリの転送方法を明確に定義することにより、検索を効率化している。

しかし、構造型トポロジを用いた P2P システムは、ピアの退出や故障などに対するネットワーク耐性の検証が完全とは言い難い。これは現状において実運用されているピア P2P 型ネットワークを用いたサービスのほとんどが、非構造型トポロジを用いた P2P システム [3][4] であることから明らかである。そのため、本研究では、非構造型トポロジを用いた P2P システムを想定する。

## 2.2 非構造型トポロジを用いた P2P システム

非構造型トポロジを用いた P2P システムは、ネットワークのトポロジ及びデータの配置に明確な方針が規定されていない。そのため、データ検索の際には、フラッディングなどの無作為な方法が用いられる。フラッディングでは、あるピアが要求データを検索する場合、TTL(Time To Live) と呼ばれる値を定めたクエリを全ての隣接ピアに送信する。このとき、クエリを受け取ったピアが、そのデータを保持していない場合は、クエリを発行したピアから論理ネットワーク上のホップ数が TTL 値を超えないかぎり、さらにそのピアの隣接ピアへクエリを転送する。データを保持している場合は、クエリが送られてきたピアを逆順にたどって、クエリを発行したピアに、返信情報を送信する。このような無作為な検索は、トラフィックや検索にかかる遅延が大きい。しかし、任意の論理ネットワークを構築できるため、システムを運用しやすいといった利点がある。

また、Adamic らは、無作為に構築された非構造型 P2P ネットワークの構成が、べき法則 (Power-Law) に従うことを示している [1]。ここで、ネットワーク上に存在する全ピアを隣接ピア数の降順に並べ、 $i$  番目のピアにおける隣接ピアの数を  $d_i$  で表す。 $d_i$  の分布が  $d_i \propto 1/i^\beta$  ( $\beta \geq 0$ ) で表されるとき、このネットワークはべき法則に従うという。

Cohen らは、文献 [2] において、非構造型トポロジ

を用いた P2P ネットワークを想定し、ネットワーク全体に配置すべき各データの複製数について議論している。非構造型トポロジを用いた P2P ネットワークの場合、フラッディングなどの無作為な検索を行う必要があるため、クエリはヒットするか、もしくはホップ数が定められた TTL を超えない限り、隣接ピアに転送され続ける。したがって、複製数が少ないデータに対して要求が頻繁に行われる場合、クエリがヒットする確率が少なくなるため、トラフィックは増加してしまう。そこで、一度の検索でクエリが伝播する範囲を平均探索サイズ (ESS) と定義し、ネットワーク全体における各データの複製数の比を、データに対するアクセス確率の平方根の比と等しくする平方根配置モデルにより、ESS が最も小さくなることを証明している。つまり、複製数の比を平方根配置モデルに近づけることで、ネットワークやピアの負荷を軽減することができ、検索効率も向上できる。

Lv らは文献 [5] において、平方根配置モデルの実現を目的とした複製配置手法として、オーナー複製法、パス複製法、ランダム複製法に関する考察を行っている。これらの複製配置手法は、検索クエリがヒットした場合に複製を作成する。また、各ピアにおいて、複製を作成できる上限数を超えてしまった場合は、保持する複製の中から FIFO (First In First Out) もしくはランダムに一つを選択して削除する。

- オーナー複製法: Gnutella で用いられる手法であり、クエリの発行元ピアがデータを受信した際に、その複製を配置する。
- パス複製法: Freenet で用いられる手法であり、クエリの発行元ピアから、要求データを保持するピアまでに、クエリが経由したパス上にある全てのピアにその複製を配置する。
- ランダム複製法: クエリが、要求データを保持するピアを発見するまでに経由したパス上にあるピア数と同じ数の複製を、検索によりクエリが伝搬した全てのピアの中から、ランダムに選んで配置する。ネットワーク上に配置される複製の数はパス複製法と同じになる。

ここで、ランダム複製法では、クエリの発行元ピアから複製を配置すべきピアへ、その旨を伝える情報を送信し、データを転送しなければならない。したがってネットワークに大きな負荷をかけてしまうため、他の手法との正当な比較を行えない。そのため、本稿ではランダム複製法は想定しない。

