社団法人電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
CPSY2009-3, DC2009-3(2009-4)

非構造化オーバーレイネットワーク構築におけるメトロポリス法を用いたアルゴリズムの評価

高村 達史† 土屋 達弘† 菊野 亨†

† 大阪大学大学院情報科学研究科
〒 565–0871 大阪府吹田市山田丘 1-5
E-mail: †{t-takamr,t-tutiya,kikuno}@ist.osaka-u.ac.jp

あらまし P2P システムの通信はオーバーレイネットワークと呼ばれる仮想的なネットワーク上で行われる．オーバーレイネットワーク上で通信コストと故障耐性を最適化する手法の一つに，モンテカルロシミュレーションで用いられるメトロポリス法を適用するアルゴリズムがある．これらのアルゴリズムは，乱数に基づき局所的な接続関係を変化させることを繰り返すことによってネットワークを最適化する．メトロポリス法の最大の利点は局所解に収束することを防ぐことにあるが，そうすることが実際に有効であるかは定量的に議論されていなかった．本研究ではメトロポリス法である Localiser を対象に，メトロポリス法を用いることがオーバーレイネットワークのパフォーマンスと故障耐性の実現に有効であるかどうかについて検討する．

キーワード 非構造化オーバーレイネットワーク，peer-to-peer，耐故障性，メトロポリス法

Evaluation of a Metropolis Algorithm for Constructing Unstructured Overlay Networks

Tatsushi TAKAMURA†, Tatsuhiro TSUCHIYA†, and Toru KIKUNO†

† Graduate School of Information Science and Technology, Osaka University
Yamadaoka 1-5, Suita-shi, Osaka, 565-0871 japan
E-mail: †{t-takamr,t-tutiya,kikuno}@ist.osaka-u.ac.jp

Abstract Peer-to-peer(P2P) systems use a virtual network called an overlay network to route messages to distinations. Some algorithms adopt the Metropolis scheme, which is a common Monte Carlo method, to optimize the communication cost and fault tolerance of an overlay network. These algorithms iteratively perform local topological changes in a randomized fashion, eventually resulting in an optimized network. The intended advantage of using the Metropolis scheme is the avoidance of getting trapped in local optima; however there has been no convincing evidence for it. In this paper we consider Localiser, which is one of these Metropolis scheme-based algorithms, and study the effects of using the Metropolis scheme on the performance and resiliency of an overlay.

Key words Unstructured overlay networks, peer-to-peer, fault tolerance, metropolis algorithm

1. まえがき

インターネットの発展に伴い，通信効率と耐故障性 (信頼性) の高い P2P 技術への要求が高くなっている．P2P システムの通信はオーバーレイネットワークと呼ばれる仮想的なネットワーク上で行われる．オーバーレイネットワークはすでに存在する下位ネットワークの上位にあたる層において仮想的，論理的ネットワークを構築し，情報探索などさまざまな機能の実現を可能にする．オーバーレイネットワークは主に構造化オーバーレイネットワークと非構造化オーバーレイネットワークに分けることができる．構造化オーバーレイネットワークはどのノードの隣にどのノードがくるか，どのコンテンツをどのノードに格納するか，などのトポロジ構成に数学的な制約が存在する．そのため，分散ハッシュテーブル (DHT) [1] [2] [3] [4] を用いて高速で正確な探索が可能であるが，ノードの参加，離脱が頻繁に行われる環境では情報を頻繁に更新する必要があり，耐故障性が低くなる．一方，非構造化オーバーレイネットワークのトポロジ構成は数学的制約に従わずに構成される．非構造型での探索は主にフラッディングやゴシップ，ランダムウォークなどランダム探索手法であり，探索成功率はやや低いが柔軟で耐故

障性が高い．しかし，一回の探索に多くのメッセージを送信し，ネットワークに負荷をかける．探索の通信コストを下げるためには隣接ノード同士は地理的に近いもの同士，つまりオーバーレイネットワークは下位ネットワークトポロジを反映させることが望ましい．

物理的な近接性を反映した非構造化ネットワークを構築する手法として，モンテカルロシミュレーションで用いられるメトロポリス法を適用し，乱数に基づいてノード間の接続を繰り返し変更する方法が知られている [5], [6].

