Vol.2011-OS-118 No.15
2011/7/28

# 構造化オーバレイにおける経路表の順序関係に基づくネットワーク近接性の考慮手法

宮尾武裕[†1]　長尾洋也[†1]　首藤一幸[†1]

Chord をはじめとする従来の分散ハッシュテーブル（DHT）では，経路表の管理をランダムに割り当てられたノード ID に基づき厳密に行っている．そのため，ノード間の通信遅延といったノード ID 以外の情報を考慮することが難しい．そこで，本研究ではノード ID 以外を考慮しやすいルーティングアルゴリズム設計方式である柔軟な経路表（FRT）に基づき，ネットワーク近接性を考慮したアルゴリズムを設計する手法（PFRT）を開発した．PFRT は，FRT における経路表の順序関係 $\leq_{\mathrm{ID}}$ を拡張した，ネットワーク近接性を考慮した順序関係 $\leq_{\mathrm{ID+PR}}$ を用いる．PFRT に基づくルーティングアルゴリズムは，この順序関係 $\leq_{\mathrm{ID+PR}}$ に基づき経路表の改良を繰り返すことで経路表を構築していく．PFRT に基づく DHT アルゴリズム PFRT-Chord は，Chord よりルーティングの平均所要時間が 40% 減少し，PFRT の有効性を示すことができた．

# Proximity-aware Structured Overlays Based on an Order of Routing Tables

Takehiro Miyao,[†1] Hiroya Nagao[†1]
and Kazuyuki Shudo[†1]

Existing DHT algorithms such as Chord prescribe which nodes are held in a routing table based on the nodes' ID. The ID restriction is a major obstacle to exploitation of network proximity. We designed Proximity-aware Flexible Routing Tables (PFRT), a routing algorithm designing methodology by expanding Flexible Routing Tables (FRT). We define an order of routing tables $\leq_{\mathrm{ID+PR}}$ for PFRT based on FRT's basic order $\leq_{\mathrm{ID}}$. PFRT updates a routing table repeatedly according to order $\leq_{\mathrm{ID+PR}}$. We designed PFRT-Chord, a DHT algorithm based on PFRT. An experiment result shows that average routing latency in PFRT-Chord is about 40% less than Chord.

## 1. はじめに

Gnutella や BitTorrent などのさまざまな P2P アプリケーションが存在する．P2P システムはスケーラビリティや耐故障性などの点で優れている．P2P ネットワークのように，物理的に接続された実ネットワーク上に独自のトポロジで構築されたアプリケーションレベルのネットワークをオーバレイネットワークと呼ぶ．オーバレイネットワークによって，非集中であっても多数のノードをうまく連携させることができる．

オーバレイネットワークは二つに分類することができる．一つは非構造化オーバレイネットワークであり，もう一つは構造化オーバレイネットワークである．非構造化オーバレイネットワークでは，オーバレイネットワーク構築時のルールが一定の数学的ルールに従わずにネットワークを構築し，その技術は Gnutella や BitTorrent などのアプリケーションに用いられている．逆に，構造化オーバレイネットワークでは一定の数学的ルールに従いネットワークを構築する．

構造化オーバレイネットワークには Chord[1)]，Kademlia[2)]，Pastry[3)] などのアルゴリズムが存在し，近年多くの研究がなされている．これらのアルゴリズムは分散ハッシュテーブル（Distributed Hash Table，DHT）と呼ばれる．DHT とは連想配列を複数のノードで管理する技術である．非構造化オーバレイネットワークでは発見のためにフラッディング等が必要となるが，これらの DHT では，ノード数を $N$ とすると $O(\log N)$ の経路長で効率のよいルーティングを行うことが可能である．大規模な P2P システムではノードが世界中に存在するため，一対一でのノード間の通信遅延はさまざまなものとなる．しかし，従来の DHT の多くではハッシュ関数で割り当てられたノード ID に基づくルーティングを行うため，ネットワーク近接性が考慮されていない．その結果，ルーティングにおいて通信遅延が大きなノードと通信を行うことが頻繁に起こり，ルーティングにおける通信遅延を小さくすることができないという問題がある．

ネットワーク近接性を考慮するための方針として，Proximity Neighbor Selection（PNS），Proximity Route Selection（PRS），Proximity Identifier Selection（PIS）が提案されてきた[4)]．PNS は経路表構築時に，PRS は経路選択時に，PIS はノード ID 決定時にそれぞ

