2008－DPS－137 (2)
2008／11／27

位置情報に基づくオーバレイネットワークにおける推計統計学を用いた負荷平準化アルゴリズムの提案

小川 和真† 斉藤 裕樹‡

† 東京電機大学大学院 工学研究科 情報メディア学専攻
‡ 東京電機大学 工学部 情報メディア学科

概要

近年，無線デバイスの高性能化や普及により広範囲な位置情報サービスへの期待が高まっている．今後，センサ類から温度や照度といった状況情報を収集し，集約することでさまざまな状況依存サービスが開発されると考えられる．このような情報量が膨大になる位置情報を扱ったサービスの技術として P2P ネットワークの応用が期待されている．しかし，従来の位置情報を扱った P2P ネットワークでは，ピアやユーザからのクエリ，センサデータの偏りなどにより特定のノードに負荷が集中してしまう問題がある．本研究では，推計統計学を用いオーバレイネットワーク全体の負荷を予測することで効率的にオーバレイネットワーク全体の負荷を平準化するアルゴリズムの提案を行なう．

Load Balancing using Inferential Statistics in Location-based Overlay Networks

Kazumasa Ogawa† and Hiroki Saito‡

†Graduate School of Engineering, Tokyo Denki University
‡Department of Information Systems and Multimedia Design, Tokyo Denki University

Abstract

Recently, expectation for wide-ranging location information service has been developed with high performance and prevalence of wireless devices. A context-based services will be developed by collecting and consolidating environmental information like temperature and illuminance from sensors. These location services would be based on P2P network because of enormous quantity of data. However, existing P2P network has some problem which loads flock to a node because of peer, some queries from users and sensor data skew. We propose a load-balancing algorithm in location-based overlay network which calculate the load of entire overlay network by using inferential statistics.

1 はじめに

近年，携帯電話や PDA (Personal Digital Assistant) などの携帯端末やセンサデバイスの高性能化や普及により広範囲な位置情報を扱ったサービスへの期待が高まっている．今後，GPS (Global Positioning System) を搭載したセンサデバイスが広く普及し，周囲の環境から温度や照度といったコンテキスト情報を収集，集約することで位置に依存したさまざまなサービスが展開できると予想される．位置情報を扱ったサービスでは，膨大な数のセンサや携帯端末がネットワークに参加するため，ネットワーク内に発生する情報量は膨大な量となる．そのため，位置情報を扱うサービスを行なうネットワークとしては，情報を集中管理するサーバ・クライアント型より情報を分散管理する P2P ネットワークで構築することが望ましい．

近年，位置情報を利用した P2P ネットワークに関する研究が盛んに行なわれている．LL-Net [1] では，実世界空間を x, y 座標によってエリアに分割し，端末間で P2P ネットワークを構築し情報の管理を行なっている．Skip Geo Network [2] では，論理ネットワークと平面ネットワークを端末間で構築している．これらの研究は，センサデータやユーザからのクエリをシンクノードで管理や処理をせず，個々の端末で処理を行なっている．それに対し，P2P データポット [3] では，センサデータをシンクノードで集中管理し，シンクノード間で P2P ネットワークを構築している．しかし，従来の実世界空間を均等にエリアに分割し，各エリアをシンクノードで分散管理するオーバレイネットワークでは，ピアやユーザからのクエリ，センサデータの偏りなどにより特定のノードに負荷が集中してしまう．

そこで，本研究では，センサデータを複数のシンクノードで管理するオーバレイネットワークを想定し，オーバレイネットワークを構成する全ノードの負荷を推計統計学的に平準化するアルゴリズムの提案を行なう．具体的には，推計統計学の中心極限定理を用いてオーバレイネットワークを構成する全ノードの負荷と信頼区間を予測し，オーバレイネットワーク上

の 1 ノードの負荷とオーバレイネットワークを構成する全ノードの負荷との大小を比較するための負荷の指標を作成する．その負荷の指標を用いてオーバレイネットワークを構成する全ノードの負荷を平準化する．

2 位置情報に基づくオーバレイネットワークの負荷分散の課題

位置情報を扱ったサービスをオーバレイネットワーク上で実現するための要件の 1 つに範囲検索がある．範囲検索を行うためには，キーの順序性を保持しなければならない．キーの順序性を保持するオーバレイネットワークアーキテクチャとして Skip Graph [4] があるが，Skip Graph は 1 次元情報しか扱うことができないため，多次元情報を 1 次元情報に写像するアーキテクチャが必要となる．多次元情報を 1 次元情報に写像するアーキテクチャとして，空間充填曲線 [5] がある．