メトロポリス法は，暫定解を乱数に基づき変化させていく手法であり，現在の状態より良い解が次の解の候補となった場合はその変化は必ず選択される．一方，良くない解が候補の場合でも，局所解におちいることを防ぐために，温度パラメータに基づいて確率的に遷移を認める．この考えに基づき，文献 [5], [6] では，局所的な接続関係の変化によって，ネットワーク全体に対する目的関数の値が改善されるなら接続の変更を実施し，改善されない場合でも確率的に変更を行うというアルゴリズムを提案している．しかしながら，局所解への収束の回避を意図したこのような考えが，非構造化ネットワークの構築において，実際に有効であるかは定量的に議論されていなかった．

そこで本研究では，メトロポリス法を利用したアルゴリズムである Localiser [5] を対象とし，温度パラメータを変化させてシミュレーションを行うことで，一時的な改悪となる遷移を認めることが，実際に有効かどうかを評価する．

以降，2 節では非構造化オーバーレイネットワークに求められる特性について述べる．3 節では Localiser アルゴリズムについて述べる．4 節ではシミュレーション実験で考えるオーバーレイネットワークの下位の物理ネットワークと初期オーバーレイネットワークについて述べ，5 節ではシミュレーションの結果を示す．最後に 6 節で本研究のまとめと今後の課題について述べる．

2. 非構造化オーバーレイネットワーク

本論文では，オーバーレイネットワークのトポロジを無向グラフ $G(V,E)$ として考える．ここで V は頂点集合，E は辺集合である．頂点はノードに対応し，辺はリンクに対応する．任意の 2 ノード i,j 間に，通信コスト $c(i,j)=c(j,i)>0$ が定められていることを仮定することで，ノード間の物理的な近接性をモデル化する．$c(i,j)$ は定数であり，リンク集合 G に依存しないものとする．

オーバーレイネットワークに要求される特性として，通信コストと故障耐性を考える．これらの特性は，それぞれ，近接性の反映と次数分布の平均化によって実現することができる

- 近接性の反映

非構造化オーバーレイネットワークでの検索はフラッディングの様に隣接ノードに問い合わせのメッセージを送り，見つからなければさらに隣接ノードへ問い合わせる．帯域を圧迫せず，遅延を減らすために，下位トポロジを反映させて隣接ノード同士は地理的に近いもの同士が接続していることが望ましい．

- 次数分布

ノードの次数はオーバーレイネットワーク上の到達可能性の指標となるだけでなく耐故障性の指標となる．障害が起こる確率がすべてのノードで一定とすると，全てのノードの次数は同じであることが望ましい．次数が小さいノードは障害が増えるにつれて孤立しやすく，次数が大きいノードに障害が起こるとネットワークへの影響が大きい．

3. Localiser

Localiser は Massoulié 等によって提案されたアルゴリズムであり，メトロポリス法に基づき通信コストと耐故障性を最適化する [5].

メトロポリス法では関数 f の値が最小になるように暫定解を繰り返し更新する．f の定義域を D とする．現在の解 $x\in D$ から次の候補解 $y\in D$ へ更新するときを考える．この移動は $f(y)$ が $f(x)$ より大きいときでもある程度の確率で行われる．この理由は局所的な最小値におちいることを避けるためである．もし更新が認められなければ解は x のままである．

Localiser では目的関数を次の式で与える．

$$f(G)=w\sum_{i\in V(G)}d_i^2+\sum_{(i,j)\in E(G)}c(i,j)$$

w は重みパラメータ，d_i はノード i の次数，$c(i,j)$ はノード i とノード j の間の通信コストを表している．

Localiser では，各ノードが独立に，図 1 のようなリンクの変更を繰り返し実行する．変更に先立ち，ノードはリンクの変更による目的関数 f の値の変化量 Δf を計算する．Δf は，ノードとその隣接ノードの情報からのみで算出でき，ネットワーク全体の情報を必要としない．Δf が 0 より小さければ，リンクの変更を実行し，0 より大きい場合は，確率的に変更の実行を決定する．以下，各ノードが実行するステップを示す．

（1） ノード i の隣接ノード j と k をランダムに選び，通信コスト $c(i,j)$ と $c(i,k)$ を求める．

（2） ノード j と k にメッセージを送り，各々の次数 $d(j)$ と $d(k)$ を受け取る．さらに通信コスト $c(j,k)$ の推定値を受け取る．

（3） リンク (i,j) をリンク (j,k) に接続しなおした場合のコストの変化量を計算する．変化量 Δf は次の式で与えられる．
$\Delta f=2w(d_k-d_i+1)+c(j,k)-c(i,j)$

（4） リンク (i,j) からリンク (j,k) リンクへの付け替えを確率 $p=\min\left(\left(e^{-\Delta f/T}\frac{d_i(d_i-1)}{d_k(d_k+1)}\right),1\right)$ で行う．