†1 東京工業大学
Tokyo Institute of Technology

れネットワーク近接性を考慮する方法である．本研究では，PNS に基づいてネットワーク近接性を考慮する．

Chord，Kademlia といった従来の DHT では，ノード ID のみによって経路表に追加するノードを厳密に選択しているので，PNS に基づいてネットワーク近接性を考慮することが難しい．そこで，ノード ID 以外を考慮しやすいルーティングアルゴリズム設計方式である柔軟な経路表（Flexible Routing Tables, FRT）[5] を利用する．FRT では，ID による経路表の順序関係 $\leq_{\mathrm{ID}}$ に基づき経路表の改良を繰り返す．FRT における経路表の順序関係 $\leq_{\mathrm{ID}}$ を拡張して，ネットワーク近接性を考慮した経路表の順序関係 $\leq_{\mathrm{ID+PR}}$ を定義し，$\leq_{\mathrm{ID+PR}}$ に基づき経路表の改良を繰り返すネットワーク近接性を考慮した柔軟な経路表（Proximity-aware Flexible Routing Tables，PFRT）を提案する．

さらに，PFRT に基づき FRT-Chord を拡張し，PFRT-Chord の実装・評価を行った．PFRT-Chord は FRT-Chord に対して，高々数ステップの拡張を行うことで実装することが可能である．また，PFRT-Chord ではルーティングの平均所要時間が Chord より約 40% 減少した．そのため，PFRT は FRT に基づく DHT アルゴリズムにおいて，ネットワーク近接性を考慮する手法として有効であることを示すことができた．

## 2. 関 連 研 究

本研究の関連研究を次に示す．

### 2.1 Chord[1)]

Chord は DHT アルゴリズムのひとつである．$m$ ビットのリング上の ID 空間にハッシュ関数により ID が割り当てられたノードを配置する．key と value を持つデータもまた key に基づき ID が割り当てられ，ID 空間に配置される．ID 空間上でデータ ID から時計回りに進んだ時，一番最初に出会うノードを担当ノードと呼び，担当ノードにそのデータを保存する．データを取り出すときは，目的のデータの key から ID を生成し，そのデータの担当ノードへたどり着くまで，各ノードが所持している経路表に基づきノードへの転送（ホップ）を繰り返す．

Chord には，successor list，predecessor，finger table の 3 つの経路表がある．ID 空間上で自ノードから時計回りに進んだ時，最初に出会うノードを successor と呼び，successor list とは時計回りに進んで出会うノードのノード情報（エントリ）を順番に一定数 $c$ だけ保持する経路表である．反時計回りに進んだ時，最初に出会うノードを predecessor と呼び，そのノードをエントリとして経路表に保持する．finger table とは自ノード $s$ から $2^i$（$i = 0, 1, ..., m-1$）の距離だけ離れた ID の担当ノードを $i$ 番エントリとして保持する経路表である．ノード $x$ からノード $y$ の ID 距離 $d(x, y)$ を次のように定義する．

$$d(x,y) = \begin{cases} y - x, & x < y \text{ のとき} \\ 2^m, & x = y \text{ のとき} \\ y - x + 2^m, & y < x \text{ のとき} \end{cases} \quad (1)$$

Chord は，successor によってルーティングにおける目的ノードへの到達性を保証し，finger table によってルーティングにおけるノードの転送回数（経路長）を減少させる．これらの経路表により，Chord はノード数を $N$ とすると，ルーティングの経路長が $O(\log N)$ で目的ノードへ到達することが可能となる．

### 2.2 LPRS-Chord[6)]

LPRS-Chord はネットワーク近接性を考慮するように Chord を拡張した DHT アルゴリズムである．LPRS-Chord では，PNS に基づく拡張，つまり経路表構築時に拡張を行っている．

Chord の 3 つの経路表のうち finger table をネットワーク近接性を考慮するように拡張する．つまり，successor によってルーティングにおける目的ノードへの到達性を保証しつつ，拡張した finger table により Chord と同等の経路長を達成し，またルーティングにおける 1 回の転送（1 ホップ）あたりの平均所要時間を減少させる．その結果，LPRS-Chord におけるルーティングの平均所要時間が減少する．LPRS-Chord の拡張した finger table は次の通りである．

(1) 他のノードと通信を行った時，ノード間の通信遅延を計測する．
(2) $2^i \leq d(s,n) < 2^{i+1}$ を満たすノード $n$ のうち，自ノード $s$ との通信遅延の最も小さいノードを finger table の $i$ 番エントリとして保持する．