本研究では，位置情報に基づくオーバレイネットワークとして，空間充填曲線を用いて位置情報を 1 次元に写像し，センサデータを収集する複数のシンクノードで位置情報を分散管理するオーバレイネットワークを想定する．図 1 に本研究で想定するオーバレイネットワークの例を示す．2 次元空間を格子状に均等に分割された空間をエリアとし，この各エリアを空間充填曲線の 1 つである Z-ordering を用いて 1 次元に写像し，オーバレイネットワーク上のシンクノードでエリアを分散管理している．図 1 中のシンクノードの中に書かれている数字はシンクノードの識別子であり，四角形の中に記されている 2 進表記の値はエリアの ID である．各シンクノードは 4 つのエリアを担当し，エリア内のデータやクエリを管理している．

想定するオーバレイネットワークでは，ピアやユーザからのクエリ，センサデータの偏りなどにより，特定のノードに負荷が集中してしまう可能性がある．例えば，平日の昼間は都心に人が集中し，ユーザからのクエリやピアは都心に集中することが予想できるが，夜間には，人の帰宅により都心の外側に人が移動し，都心の周囲に人が集中するため，都心よりも都心の周囲にユーザからのクエリやピアが集中することが予想できる．また，長期休暇では，レジャー施設，海や山といった場所にピアやユーザからのクエリが偏ることが考えられる．このように，人の流れの変化によりユーザからのクエリやピアの数は変化するため，ノードの負荷を動的に平準化する必要がある．オーバレイネットワークを構成する全ノードの負荷を動的に平準化するためには，オーバレイネットワークを構成する全てのノードの負荷を知る必要があるが，オーバレイネットワークでは何千，何万といった膨大なノードがネットワークに接続することを想定するため，全てのノードの負荷を知るために全てのノードにクエリを送ることは非現実的である．

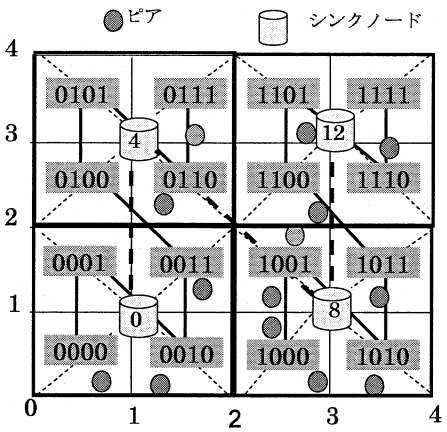

図 1 想定オーバレイネットワーク

そこで，本研究ではオーバレイネットワーク上の各ノードで推計統計学の中心極限定理を用いてオーバレイネットワーク全体の負荷の平均と信頼区間を予測し，負荷の指標を作成するアルゴリズムを提案する．また，得られた負荷の指標を用いてキーの順序性を保持した負荷平準化アルゴリズムの検討を行なう．

3 推計統計学を用いた負荷指標作成アルゴリズム

3.1 基本設計

想定するオーバレイネットワークのシンクノードは管理するキーに範囲があり，負荷平準化により管理するキーの範囲を動的に変更し，そのキーに対応するデータを移動することで負荷を分散する．負荷を平準化するためにオーバレイネットワーク上の全ノードにクエリを送り，全てのノードの負荷情報を得て高負荷ノードと低負荷ノードの負荷分散組み合わせを作るのは非現実的であるので，推計統計学を用いてオーバレイネットワーク上の全ノードの負荷の平均が含まれる範囲を推定し，オーバレイネットワーク全体の負荷を平準化する．

まずはじめに，高負荷ノードと低負荷ノードに分類するために負荷の指標を作成する．次に，各ノードは中心極限定理を用いて標本平均と信頼区間を算出し，それらを負荷の指標とする．なお本研究では負荷 L の定義を，蓄積データ量 D と送られてきたクエリ数 Q の積 ($L = DQ$) とする．

3.2 負荷指標作成アルゴリズム

負荷の大小を比較するための指標を算出するアルゴリズムを図 2 に示す．

はじめに，各ノードはランダムサンプリングを2回行う．1回目のランダムサンプリングでノードの識別子と負荷の情報を取得し，標本平均を計算する．このとき，中心極限定理より，オーバレイネットワークを構成する全ノードの負荷がどのような分布であってもオーバレイネットワークを構成するノード数が十分に多ければオーバレイネットワークを構成する全ノードの負荷の標本平均は近似的に正規分布に従うことが知られている．このため，標準誤差を用いた区間推定を行うことができる．

