

# スーパーピア型P2Pネットワークにおける インデックス再構成手法

グエンクアンキー<sup>†</sup> 黄宏軒<sup>†</sup> 川越恭二<sup>†</sup>

<sup>†</sup>立命館大学情報理工学部

## 1 はじめに

P2P ネットワーク技術の一つに、ピア (Peer) が特定のスーパーピア (Super-peer) に接続されている Super-peer 型 P2P ネットワーク [1] がある。この技術はクラウドや分散データを管理する手法などに実際に利用されている。Super-peer 型 P2P ネットワークでは、各 Peer に格納されているコンテンツを検索する際、コンテンツの件数が増加するにつれ、データを検索する時間が増える。このため、ネットワーク内のデータを効率的に管理することが重要である。Nasrin ら [2] は Super-peer 型 P2P ネットワークに適用する  $SG^2$  と呼ぶインデックスを提案した。 $SG^2$  インデックスは多次元データの管理と範囲クエリに強く対応できる。しかし、 $SG^2$  インデックスはネットワーク内のデータ分布を考慮しないため、コンテンツの登録・更新・削除によりデータ管理に必要なインデックスサイズが増加するという問題がある。

そこで本研究では、 $SG^2$  インデックスに対応するデータ分布を考慮したインデックス再構成手法を提案する。コンテンツとそれを保持する Peer に関するインデックスを再構成することにより、データ検索で参照する Super-peer 数を減らす。その結果、ネットワーク内のデータへ効率的にアクセスが可能になる。

## 2 $SG^2$ インデックス

$SG^2$  インデックスの構成を図 1 に示す。 $SG^2$  インデックスは Global Index, Group Index, Synopsis Index から構築される。Global Index はネットワーク全体にどのようなデータが存在するかを表現する。理想的には、各 Super-peer がいつも最新の Global Index を保持していることである。Group Index は各 Super-peer に接続する Peer がどのようなデータを保持しているかを管理する。Synopsis Index は各 Super-peer 間にデータの更新情報を送るためのインデックスである。Global Index は Synopsis Index を基に更新する。

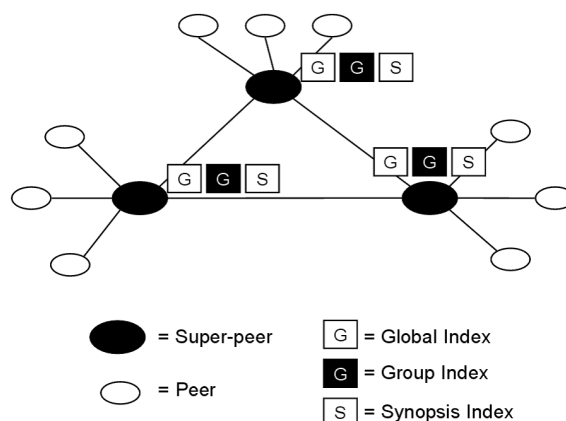


図 1:  $SG^2$  インデックスの構成

## 3 インデックス再構成手法

### 3.1 基本的考え方

本研究では、 $SG^2$  インデックスを再構成するためにネットワーク内に類似するデータを持つ Peer を特定の Super-peer に接続されるようにする。このことで、特定の Super-peer で類似するデータを管理できるようになる。その結果、利用者は少数の Super-peer にアクセスするだけでデータを検索することができる。同時にシステムの負荷が減らす。

### 3.2 再構成手法の手順

【Step1】再構成処理を適用する範囲内の複数の Peer が持つデータの Group Index を予め決められた特定の Super-peer に収集する。ここで本稿では、再構成処理を実行する Super-peer を「担当者」と呼ぶ。

【Step2】担当者は Peer が管理するデータの Group Index を基に、その Peer を特定するための「Peer インデックス」を作成する。Peer インデックスの作成方法は 3.3 節で説明する。担当者は Peer インデックスの集合にクラスタリング手法を適用する。その結果、同一クラスターには類似した Peer インデックスが割り当てられることになる。本研究では、予備実験の実験結果から、最も良い精度を得られた階層的クラスタリングの Single-link 手法を

**Index Reconstruction Method in a Super-peer P2P Network**  
Quang-Ky NGUYEN, Hung-Hsuan HUANG, Kyoji KAWAGOE  
College of Information Science and Engineering, Ritsumeikan University.  
{nguyen, kawagoe}@coms.ics.ritsumeikan.ac.jp  
huang@fc.ritsumeikan.ac.jp