### 2.3 柔軟な経路表[5)]

柔軟な経路表（Flexible Routing Tables，FRT）は，ID による経路表の順序関係 $\leq_{\mathrm{ID}}$ に基づき経路表の改良を繰り返すことで経路表を構築する方式である．

#### 2.3.1 FRT-Chord

FRT-Chord は FRT に基づき設計された DHT アルゴリズムである．Chord と同様に $m$ ビットのリング上の ID 空間を持ち，担当ノードの決定法も同じであるが，経路表の構築方法が異なる．FRT-Chord には，Chord の successor list，predecessor，finger table の 3 つの経路表を 1 つに統合した経路表 $E$ がある．

経路表 $E$ は経路表エントリ $e_i$ の集合 $\{e_i\}$ である．ノード $s$ の経路表エントリ $e_i$ はノード ID $e_i$.id および IP アドレス $e_i$.address を保持している．$d(s, e_i) < d(s, e_{i+1})(i = 1, 2, ..., |E|-1)$ を満たしている．このとき，$e_1$ は successor，$\{e_i\}_{i=1,2,...,c}$ は successor list，$e_{|E|}$ は predecessor を表している．

#### 2.3.1.1 $\leq_{\mathrm{ID}}$ の定義

FRT-Chord では，ノード $s$ が経路表エントリ $e_i$ へフォワーディングを行った際，残り ID 距離の短縮倍率の最悪値を $r_i(E)$ とする．

$$r_i(E) = \frac{d(e_i, e_{i+1})}{d(e_s, e_{i+1})} \quad (i = 1, 2, ..., |E| - 1) \tag{2}$$

$\{r_i(E)\}$ を降順に並べた列を $\{r_{(i)}(E)\}$ とすると，次のように ID に基づく経路表の順序関係を定義する．

$$E \leq_{\mathrm{ID}} F \Longleftrightarrow \{r_{(i)}(E)\} \leq_{\mathrm{dic}} \{r_{(i)}(F)\} \tag{3}$$

ただし，$\leq_{\mathrm{dic}}$ とは辞書式順序である．このとき，次の式を満たす経路表 $\tilde{E}$ を最良経路表と呼ぶ．

$$\tilde{E} \leq_{\mathrm{ID}} E \quad \text{for} \forall \mathrm{E} \tag{4}$$

#### 2.3.1.2 到達性保証操作

FRT-Chord では，ルーティングにおける目的ノードへの到達性を保証するための操作を到達性保証操作と呼ぶ．これは Chord の stabilize 操作にあたる操作である．この操作により successor を常に最新に保ち，Chord と同等の到達性を保証する．

#### 2.3.1.3 エントリ情報学習操作

FRT-Chord では，エントリを学習する操作をエントリ情報学習操作と呼ぶ．この操作により，ノードが知りえたエントリをすべて経路表に追加する．エントリ情報学習操作は次の操作に分けられる．

- ネットワーク参加時に他のノードから経路表の全エントリを学習する．
- ノードの探索時等において通信したノードを学習する．
- 能動学習探索を行い，通信したノードを学習する．

能動学習探索は，現在の経路表より優れた経路表を構築するために，必要なエントリを学習するルーティングである．現在の経路表から計算された ID へのルーティングを行う．

#### 2.3.1.4 エントリ厳選操作

FRT-Chord では，経路表 $E$ のエントリ数 $|E|$ が経路表サイズ $L$ を超えるとき，$|E| \leq L$ になるように経路表のエントリを厳選する操作をエントリ厳選操作と呼ぶ．エントリ厳選操作では，次の式を満たすエントリ $e_{\mathrm{remove}}$ の削除を行う．

$$E - \{e_{\mathrm{remove}}\} \leq_{\mathrm{ID}} E - \{e\} \quad \text{for} \forall \mathrm{e} \in \mathrm{E} \tag{5}$$

削除エントリ $e_{\mathrm{remove}}$ を効率よく見つけるために，次のように正規化間隔 $S_i^E$ を定義する．

$$S_i^E = \log \frac{d(s, e_{i+1})}{d(s, e_i)} \tag{6}$$

$e_{\mathrm{remove}}$ は $S_{i-1}^E + S_i^E$ が最小となるエントリである．$S_{i-1}^E + S_i^E$ を昇順として保持しておくことで，効率よく削除エントリ $e_{\mathrm{remove}}$ を見つけることができる．