このアルゴリズムは $f(G)$ が増加する場合でもある程度の確率で解の更新を行い，極小値におちいることを避ける．また，リンクが付け替えられるだけなので，ネットワーク全体の辺の数は不変である．

このアルゴリズムは w と T のパラメータを持つ．

（1） パラメータ w を大きくすると次数を均一化することに重点を置くようなる．$w=0$ の場合，通信コストのみを考慮に入れてネットワークを最適化する．

（2） パラメータ T(遷移する確率を計算するために用いる)

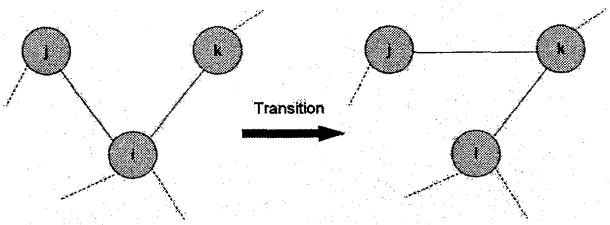

図 1　Localiser アルゴリズム

は温度と呼ばれ，温度を小さくすると極小値から別の極小値へ移動することが難しくなり，大きくすると極小値に収束しない．

4. シミュレーションモデル

4.1 物理ネットワーク

下位物理ネットワークのトポロジは Waxman モデル [7] と Transit-Stub モデル [8] を用いてモデル化する．ノードは 100 台とし，そこにエンドノードがランダムに接続する．隣接する 2 つのノード間の通信コストを 50ms とし，オーバーレイネットワーク上の 2 エンドノード間の通信コストを物理ネットワークでそれらが接続している 2 ノード間の最短パスにおける通信コストの合計値とする．

4.1.1 Waxman モデル

Waxman モデルは実際のネットワークの性質をランダムグラフに加えたものである．このモデルでは距離が近いノード同士が接続しやすい傾向がある．ノード u がノード v に接続する確率を $P(u,v)$ とする．$0 < \alpha,\beta \leq 1$ とし，ノード u とノード v の距離を d，最も離れた二つのノードの距離を L としたとき，確率 $P(u,v)$ は次の式で与えられる．

$$P(u,v) = \alpha e^{-d/(\beta L)}$$

今回の実験では $\alpha = 0.2$，$\beta = 0.15$ とし，物理ネットワークの平均次数を 3.5 にした．

4.1.2 Transit-Stub モデル

Transit-Stub モデルは，インターネットにおける特徴的な性質を有するネットワークを構築する．ネットワークのいくつかの部分をランダムグラフで構築し，それらを組み合わせることによって全体のネットワークを構築する．インターネットは，共通の管理や経路情報を持つノードのグループ (ドメイン) から構成されていると見なすことができる．このモデルでは，ドメインはトランジットドメインとスタブドメインに分けられ，トランジットドメインに含まれるノードをトランジットノード，スタブドメインに含まれるノードをスタブノードと呼ぶ．スタブドメイン中のノードは同じドメイン中のノードと連結している．トランジットドメインはスタブドメイン同士を相互に連結させる．

今回の実験では平均次数を 3.5 とした．4 つのトランジットノードを持つ 1 つのトランジットドメイン，トランジットノード 1 つ毎に 3 つのスタブドメインが接続されており，スタブドメインの中には 8 つのスタブノードがある．