文献 [5] では、パス複製法を用いることにより、各データの複製数の比を平方根配置モデルに近づけら

れることを示している．しかし，パス複製法によって作成される複製数は，平方根配置モデルに基づいた複製数に十分近いとは言い難い．詳しくは，後述の3.2節で議論する．

### 3 準備実験

本章では，各データの複製数を平方根配置モデルに近づけることによる，検索性能への影響を調査するために行った準備実験とその結果を示す．

#### 3.1 シミュレーション環境

P2P ネットワークに参加するピア数は3000とした．使用するネットワークは，べき法則に従って各ピアの隣接ピア数を決定し，ピア同士をランダムに接続したPLRG(Power-Law Random Graph)とした．本稿では，隣接ピア数を降順に並べたときの*i*番目のピアの隣接ピア数を*d<sub>i</sub>*とし，*d<sub>i</sub>*を次式を用いて与えた．

$$d_i = \lfloor \omega / i^\beta \rfloor \quad (\beta \geq 0) \quad (1)$$

ここで， $\omega$ は最大隣接ピア数を示す．また， $\beta$ はネットワークの粗密を決定するパラメータである．検索対象となる総データ数は500とし，初期状態では，全ピアの中から各データのオリジナルデータを所持するピアを1つ，複製データを所持するピアを2つ，いずれもランダムに選択した．本シミュレーション環境では，オリジナルデータの更新は発生しないものとし，複製データは常にオリジナルデータと同じ内容であるとした．さらに，各データのサイズは全て等しく，各ピアがオリジナルデータおよび複製を保持可能な数を5とした．また，オリジナルデータは削除されないものとした．ピアの参加・退出は考慮しない．

各ピアがあるデータに対して，クエリを発行する確率は全てのピアで等しいものとし，各タイムスロットで0.1とした．また，各ピアがデータ*j*に対してクエリを発行する確率*q<sub>j</sub>*は，以下のようなZipf法則を用いて与えた．ここで*k*は総データ数を表す．

$$q_j = \frac{j^{-\alpha}}{\sum_{m=1}^k m^{-\alpha}} \quad (2)$$

$\alpha$ は，アクセス確率を決定するためのZipf係数と呼ばれるものであり，この値が大きいほど一部のデータが頻繁に要求される．式(2)では，識別子が小さいデータに対して頻繁にクエリが発行される．

また，式(2)を用いた場合， $q_1 \geq q_2 \geq \dots \geq q_k$ となるので，データの識別子とデータアクセス頻度の順位は等しくなる．検索クエリの伝搬には，フラッディングを用い，TTLは4とした．検索において，

要求データを保持するピアを複数発見した場合，それらのうち論理ネットワーク上のホップ数が最も近いピアにアクセスするものとした．

以上のようなシステム環境において，4,000タイムスロットのシミュレーション実験を行った．

#### 3.2 検索に与える影響の調査

パス複製法を用いた場合のネットワーク全体の複製数の分布を調べた結果を図2に示す．ここで， $\omega$ を500， $\beta$ を0.8とし，Zipf係数 $\alpha$ を1.0とした．また，グラフの縦軸が総複製数に対する各データの複製数の割合，横軸がデータの識別子を表す．両者共に対数表示している．さらに，平方根配置モデルに基づく理想値を“square root”とし，パス複製法にもとづいた結果を“path”と示している．

図1から，パス複製法は平方根配置モデルに対して，識別子が50以下のデータの複製が多く配置されている．特に，識別子が小さくなるにつれてその差は大きくなっている．反対に，識別子が50以上のデータに関しては，識別子が大きくなるにつれて，平方根配置モデルよりも複製数が少なくなっている．これは，人気のあるデータに対してクエリが頻繁に発行されるため，パス複製法では，それらの複製が非常に多くネットワーク上に配置されるためである．つまり，クエリの発行頻度の低いデータは，その影響を受け，削除される確率が高くなる．