FRT-Chord では，Chord における successor list および predecessor にあたる経路表エントリを sticky entry と呼び，エントリ厳選操作において経路表から削除しない．FRT-Chord におけるエントリ厳選操作は次の通りである．

(1) 削除候補のエントリ集合 $C$ に経路表 $E$ の全エントリを追加する．
(2) $C$ から sticky entry を取り除く．
(3) $C$ のエントリの中で $S_{i-1}^E + S_i^E$ が最小となるエントリ $e_i = e_{\mathrm{remove}}$ を経路表から削除する．

## 3. PFRT: ネットワーク近接性を考慮した柔軟な経路表

ネットワーク近接性を考慮した柔軟な経路表（Proximity-aware Flexible Routing Tables, PFRT）は，ネットワーク近接性を考慮した経路表の順序関係 $\leq_{\mathrm{ID+PR}}$ に基づき経路表の改良を繰り返すことで経路表を構築する方式である．ただし，$\leq_{\mathrm{ID+PR}}$ は，FRT の ID に基づく経路表の順序関係をネットワーク近接性を考慮するように拡張した順序関係である．

### 3.1 PFRT-Chord

PFRT-Chord は FRT-Chord を PFRT に基づき拡張した DHT アルゴリズムである．経路表を構築する方法以外は FRT-Chord のアルゴリズムを用いている．ノード $s$ の経路表エントリ $e_i$ は ID および IP アドレスの他にノード $s$ と $e_i$ との間の通信遅延 $e_i$.latency を保持している．

#### 3.1.1 $\leq_{\mathrm{ID+PR}}$: ネットワーク近接性を考慮した経路表の順序関係

PFRT では，経路長を短くするために ID を考慮するだけでなく，ノード間の通信遅延も考慮する．そこで，まずノード間の通信遅延に基づく経路表の順序関係 $\leq_{\mathrm{PR}}$ を定義する．次に，$\leq_{\mathrm{ID}}$ および $\leq_{\mathrm{PR}}$ を用いることで，ネットワーク近接性を考慮した経路表の順序関係 $\leq_{\mathrm{ID+PR}}$ を定義する．

#### 3.1.2 $\leq_{\mathrm{PR}}$: ノード間の通信遅延に基づく経路表の順序関係

$\leq_{\mathrm{PR}}$ はノード間の通信遅延に基づき経路表の優劣を表す順序関係である．$\leq_{\mathrm{PR}}$ において優れた経路表ほどルーティングにおける 1 ホップあたりの平均所要時間が小さくなる．ノード間の通信遅延に基づく経路表の順序関係 $\leq_{\mathrm{PR}}$ を次のように定義する．

$$E \leq_{\mathrm{PR}} F \Longleftrightarrow l(E) \leq l(F) \tag{7}$$

ただし，$l(E)$ は経路表 $E$ 上の各エントリ $e_i$ と自ノードとの通信遅延 $e_i$.latency の平均値であり，次のように定義される．

$$l(E) = \frac{1}{|E| - c - 1} \sum_{i=c+1}^{|E|-1} e_i.\mathrm{latency} \tag{8}$$

$l(E)$ には，sticky entry の通信遅延を含まない．

#### 3.1.3 $\leq_{\mathrm{ID+PR}}$ の定義

$\leq_{\mathrm{ID+PR}}$ はネットワーク近接性を考慮した経路表の優劣を表す順序関係である．$\leq_{\mathrm{ID+PR}}$ において優れた経路表ほどルーティングの平均所要時間が小さくなる．ルーティングの平均所要時間を減少させるための十分条件は，経路長を短くし，1 ホップあたりの平均所要時間を小さくすることである．そこで，$\leq_{\mathrm{ID+PR}}$ を次のように定義する．

$$E \leq_{\mathrm{ID+PR}} F \Longleftrightarrow (E \leq_{\mathrm{ID}} F) \cap (E \leq_{\mathrm{PR}} F) \tag{9}$$

$E \leq_{\mathrm{ID}} F$ および $F \leq_{\mathrm{PR}} E$ の両方を満たす経路表 $E, F$ が存在するとき、経路表 $E, F$ を $E \leq_{\mathrm{ID+PR}} F$ に基づき比較することができない。つまり、$\leq_{\mathrm{ID+PR}}$ は半順序関係である．