適用する。

【Step3】各クラスタの Peer インデックスに対応する Peer を同じ Super-peer に接続されるように Peer の Super-peer を変更する。

【Step4】 Super-peer は保持している SG<sup>2</sup> インデックスを SG<sup>2</sup> インデックスの作成方法を基に再構成する。

### 3.3 Peer インデックスの作成方法

Peer は k 件のデータを持っているとする。その内、i 番目のデータに対応する Group Index は  $[x_{i1}, x_{i2}, \dots, x_{in}]$  である。ここで、n は Group Index の次元数である。  $x_{ij}$  は i 番目のデータに対応する Group Index の j 番目の成分であり、データの j 番目の属性をハッシュしたビット列である。Peer インデックスは Group Index と同一の構成を持ち、  $[y_1, y_2, \dots, y_n]$  である。Peer インデックスの成分は以下の数式で算出する。

$$y_j = \text{binary} \left( \frac{\sum_{i=1}^k \text{decimal}(x_{ij})}{k} \right)$$

このように平均を計算することによって Peer が持つデータの特徴を集約化することが可能になる。

### 3.4 集中型手法と分散型手法

再構成処理を実行する際に【Step1】の適用範囲により集中型手法と分散型手法の2つの手法を説明する。本稿で提案する手法は分散型手法である。

【集中型手法】担当者は一個だけでネットワーク内の全データの再構成処理を実行する。

【分散型手法】複数個の担当者を予め決めておき、個々の担当者は、各々の範囲内のデータの再構成処理を実行する。

## 4 評価実験

### 4.1 実験条件

本研究では、7680 件のデータ、768 個の Peer、96 個の Super-peer のネットワークの環境で本再構成手法の有効性を示す実験を行った。なお、分散型手法は再構成処理を適用する範囲個数が 4, 8, 12 個の3種類で評価した。ここで、集中型手法とこれらの分散型手法の有効性を以下の二点の指標で評価する。

【クエリアクセスコスト】一つの検索クエリに参照する Super-peer の平均個数である。

【再構成コスト】インデックス再構成するためにかかる Peer へのアクセス回数である。本稿では 3.2 節の再構成手法の手順の【Step1】と【Step3】においての Peer へのアクセス回数である。

### 4.2 実験結果

実験結果を図2と表1に示す。図2は各再構成手法の再構成コストを示す。表1は各再構成手法を用いた後と用いる前のクエリアクセスコストの割合を示す。以上の結果から以下のことが言える。

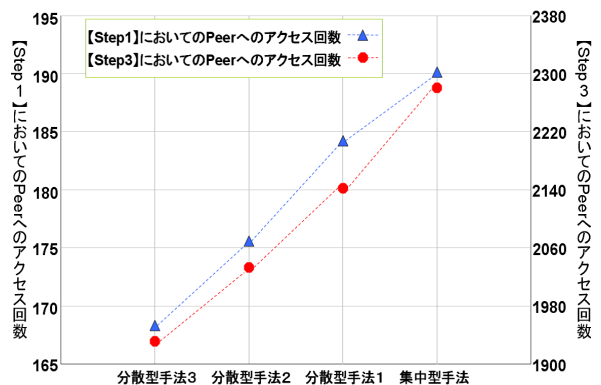


図 2: 再構成コスト

手法名	割合 (%)
集中型手法	64.915
分散型手法 1	89.750
分散型手法 2	93.613
分散型手法 3	95.696

表 1: クエリアクセスコストの割合

1. 分散型手法は集中型手法と比べてクエリアクセスコストは高いが再構成コストは低い。
2. 分散型手法は範囲個数が増加するとクエリアクセスコストは上がるが再構成コストは下がる。

## 5 おわりに

本稿では、Super-peer 型 P2P ネットワーク内のデータ分布を考慮し、インデックス再構成手法を提案した。分散型手法は再構成コストが低いという利点を持つが、アクセスクエリコストが高いという問題がある。今後は、さらに再構成コストを低く維持しながらクエリアクセスコストを低くする改良分散型手法を開発する予定である。現段階で以下の3つ方法を検討している。

1. Peer を Super-peer に再割り当てると共に部分的にもデータの再割り当てを行う。
2. 担当者間の通信を行うことでクエリアクセスコストを減らす。
3. 担当者間で Peer インデックスを共有する。

## 参考文献

- [1] B. Yang, H. Garcia-Molina, "Designing a Super-Peer Network", ICDE 2003.
- [2] N. Salma, K. Kawagoe, "SG<sup>2</sup>: A novel index structure for efficient data management in Super-peer architecture", ICUFN 2010.