複製数を平方根配置モデルに近づけるためには，アクセス頻度の高いデータの複製によってアクセス頻度の低いデータの複製が削除されすぎないようにする必要がある．そこで本研究では，パス複製法のように要求データを保持するピアからクエリの経由した全てのピアに複製を配置するのではなく，検索対象となる各データのアクセス頻度の順位に応じて複製を配置する確率を変化させることを考え，下記のような試行を行った．

試行1:

クエリのパス上のピアにおいて，複製を配置する確率(複製配置確率)*R<sub>j</sub>*を次の式で与えた．

$$R_j = \frac{j}{k} \quad (3)$$

式(3)を用いた場合の，ネットワーク全体の複製数の分布を調べた結果を図3に示す．この結果より，式(3)を用いた場合には，全ての複製数がほぼ同じになっている．したがって，式(3)では，複製配置確率の差が大きすぎることが分かる．

試行2:

試行1の結果から，平方根配置モデルに近づけるためには，式(3)よりもアクセス頻度の高いデータの複製配置確率を高くする必要があると考え，複製配置確率*R<sub>j</sub>*を次式で与えた．

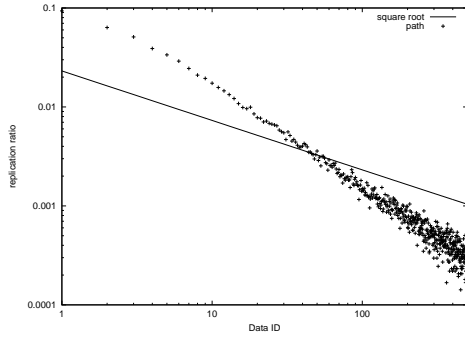


図 2: パス複製法における複製の分布

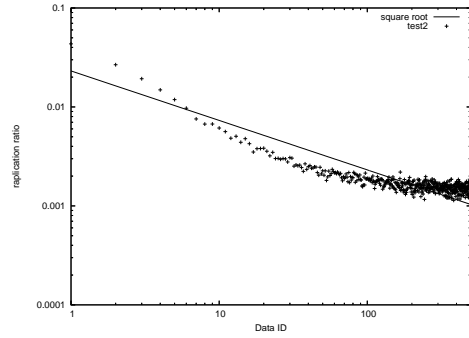


図 4: 式 (4) を用いた場合の複製の分布

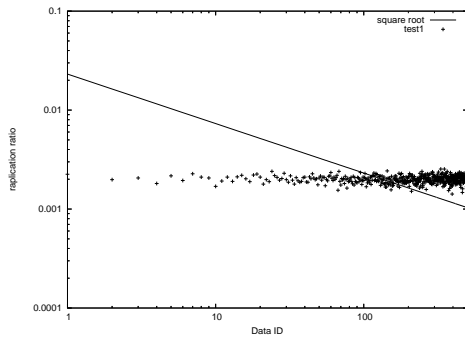


図 3: 式 (3) を用いた場合の複製の分布

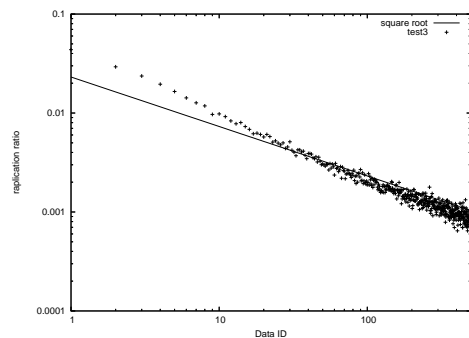


図 5: 式 (5) を用いた場合の複製の分布

$$R_j = 0.05 + 0.95 \times \frac{j}{k} \quad (4)$$

式 (4) は、アクセス頻度が最下位のデータの複製は必ず作成され ( $R_j = 1$ )、アクセス頻度の最上位のデータの複製もある程度は作成されるように考慮した ( $R_j \geq 0.05$ ) ものである。式 (4) を用いた場合の、ネットワーク全体の複製数の分布を調べた結果を図 4 に示す。この結果より、式 (4) を用いた場合には、アクセス頻度が高いデータと低いデータの複製は、平方根配置モデルよりも多く作られ、アクセス頻度が中間のデータの複製は、平方根配置モデルより少なくなることが分かる。