#### 3.1.4 到達性保証操作

FRT-Chord の到達性保証操作と同じである．

#### 3.1.5 エントリ情報学習操作

PFRT-Chord ではエントリの学習を行うために，FRT-Chord のエントリ情報学習操作を拡張した操作を行う．PFRT-Chord では，経路表エントリ $e_i$ に $e_i$.latency を追加しているので，エントリを学習する際に，$e_i$.latency を記録する。通信遅延が自然に得られない場合、ノード間の通信遅延を測定する。エントリ情報学習操作は次の通りである．

(1) エントリ情報学習操作を行う．
(2) 最新追加エントリ $e_{\mathrm{add}}$ を (1) で経路表に追加したエントリに更新する．

$e_{\mathrm{add}}$ はエントリ厳選操作において利用される．

#### 3.1.6 エントリ厳選操作

PFRT-Chord ではエントリの厳選を行うために，FRT-Chord のエントリ厳選操作を拡張した操作を行う．$\leq_{\mathrm{ID}}$ は全順序関係であるため，FRT-Chord では最善の削除するエントリ $e_{\mathrm{remove}}$ を決定することができた．しかし，$\leq_{\mathrm{ID+PR}}$ は半順序関係であるため，PFRT-Chord では一般に最善の削除エントリ $e_{\mathrm{remove}}$ を決定することができない．そこで，まず経路表エントリのうち，$\leq_{\mathrm{ID+PR}}$ および $e_{\mathrm{add}}$ を用いて削除するエントリの候補集合を $E_{\leq_{\mathrm{ID+PR}},\mathrm{add}}$ に絞る．

$$E_{\leq_{\mathrm{ID+PR}},\mathrm{add}} = \{e \in E \mid E - \{e\} \leq_{\mathrm{ID+PR}} E - \{e_{\mathrm{add}}\}\} \tag{10}$$

次に，$E_{\leq_{\mathrm{ID+PR}},\mathrm{add}}$ の中で $\leq_{\mathrm{ID}}$ を用いて削除するエントリ $e_{\mathrm{remove}} \in E_{\leq_{\mathrm{ID+PR}},\mathrm{add}}$ を決定する．

$$E - \{e_{\mathrm{remove}}\} \leq_{\mathrm{ID}} E - \{e\} \quad \text{for}\, \forall \mathrm{e} \in \mathrm{E}_{\leq_{\mathrm{ID+PR}},\mathrm{add}} \tag{11}$$

$e_{\mathrm{add}} \in E_{\leq_{\mathrm{ID+PR}},\mathrm{add}}$ より $E_{\leq_{\mathrm{ID+PR}},\mathrm{add}} \neq \phi$ である．つまり，$e_{\mathrm{remove}}$ は必ず存在する．$e_{\mathrm{remove}}$ を削除した経路表は，最新追加エントリ $e_{\mathrm{add}}$ を追加する前の経路表よりネットワーク近接性を考慮した経路表として優れているか，または同じ経路表である．

エントリ厳選操作の手順は次の通りである．ただし，sticky entry のうち successor list サイズを $c$ とする．

(1) 削除候補のエントリ集合 $C$ に経路表 $E$ の全エントリを追加する．
(2) $C$ から sticky entry を取り除く．
(3) ノード間の通信遅延の閾値となるエントリを閾値エントリ $e_{\mathrm{threshold}}$ とすると，

$$e_{\mathrm{threshold}} = \begin{cases} e_{c+1}, & e_{\mathrm{add}} \in \{e_i\}_{i=1,2,\ldots,c} \text{ のとき} \\ e_{|E|-1}, & e_{\mathrm{add}} = e_{|E|} \text{ のとき} \\ e_{\mathrm{add}}, & \text{otherwise} \end{cases} \tag{12}$$

(4) $e_{\mathrm{threshold}} \in E$ ならば $C = C - \{e \in E \mid e.\mathrm{latency} < \mathrm{e_{threshold}}.\mathrm{latency}\}$
(5) $C$ のエントリの中で $S^E_{i-1} + S^E_i$ が最小となるエントリ $e_i = e_{\mathrm{remove}}$ を経路表から削除する．

最新追加エントリ $e_{\mathrm{add}}$ が sticky entry の場合は $e_{\mathrm{add}}$ を削除することができないので，$e_{\mathrm{add}}$ を追加することで sticky entry からあふれたエントリを閾値エントリ $e_{\mathrm{threshold}}$ にする (手順 (3))．経路表サイズを変更したいとき，エントリ厳選操作を連続して実行する場合がある．その場合，$e_{\mathrm{threshold}} \notin E$ が起こりうるので，通信遅延に基づき $C$ からのエントリ削除を行わない（手順 (4)）．