2回目のランダムサンプリングでノードの識別子と負荷と1回目のランダムサンプリングで算出した標本平均の情報を取得し，標本平均のさらに平均を計算する．以降，本論文では標本平均の平均を中心極限平均と呼ぶ．このとき，各ノードはランダムサンプリングで取得したノードの識別子と負荷の情報を保持する．次に，ノードの負荷を l，中心極限平均を v，2回目のランダムサンプリング数を n とし式 (1) を用いて不偏分散 u を計算する．

$$u = \frac{\sum_{i=1}^{n}(l_i - v)^2}{n-1} \tag{1}$$

得られた不偏分散 u から式 (2) を用いて標準誤差 s を計算する．

$$s = \sqrt{\frac{u}{n}} \tag{2}$$

信頼区間の最大値を最大信頼区間 p，信頼区間の最小値を最小信頼区間 q とし，得られた標準誤差 s と中心極限平均 v から式 (3) を用いて信頼区間の計算を行う．なお，式 (3) の t は t 分布表に記載されている値を用いる．

$$\begin{cases} p = v + t \cdot s \\ q = v - t \cdot s \end{cases} \tag{3}$$

最終的に得られた最大信頼区間と最小信頼区間と中心極限平均を負荷の指標として用いる．

以下に例を記述する．あるノードがランダムサンプリングして得た標本平均が {19, 109, 127, 52, 43, 39, 67, 111, 22, 51} とする．標本平均の合計が 640 でサンプリングの個数 10 であるから，中心極限平均 v は 64 となる．不偏分散 u は式 (1) より 1324.78 となり，式 (2) より標準誤差 s は 12.13 となる．信頼区間は 95% 信頼区間を用いると t=2.262 であるから式 (3) より $p = 91.44$，$q = 36.57$ となる．

4 負荷平準化アルゴリズム

実世界では人口やピアの数によりデータ数やクエリ数に偏りがあり，エリアを管理するシンクノードの負荷に偏りがあると予想できる．負荷の偏りには局所性があると予想できるので，シンクノードの隣接間でクラスタリングを行い，局所的に処理をすることで負荷を平準化のために送るメッセージ数や負荷平準化の効果がより高くなると予想できる．

負荷平準化アルゴリズムの概略は，次のようである．低負荷ノードがネットワークから離脱し，高負荷ノードの隣接ノードとしてネットワークに再参加し，データとキーを受け取ることで負荷を分散する．次節より負荷平準化アルゴリズムの詳細な手順を説明する．

Algorithm1. 負荷指標作成アルゴリズム

```
load barometer begin
while (n < k) do
   randomsapling := samplingnode(i)
   loadsum += samplingnode(i).load
   n++
   i++
done
   samplingaverage = loadsum / k
while(n < samplesize) do
   randomsampling− > samplingnode
   n++
done
/*標本平均の計算*/
      limitedaverage = samplingnode.sampleaverage/samplesize
/*不偏分散の計算*/
while(n < samplesize) do
     sum += (samplingnode.load − limitedaverage)^2
done
   unbiasedvariance = sum/(m -1)
/*標準誤差の計算*/
    noromalerror = (unbiasedvariance /samplesize)^1/2
/*信頼区間の計算*/
   reliablemax = limitedaverage + t × normalerror
   reliablemax = limitedaverage - t × normalerror
end
```

図2 負荷指標作成アルゴリズム

4.1 負荷の分類

3.2 節で得られた中心極限平均と信頼区間から負荷を以下の5つに分類し，クラスタリングを行うことで負荷平準化の効率化と負荷平準化に必要なノードの取得の効率化を図る．

i. 最小信頼区間の半分以上かつ最小信頼区間未満

ⅱ. 最小信頼区間の半分未満
ⅲ. 信頼区間内
ⅳ. 最大信頼区間の 2 倍以上
ⅴ. 最大信頼区間より大きいかつ最大信頼区間の 2 倍未満

4.2 ノードのクラスタリング

4.1 節の i，ⅱ，ⅴ に該当するノードは一定間隔ごとにクラスタリングを行なう．i，ⅱ に該当するノードのクラスタは負荷分散に必要なノードの取得を効率的にするために行う．ⅴ に該当するノードは，単一で低負荷ノードと負荷分散を行うと負荷が信頼区間外になってしまう可能性がある．例えば，信頼区間が 600 ～800 で負荷が 1000 のノードが，再参加したノードと負荷を 2 分すると，少なくともどちらかのノードの負荷は信頼区間外となる．そのため，ⅴ に該当するノードでクラスタリングを行い，複数のノードの負荷を処理することで負荷が信頼区間内に収まるようにする．以下に，i，ⅱ，ⅴ に該当するノードのクラスタリング条件を示す．