試行 3:

試行 2 の結果を考慮して、式 (4) の問題を補正するために、各データの複製配置確率  $R_j$  を次の式で与えた。

$$R_j = 0.05 + 0.95 \times \left(\frac{j}{k}\right)^{\frac{1}{2}} \quad (5)$$

式 (5) を用いた場合の、ネットワーク全体の複製数の分布を調べた結果を図 5 に示す。この結果より、式 (5) を用いた場合、これまでの試行に比べて、より平方根配置モデルに近づいていることが分かる。つまり、アクセス頻度の高いデータの複製配置確率を適切に低くすることで、アクセス頻度の低いデー

タの複製配置確率が相対的に高くなっている。そのため、パス複製法に比べてアクセス頻度の高いデータの複製数が減少し、逆にアクセス頻度の低いデータの複製数が増加している。

次に、平方根配置モデルに近づけたときのデータ検索性能への影響を調べるために、パス複製法および式 (5) を用いた複製法における検索成功率を、それぞれ図 6、図 7 に示す。ここで、グラフの縦軸は、各データに対するクエリがヒットした割合（検索成功率）を表し、横軸はデータの識別子を表す。これらの結果を比較すると、アクセス頻度が高いデータの検索成功率は両者に大きな差は見られないが、パス複製法を用いた場合、アクセス頻度が低いデータに検索成功率の低いものが目立つ。それに対し、式 (5) を用いた複製法では、全てのデータに対して高い検索成功率を示している。このように、式 (5) を用いた複製法は、パス複製法と比べて、アクセス頻度が高いデータの複製配置確率を低くすることで、アクセス頻度の低いデータの検索成功率を向上させている。これらの結果から、ネットワーク全体における複製数の比を平方根配置モデルに近づけたとき、検索効率、特にアクセス頻度の順位が下位のものに対する検索成功率が向上することが分かる。

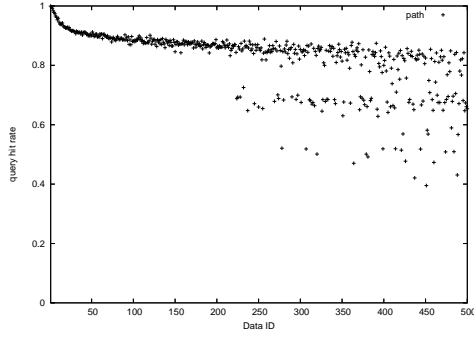


図 6: パス複製法による検索成功率

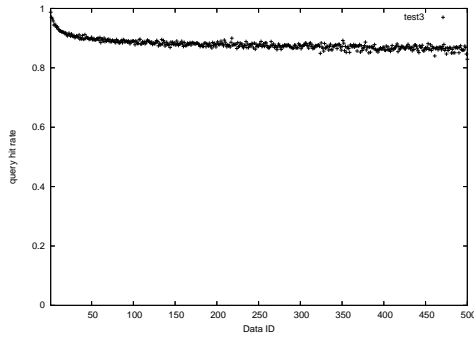


図 7: 式 (5) を用いた場合の検索成功率

## 4 アクセス頻度を考慮した複製配置手法

本章では、複製数の比を平方根配置モデルに近づけることを目的とし、検索対象データのアクセス頻度の順位を考慮した複製配置手法を提案する。さらに、提案手法の性能評価を行う。

### 4.1 提案手法

提案手法では、パス複製法と同様に、クエリの発行元ピアから、要求データを保持するピアまでクエリが経由した全てのピアを、複製配置の対象のピアとする。このとき、配置対象の各ピアにおいて複製を作成するか否かを決定するために、データに対するアクセス頻度の順位に基づいた確率を用いる。複製を配置する確率  $R_j$  は、3章の準備実験の結果を参考にして、次式で与える。

$$R_j = a + (1.00 - a) \left(\frac{j}{k}\right)^c \quad (6)$$

ここで  $a$  と  $c$  は、アクセス頻度が高いデータと低いデータの複製配置確率の差を決定するパラメータである。 $a$  や  $c$  の最適値は、システム環境に依存して変化する。各パラメータの最適値の検証は、4.2.1 節で行う。