手順 (1)，(2) および (5) は FRT-Chord のエントリ厳選操作である．PFRT-Chord により拡張した手順は手順 (3) および (4) である．$S^E_{i-1} + S^E_i$ を昇順に保持し，上から $e_{\mathrm{threshold}}$

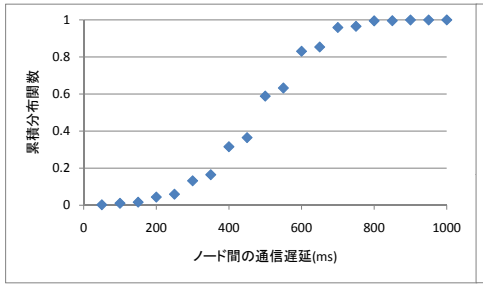


**図 1** ノード間の通信遅延の分布
Fig. 1 CDF of communication latency

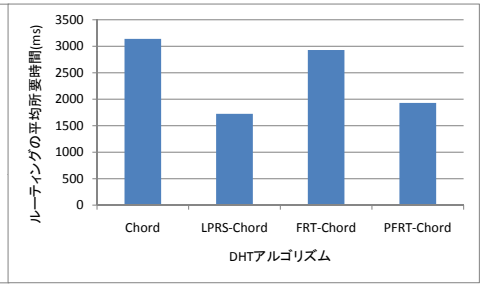


**図 2** 各 DHT アルゴリズムの平均所要時間
Fig. 2 Average route duration of Chord, LPRS-Chord, FRT-Chord and PFRT-Chord

の通信遅延以上であるかを調べることで，削除エントリ $e_{\text{remove}}$ の探索をすることができる．$e_{\text{remove}}$ の探索は $O(L)$ で終了する．

## 4. 評　　価

オーバレイ構築ツールキットである Overlay Weaver[7][8] 上に，PFRT-Chord を実装した．一台のマシン上でエミュレーションによる実験・評価を行った．ルーティングの所要時間は，全ホップの通信遅延の総和である．一対一でのノード間の通信遅延は 4.1 節で設定する．実験に使用したマシンは次の通りである．

- Overlay Weaver 0.10.1
- OS: Windows 7 Professional 32 bit
- CPU: Intel 2.67 GHz Core 2 Quad Q9400
- メモリ: 4 GB
- Java SE 6 Update 22

### 4.1 ノード間の通信遅延の設定

ノード間の通信遅延を設定するために，Transit-Stub モデル（TS モデル）[9] を利用した．TS モデルにより，実ネットワークのトポロジをシミュレートを行った．TS モデルはトランジットノードとスタブノードの 2 種類のノードで構成されている．トランジットノード間，トランジット・スタブノード間およびスタブノード間の通信遅延を 100 ms，20 ms，5 ms とし，一対一でのノード間の通信遅延を決めた．その結果，ノード間の通信遅延の分布は**図 1** となり，通信遅延の平均値は 470 ms，最大値は 1000 ms となった．

### 4.2 実 験 結 果

4.2.1 節で PFRT の有効性を示し，4.2.2 節で想定した通りに経路表の改良がされていることを示す．これらの実験に使用したパラメータは次の通りである．

- ノード数: $N = 10000$
- 探索回数: 1 ノードあたり 100 回
- FRT-Chord および PFRT-Chord の経路表サイズ: $L = 16$
- successor list サイズ: $c = 4$

#### 4.2.1 PFRT の有効性

**図 2** に，PFRT-Chord と他の DHT アルゴリズムのルーティングにおける平均所要時間を示す．他の DHT アルゴリズムと平均所要時間を比較することで，PFRT-Chord の有効性を示す．比較対象は Chord，LPRS-Chord および FRT-Chord である．

Chord および LPRS-Chord では finger table のサイズは変更できない．そこで，FRT-Chord および PFRT-Chord の経路表サイズを Chord，LPRS-Chord の経路表サイズに合わせた．finger table のサイズは $\log N$ と計算できるので，約 13 である．successor list および predecessor を加えると Chord および LPRS-Chord の経路表サイズは約 18 となる．PFRT-Chord が有利にならないようにするために，FRT-Chord および PFRT-Chord では経路表サイズ $L = 16$ とした．