i に該当するノードは，隣接ノードが i に該当するノードで，ノード数 g でクラスタ内の負荷合計を c とし，式 (4) を満たすクラスタを作成する．クラスタ内の最大ノード数を k としたとき，最大ノード数 k でクラスタを作成できない場合，クラスタは作成しない．

$$q < \frac{c}{(g-1)} < p \tag{4}$$

ⅱ に該当するノードは，隣接ノードがⅱに該当するノードのとき，式 (5) を満たすようにクラスタリングを行なう．

$$q < c < p \tag{5}$$

ⅴ に該当するノードは，隣接ノードがⅴに該当するノードのとき，最大ノード数 k 個でクラスタリングを行なう．クラスタ内の各ノードはクラスタ内の全てのノード ID や負荷の情報を保持する．

4.3 負荷平準化に必要なノードの選択

4.1 節のⅳ，ⅴ に該当するノードは負荷が信頼区間内に収まるように負荷を平準化する．そのために，負荷平準化に必要なノード数の計算を行なう．それぞれⅳに該当するノードは式 (6)，ⅴ に該当するクラスタで最も負荷が低いノードは式 (7) を用いて負荷分散に必要なノード数 m を計算する．

$$m = \left(\frac{l}{v} - 1\right) \tag{6}$$

$$m = \frac{\sum_{i=1}^{k}(l_i - p)}{v} \tag{7}$$

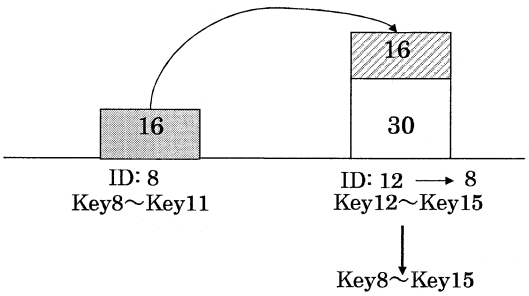

図 3 離脱時のデータの移動

次に，ランダムサンプリングした結果から 4.1 節の i，ⅱ に該当するノードを選択し，そのクラスタから m 個のノードをノード ID の昇順で選択する．これを m 個のノードが選択できるまで繰り返す．

4.4 ノードの離脱

負荷分散に必要なノードとして選択されたノードはネットワークからの離脱を行なう．離脱の順番はクラスタ内でノード ID の昇順で行う．離脱は，右側のリンクのノードにキーとそのキーに対応するデータを移動する．

図 3 に離脱時の負荷の移動の例を示す．ノード ID8 のノードがネットワークからの離脱を行なう．ノード ID8 のノードは右側のリンクのノード ID12 のノードに管理しているキーとそのキーに送られてきたデータを移動する．送られてきたデータとキーからノード ID12 のノードの負荷は 46 となり，管理するキーの最小値は 8 になり，ノード ID は管理するキーの最小値の変更に伴い 8 になる．

4.5 ノードの再参加による負荷の分散

離脱したノードは高負荷ノードの隣接ノードとしてネットワークに再参加する．

ⅳ に該当するノードの負荷分散は，キーが 1 つの場合とキーが 1 つ以上の場合に分けて負荷分散を行なう．キーが 1 つの場合，再参加するノードはⅳに該当するノードの複製ノードとなる．そのため，再参加するノードのノード ID とキーはⅳに該当するノードと同じになる．再参加するノードの負荷は，高負荷ノードの負荷を必要ノード数で除算した商となるように再参加するノードにデータを移動する．キーが 1 つ以上の場合，再参加するノードの負荷が最大信頼区間未満かつ最小信頼区間以上になるようにキーとデータを移動する．再参加ノードは高負荷ノードから送られてきたキーの最小値を ID とする．