### 4.2 性能評価

本稿では、提案手法において、各データの複製数の比が平方根配置モデルにどれだけ近いかを表す指標として、次式で表される評価値  $V$  を用いる。

$$V = \frac{1}{k} \sum_{j=1}^k \frac{|r_j - s_j|}{s_j} \quad (7)$$

ここで、データ  $j$  の複製数の比を  $r_j$ 、平方根配置モデルにおける複製数の比を  $s_j$  としている。評価値  $V$  が小さいほど、複製数の比が平方根配置モデルに近いことを示す。

#### 4.2.1 アクセス分布と複製配置確率の関係

アクセス確率の分布を決定するパラメータ  $\alpha$  を変化させたときに、評価値  $V$  が最も小さくなる  $a$  と  $c$  の値をシミュレーション実験によって調べた。ここで、 $\alpha$  を 0.0 から 1.2 (0.1 刻み)、 $a$  を 0.00 から 1.00 (0.01 刻み)、 $c$  を 0.00 から 1.00 (0.05 刻み) の範囲で変化させた。また、 $\omega$  を 500、 $\beta$  を 0.6 とする PLRG 型 P2P ネットワークを用いた。

実験結果を表 1 に示す。この結果から、アクセス頻度の分布を決定する  $\alpha$  と、 $a$  および  $c$  には相関性があることが分かる。 $\alpha$  が小さい場合、 $a$  が比較的大きい値の方が平方根配置モデルに近くなり、逆に  $\alpha$  が大きくなるにつれて  $a$  が小さい値の方が近くなる。また  $c$  については、 $\alpha$  が小さいときは小さく、 $\alpha$  が大きくなるにつれて  $c$  も大きくすることで平方根配置モデルに近くなること分かる。

ここで、参考のために、パス複製法を用いた場合の、評価値  $V$  に対する  $\alpha$  の影響を図 8 に示す。この結果より、 $\alpha$  が小さいときは、平方根配置モデルに近い数の複製を配置しており、 $\alpha$  が大きくなるにつれて、平方根配置モデルから遠ざかることが分かる。つまり、パス複製法による配置は、 $\alpha$  が小さいときは良い性能を示すが、 $\alpha$  が大きくなるにつれて、性能低下が著しい。そのため、提案手法を用いた場合、 $\alpha$  が大きいときは、アクセス頻度の高いデータと低いデータの複製配置確率を大きく調整した方が良い結果を示している。

ここで、 $\alpha$  の値と良好な性能を与える  $a$ 、 $c$  の値の関係について検討する。表 1 の結果から、最良の性能を示す  $a$  と  $c$  は、 $\alpha$  を用いて次式のように近似できる。

$$a = \frac{1}{(\alpha \times 10 + 1)^{\sqrt{2}}} \quad (8)$$

$$c = \frac{\alpha}{2} \quad (9)$$

これらの近似式を用いて  $a$ 、 $c$  を決定した場合の評価値  $V$  を表 2 に示す。表 2 の結果より、近似値を用いた場合の評価値  $V$  は最良値に十分近いものと判

表 1: 各  $\alpha$  における最適な  $a, c$  と評価値  $V$

$\alpha$	$a$	$c$	$V$
0.0	0.96	0.05	0.0682
0.1	0.68	0.20	0.0676
0.2	0.34	0.15	0.0691
0.3	0.01	0.15	0.0711
0.4	0.04	0.20	0.0725
0.5	0.00	0.25	0.0747
0.6	0.06	0.30	0.0770
0.7	0.04	0.35	0.0820
0.8	0.07	0.40	0.0890
0.9	0.02	0.40	0.0938
1.0	0.02	0.45	0.1029
1.1	0.02	0.55	0.1124
1.2	0.02	0.60	0.1227