PFRT-Chord でのルーティングにおける平均所要時間は 1930 ms であった．まず，Chord と比較する．Chord の平均所要時間は 3140 ms であることから，PFRT-Chord により約 40% 減少した．この結果より，PFRT-Chord はルーティングの平均所要時間を短縮したといえる．

次に FRT-Chord と比較する．FRT-Chord の平均所要時間は 2930 ms であることから，約 35% 減少した．この結果より，PFRT は FRT に基づく DHT アルゴリズムをネットワーク近接性を考慮するように拡張したといえる．

最後に LPRS-Chord と比較する．LPRS-Chord の平均所要時間は 1724 ms であることから，約 10% 増加した．しかし，平均所要時間の差は約 200 ms であり，これはノード間の通信遅延の平均値の半分以下に過ぎない．また，LPRS-Chord の経路表は finger table であり，経路表サイズは変更できないが，一方 PFRT-Chord は FRT の特長を引き継ぎ，経路表サイズの変更が可能なので，経路表サイズを大きくすることで平均所要時間を短縮することも容易である．さらに，PFRT-Chord は他にも FRT の特長を引き継いでいるため，PFRT-Chord は LPRS-Chord より優れた DHT アルゴリズムであるといえる．

Vol.2011-OS-118 No.15
2011/7/28

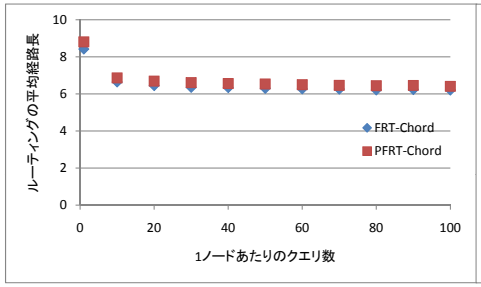


**図 3** 実行クエリ数と平均経路長の関係
Fig. 3 Correlation between number of lookups and average route length

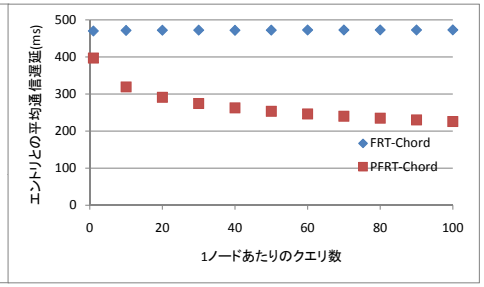


**図 4** 実行クエリ数と経路表エントリの平均通信遅延の関係
Fig. 4 Correlation between number of lookups and average communication latency to entry node

#### 4.2.2 PFRT-Chord の経路表改良の効果

PFRT-Chord では，経路表を改良するごとにルーティングの経路長が減少し，エントリの自ノードとの通信遅延が小さくなる．**図 3** および**図 4** は，クエリ実行回数に対する平均経路長および経路表上の各エントリと自ノードとの通信遅延の平均値（以下，平均通信遅延）を示している．1 ノードあたり 10 クエリ実行する毎にルーティングの結果を測定している．

まず，図 3 からルーティングの平均経路長の変化を確認する．PFRT-Chord では，実行したクエリ数が増加するにつれて，平均経路長が減少している．よって，エントリ厳選操作によって平均経路長が減少していることがわかる．また，10 番目のクエリを実行するまでの間に平均経路長が大きく減少し，それ以降のクエリでは緩やかに減少し続け，100 番目のクエリを実行する時にはほとんど減少していない．つまり，PFRT-Chord の経路表は収束していないが，少ないクエリ数で収束後の経路表に大きく近づくことができる．さらに，FRT-Chord とほぼ同じ変化であるので，PFRT-Chord に拡張することによる経路長の損失がほとんどないといえる．

次に，図 4 から平均通信遅延の変化を確認する．PFRT-Chord では，実行したクエリ数が増加するにつれて平均通信遅延が減少している．このことから，エントリ厳選操作によって平均通信遅延が減少していることがわかる．また，10 番目のクエリを実行するまでの間に平均通信遅延が大きく減少し，それ以降のクエリでは緩やかに減少している．つまり，PFRT-Chord の経路表は平均通信遅延が収束するために，平均経路長が収束するよりたくさんのクエリを実行する必要があるが，少ないクエリ数で収束後の経路表に大きく近づき，クエリ数を増やすことでより経路表をより改良することができる．さらに，FRT-Chord より常に平均通信遅延が小さく，100 番目のクエリを実行する時に約 50% 減少しているので，PFRT-Chord に拡張することで大きく減少させることができるといえる．