ⅴ に該当するノードは式 (8) を用いて負荷分散に必要なノード数参加後のクラスタ内の負荷平均 f を計算し，クラスタ内の各ノードの負荷が f になるよう

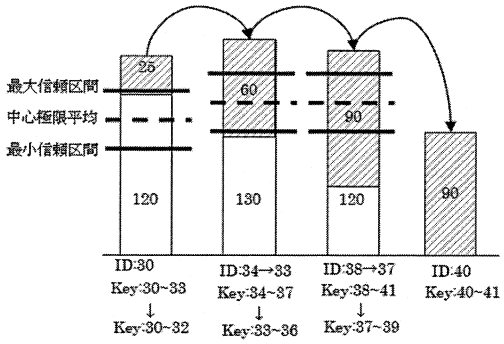

図 4　v に該当するノードの負荷分散

に負荷を分散する．

$$f = \frac{\sum_{i=1}^{k} l_k}{(k+m)} \tag{8}$$

図 4 に v に該当するノードの負荷分散の例を示す．図 4 は縦軸を負荷値とし，クラスタ内の負荷の合計が 370，負荷分散に必要なノード数を 1 としている．ノードの参加後はクラスタ内のノード数が 4 となり，クラスタ内の負荷平均が 92.5 となる．ノード ID38 のノードは再参加ノードの負荷がクラスタ内の負荷平均に近づくようにキーとデータを移動する．ノード ID34 のノードも同じように負荷が 92.5 に近づくようにデータとキーをノード ID38 のノードに移動し，ノード ID30 のノードも同様に ID34 のノードにデータとキーを移動する．データとキーを移動後，送られてきたキーから管理するキーの最小値と最大値を設定し，キーの最小値を ID として設定する．

5　シミュレーション実験

提案手法の有用性を示すために負荷平準化の精度の評価とクラスタの有無による平準化コストの比較の評価，サンプリングノード数の変化による信頼区間算出の精度の評価の 3 つについてシミュレーション実験を行った．

5.1　評価パラメータ

シミュレーションに用いたデータセットは，クエリとデータの発生確率が領域の中央を中心とした指数分布とし，以下のシミュレーション条件で行った．

領域の大きさ $2^{10} \times 2^{10}$，ノード数:1024, 全体のクエリ数:100000，全体のデータ数:100000，オーバレイネットワーク全体の負荷平均:719.45，1 回目のランダムサンプリング数:100 回，2 回目のランダムサンプリング数:50 回，信頼区間は 95% 信頼区間とした．

5.2　負荷平準化の精度の評価

本シミュレーション実験は，負荷の平準化の効果を評価するために行った．

図 5 に各ノードの ID ごとの平準化前の負荷の値と平準化後の負荷の値を示す．平準化前は，負荷が 200 以下のノードが 65% 以上で負荷が 1600 以上のノードが 20% 以上と，負荷の分布に偏りが見られる．平準化後は，約 65% のノードの負荷が 600〜800 に収り，負荷が平準化されていることが確認できた．少数のノードの負荷が平準化されていないが，これは低負荷ノードがサンプリングにもれたことが原因と考えられる．

全体の負荷の平均が 759.45 で各ノードの最小信頼区間の平均値が 660.55, 最大信頼区間の平均値が 844.33 であることから全てのノードの負荷が 660〜840 の間の値になるのが望ましい．シミュレーション結果では，約 65% のノードがその区間内に収まった．

5.3　クラスタの有無による比較評価

本シミュレーション実験は，クラスタの有用性を示すたにクラスタリングを行う場合と行わない場合についての負荷平準化の精度と負荷平準化コストの比較を行う．本シミュレーションでは，オーバレイネットワーク上の全ノードが負荷平準化のためにノードに送ったメッセージ数の総和を総メッセージ数としている．また，総メッセージ数を平準化コストとしている．クラスタリングを行い局所的に処理を行うことで，負荷平準化のために送るメッセージ数や負荷平準化の効果をより高くできると予想できる．

図 6 にクラスタの有無によるノードの負荷の頻度分布を示す．負荷平準化前と比較するとクラスタの有無に関わらずノードの負荷が 600〜800 に多くのノードが集中し負荷が平準化されていることがわかる．クラスタを行った場合では，約 65% のノードの負荷が 660〜840 の間に収まり，クラスタを行わない場合では，約 52% のノードの負荷が 640〜840 の間に収まった．結果として，クラスタリングを行った場合，クラスタリングを行わなかった場合と比べて負荷平準化の精度は約 10% 向上した．

図 7 にクラスタの有無による平準化コストを示す．クラスタリングを行った場合とクラスタリングを行わなかった場合を比較するとクラスタリングを行うことで総メッセージ数は約 30% 削減されることがわかった．また，負荷を分散に必要なノードを選択するために送るメッセージ数は全体で約 60% 削減された．以上のことから，クラスタリングは平準化コストにも有用であることがわかる．