表 2: 各  $\alpha$  における近似値  $a, c$  と評価値  $V$

$\alpha$	$a$	$c$	$V$
0.0	1.00	0.00	0.0701
0.1	0.38	0.05	0.0748
0.2	0.21	0.10	0.0750
0.3	0.14	0.15	0.0757
0.4	0.10	0.20	0.0781
0.5	0.08	0.25	0.0793
0.6	0.06	0.30	0.0770
0.7	0.05	0.35	0.0877
0.8	0.04	0.40	0.0911
0.9	0.04	0.45	0.1028
1.0	0.03	0.50	0.1032
1.1	0.03	0.55	0.1191
1.2	0.03	0.60	0.1326

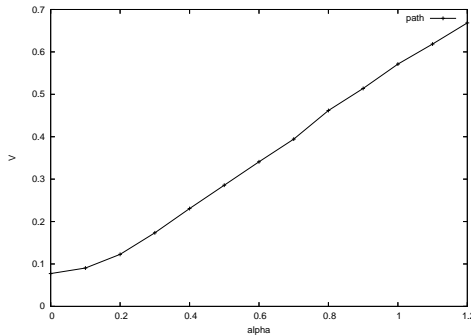


図 8:  $\alpha$  と評価値  $V$  (パス複製法)

断できる．したがって，以降のシミュレーション実験ではこれらの近似式を用いて， $a, c$  を決定した．

これらの近似式は，今回のシミュレーション環境にのみ適用可能であり，環境が変われば，近似式も異なるものとなる．しかし，アクセス頻度が Zipf 分布で与えられる環境であれば，本稿のような準備実験を行うことにより，その環境に適した  $a, c$  の近似式を算出できるものと考えられる．

#### 4.2.2 提案手法の優位性の検証

実際の検索における提案手法の優位性を検証するため，TTL を 1 から 7 まで変化させたときの，検索成功率と各ピアが受信した平均クエリ数をシミュレーション実験により調べた．ここで， $\omega$  は 30， $\beta$  は 0.4 とした．データアクセス頻度を決定する Zipf 係数  $\alpha = 0.4$  のときの検索成功率を図 9，平均クエリ数を図 10 に示し， $\alpha = 1.2$  のときの結果を図 11，図 12 に示す．また，各図において，パス複製法を用いた場合の結果を“path”，提案手法を用いた場合の結果を“rank”と表している．

図 9 と図 11 から，TTL が 1 の場合を除いて，提案手法を用いた場合の方が，パス複製法よりも検索成功率が高いことが分かる．特に， $\alpha = 1.2$  の場合，パス複製法に対する性能の差が大きい．これは，パス複製法では， $\alpha$  が大きくなるにつれて複製配置数の比が平方根配置モデルに対して大きくなるのに対して，提案手法が常に平方根配置モデルに近い配置を行えるためである．一方，図 10 と図 12 から，各ピアが受信した平均クエリ数は，提案手法とパス複製法でほとんど変わらないことが分かる．つまり，提案手法はトラフィックや各ピアへの負荷を増加させることなく，検索効率を向上させている．

以上の結果から，提案手法では，複製配置数の比を平方根配置モデルに近づけることにより，検索性能を向上させることを確認した．

## 5 まとめ

本稿では，非構造型 P2P ネットワークにおけるデータ共有サービスを想定し，負荷分散と検索効率の向上を目的とした複製配置手法を提案した．提案手法では，ピアにおける複製の配置確率を，各データに対するアクセス頻度の順位を考慮して決定することにより，パス複製法よりも最適値により近い数の複製を作成できる．さらに本稿では，シミュレーション実験により，提案手法の性能評価を行った．その結果から提案手法は，パス複製法よりも検索成功率の点で良い結果を示すことを確認した．

提案手法は，ピアの各データに対するアクセス確率の分布は既知のものとしているが，実環境におけるピア P2P 型ネットワークでは，これらの情報を取得することは容易ではない．そこで今後は，周辺

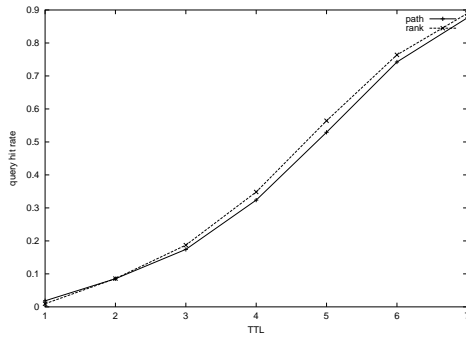


図 9: TTL と検索成功率 ( $\alpha = 0.4$ )