## 5. まとめ・今後の課題

ノード ID 以外を考慮しやすいルーティングアルゴリズム設計方式である柔軟な経路表（FRT）を拡張した，ネットワーク近接性を考慮した柔軟な経路表（PFRT）を提案した．PFRT は，ネットワーク近接性を考慮した経路表の順序関係 $\leq_{\mathrm{ID+PR}}$ に基づき経路表の改良を繰り返すことで経路表を構築する方式である．また，PFRT に基づき FRT-Chord を拡張した PFRT-Chord を提案し，実験・評価を行った．

PFRT-Chord はルーティングの平均所要時間が Chord より約 40% 減少し，FRT-Chord より約 35% 減少した．また，平均所要時間が LPRS-Chord よりわずかに増加したが，PFRT-Chord は FRT の他の特長（たとえば，経路表サイズの変更可能など）を持っている．

PFRT-Chord は，FRT-Chord を高々数ステップ拡張したにすぎない．その数ステップの拡張により，経路表エントリの自ノードとの通信遅延を大きく減少させることができた．また，FRT-Chord を PFRT-Chord に拡張することにより経路長やエントリの自ノードとの通信遅延の損失がほとんどないことがわかった．

PFRT-Chord では $\leq_{\mathrm{ID+PR}}$ の定義から，ID に基づく順序関係 $\leq_{\mathrm{ID}}$ およびノード間の通信遅延に基づく順序関係 $\leq_{\mathrm{PR}}$ の両方において優れている経路表に改良している．しかし，$\leq_{\mathrm{ID}}$ または $\leq_{\mathrm{PR}}$ のどちらか一方のみが優れている経路表に改良したときの方が，ルーティングの所要時間が減少する場合が考えられる．それらの改良が行われないことによる損失を分析することで，さらに良い順序関係 $\leq_{\mathrm{ID+PR}}$ を定義することができる．

本研究では，$\leq_{\mathrm{PR}}$ を用いることでノード間の通信遅延を考慮し，ネットワーク近接性を考慮した経路表の順序関係 $\leq_{\mathrm{ID+PR}}$ を定義したが，ID 以外の情報（たとえばネットワークに生存している時間）を考慮するために，$\leq_{\mathrm{EX}}$ を用いて本研究と同じ方式で経路表の順序関係 $\leq_{\mathrm{ID+EX}}$ を定義することが可能である．単独または複数の ID 以外の情報を考慮するように柔軟な経路表を拡張することが課題である．

**謝辞** 本研究は科研費（22680005）の助成を受けたものである．

Vol.2011-OS-118 No.15
2011/7/28

## 参 考 文 献

1) Stoica, I., Morris, R., Karger, D., Kaashoek, F. and Balakrishnan, H.: Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications, *Proc. ACM SIGCOMM '01*, pp.149–160 (2001).
2) Maymounkov, P. and Mazieres, D.: Kademlia: A Peer-to-peer Information Systems Based on the XOR Metric, *Proc. IPTPS 2002*, pp.53–65 (2002).
3) Rowstron, A. and Druschel, P.: Pastry: Scalable, Decentralized Object Location, and Routing for Large-scale Peer-to-peer Systems, *Proc. IFIP/ACM Middleware 2001*, pp.329–350 (2001).
4) Gummadi, K., Gummadi, R., Gribble, S., Ratnasamy, S., Shenker, S. and Stoica, I.: The Impact of DHT Routing Geometry on Resilience and Proximity, *Proc. ACM SIGCOMM '03*, pp.381–394 (2003).
5) Nagao, H. and Shudo, K.: Flexible Routing Tables: Designing Routing Algorithms for Overlays Based on a Total Order on a Routing Table Set, *Proc. IEEE P2P '11* (2011) (to be published).
6) Zhang, H., Goel, A. and Govindan, R.: Incrementally Improving Lookup Latency in Distributed Hash Table Systems, *Proc. ACM SIGMETRICS '03* (2003).
7) 首藤 一幸：Overlay Weaver, http://overlayweaver.sourceforge.net/.
8) Shudo, K., Tanaka, Y. and Sekiguchi, S.: Overlay Weaver: An Overlay Construction Toolkit, *Computer Communications*, Vol.31, No.2, pp.402–412 (2008).
9) Zegura, E.W., Calvert, K.L. and Bhattacarjee, S.: How to Model an Internetwork, *Proc. IEEE INFOCOM '96*, pp.592–602 (1996).