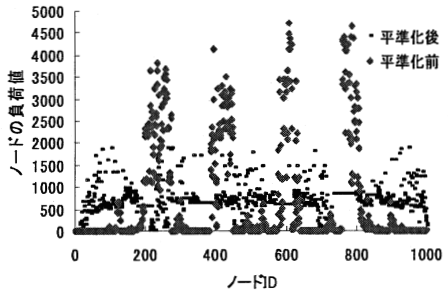

図 5　ノード ID ごとの負荷の分布

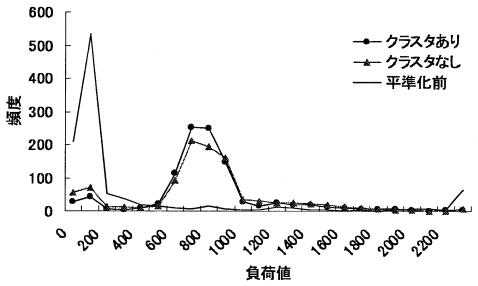

図 6　ノードの負荷の頻度分布

5.4　サンプリングノード数による信頼区間の精度の評価

表 1 はサンプリングノード数と信頼区間の精度の比較である．信頼区間は，オーバレイネットワーク上の全ノードの平均を取った値である．サンプリング数が増加すると信頼区間の範囲は狭まる．サンプリング数を増やすことで信頼区間の範囲が狭まるのは，標準誤差が小さくなるためだと考えられる．信頼区間の範囲が狭まることにより平準化後の各ノードの負荷がよりオーバレイネットワーク全体の負荷により近似する．5.2 節の実験と同条件でサンプリングノード数を 800 とし，平準化を行うと約 78% のノードの負荷が 700〜800 の間に収まった．以上のことからサンプリングノード数を増やすことで信頼区間の精度と平準化の精度を向上させることができることがわかった．

6　まとめ

本論文では，センサデータを収集するシンクノードで構成された位置情報に基づくオーバレイネットワークのための負荷分散アルゴリズムとして推計統計学を用いた負荷平準化手法を提案した．また，シミュレーション実験を行うことで提案手法の有用性を示した．今後は，信頼区間の整合性をとることで負荷分散をより効率的に行うことを検討する．

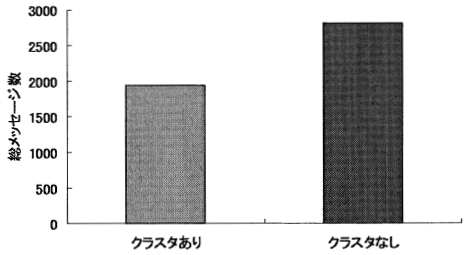

図 7　クラスタリングによるコストの比較

表 1　サンプリングノード数と信頼区間の関係

サンプリングノード数	最小信頼区間	最大信頼区間
50	619.56	812.04
250	670.54	753.81
500	696.13	759.62
750	694.67	742.61
1000	702.57	744.39

謝辞

本研究は，東京電機大学総合研究所研究課題 Q07J-02 として行ったものである．

参考文献

[1] 金子 雄, 福村 真哉, 春本 要, 下條 真司, 西尾 章次郎, "モバイル環境における端末の位置情報に基づく P2P ネットワークの提案と評価, " 電子情報通信学会第 15 回データ工学ワークショップ論文集, 2004.

[2] 奥 智照, 坪井 新治, 大西 真晶, 上島 紳一, "P2P 型ジオキャストのための階層ネットワークの提案と評価, " 電子情報通信学会第 19 回データ工学ワークショップ論文集, 2008.

[3] 藤崎 友樹, 鈴来 和久, 横田 裕介, 大久保 英嗣, "P2P データポット:センサネットワーク向け分散マイクロストレージアーキテクチャ, " 電子情報通信学会第 18 回データ工学ワークショップ論文集, 2007.

[4] James Aspnes, Gauri Shah: "Skip Graphs, " ACM SIAM Symposium on Discrete Algorithms, pp. 384-393 Jan. 2003.

[5] Bongki Moon, H. V. Jagadish, Christos Faloutsos, and Joel H. Saltz, "Analysis of the Clustering Properties of the Hilbert Space-Filing Curve, " IEEE Transaction on Knowledge and Data Engineering Vol. 13, No.1, pp. 124-141, 2001.