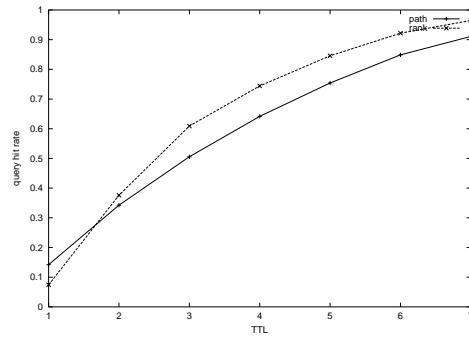


図 11: TTL と検索成功率 ( $\alpha = 1.2$ )

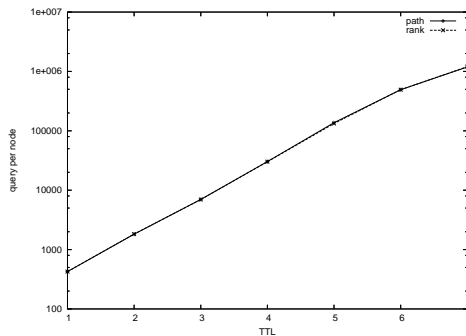


図 10: TTL と平均負荷 ( $\alpha = 0.4$ )

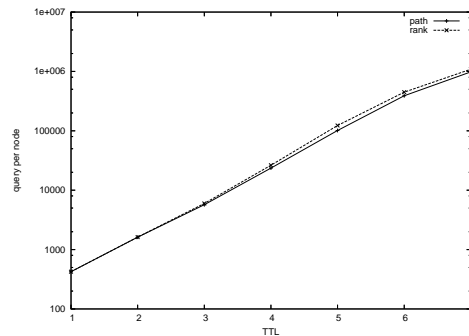


図 12: TTL と平均負荷 ( $\alpha = 1.2$ )

のピアから送信される検索クエリを監視し、アクセス確率の分布を推測することで、自律的に複製の配置確率を決定する手法について検討する予定である。

文献 [5] において Lv らは、ネットワークトラフィックの負荷減少を目的として、エキスパンディングリングやランダムウォークを提案している。これらの手法では、フラッディングの TTL を初めは少なく設定し徐々に大きくしたり、隣接ピアのうち、いくつかのピアに限ってクエリを送信することにより、フラッディングのトラフィックを軽減している。今後は、このような手法を用いた場合でも高い検索成功率を得られるように、本稿の提案手法を拡張する予定である。

謝辞 本研究は、文部科学省 21 世紀 COE プログラム「ネットワーク共生環境を築く情報技術の創出」および文部科学省特定領域研究 (16016260) の研究助成によるものである。ここに記して謝意を表す。

## 参考文献

[1] L.A. Adamic, R.M. Lukose, A.R. Puniyani, and B.A. Huberman, “Search in Power-Law Networks,” *Physical Review E*, Vol 64, No. 4, 046135, 2001.

[2] E. Cohen and S. Shenker, “Replication Strategies in Unstructured Peer-to-Peer Networks,” *Proc. ACM SIGCOMM’02*, pp.177–190, 2002.

[3] Freenet, <URL: <http://freenetproject.org>>.

[4] Gnutella, <URL: <http://www.gnutella.com>>.

[5] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker, “Search and Replication in Unstructured Peer-to-Peer Networks,” *Proc. 16th Int’l Conf. on Supercomputing*, pp.84–95, 2002.

[6] Napster, <URL: <http://www.napster.com>>.

[7] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, “A Scalable Content-Addressable Network,” *Proc. ACM SIGCOMM’01*, pp.161–172, 2001.

[8] A. Rowstron and P. Druschel, “Pastry: Scalable, Distributed Object Location and Routing for Large-scale Peer-to-Peer Systems,” *Proc. IFIP/ACM Middleware 2001*, pp.329–350, 2001.

[9] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, “Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications,” *Proc. ACM SIGCOMM’01*, pp.149–160, 2001.