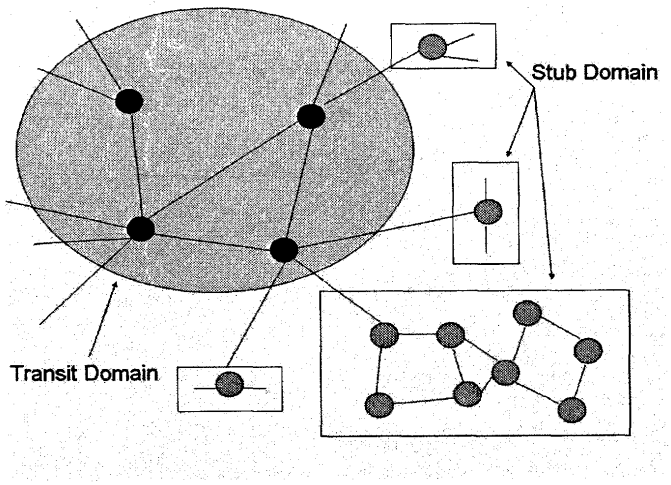

図 2　トランジットスタブモデルの例

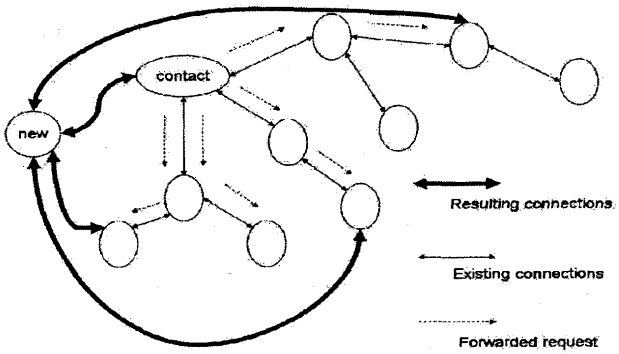

図 3　Scamp アルゴリズム

4.2 初期オーバーレイネットワーク

4.2.1 Scamp

Scamp [9] は非構造化オーバーレイネットワークを構築するプロトコルである．このプロトコルではそれぞれのノードが保持する情報は隣接ノードのアドレスだけであり，ネットワーク全体の情報を必要としない．

Scamp プロトコルは次の通りである．ただし，論文 [9] では単方向のリンクであったが，本論文では双方向のリンクに変更している．

（1） コンタクト

システムに新しく参加するノード (new) はシステムの任意のメンバー (contact) にサブスクリプションリクエストを送り，そのノードと接続する．

（2） 参加

サブスクリプションリクエストを受け取ったノードは新しく参加するノードの ID を含んだサブスクリプションリクエストを全ての隣接ノードへ送る．さらに隣接ノードからランダムに $a-1$ 個を選び，サブスクリプションリクエストを送る．(a はデザインパラメータであり，耐故障性に影響を与える．)

（3） 参加要求の転送

サブスクリプションリクエストを受け取ったノードは確率 p

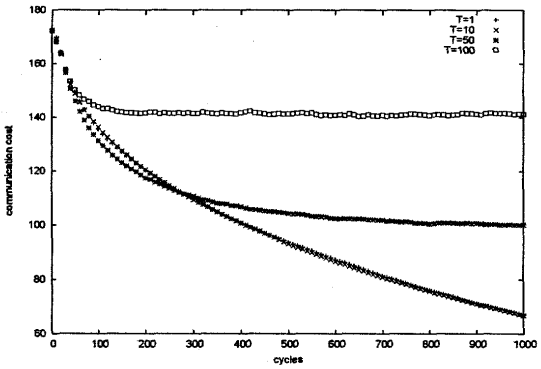

図 4　通信コスト (Waxman モデル，ノード数 5000)

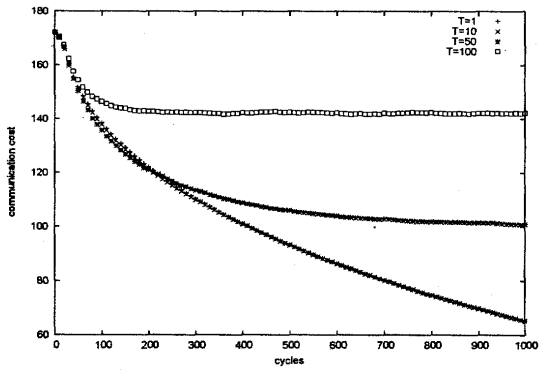

図 5　通信コスト (Waxman モデル，ノード数 10000)

でメッセージを保持する．p は隣接ノード数に依存し，隣接ノード数が少ないときは新しいノードとより接続しやすく，多いときは接続しにくい．今回は p は次数の逆数とした．メッセージを保持したノードは確率 1/2 で新しく参加するノード new とリンクをはる．リンクをはらない場合は，メッセージを破棄する．メッセージを保持しなければ，さらに隣接ノードからランダムにノードを一つ選び，サブスクリプションリクエストを転送する．

$E[M_{n-1}]$ をノード数 $n-1$ の辺の数とすると，平均次数は $E[M_{n-1}]/(n-1)$ で求まる．このとき，新しいノードがメンバーに加わったときのネットワークの辺の数は次の式で求まる．

$$E[M_n] = E[M_{n-1}] + \frac{E[M_{n-1}]}{n-1} + a + 1$$

これより $E[M_n] \approx (a+1)n\log n$ と近似できる．

各ノードが全ノード数を知らなくても平均次数は $(a+1)\log(n)$ になる．

5. 評価実験

この節ではシミュレーションによる結果を述べる．シミュレーターとして PeerSim [10] を用いた．10 回の実行のうち平均的なデータを載せる．パラメータ w を 100 に，a を 1 に設定する．下位物理ネットワークとして Waxman モデルと Transit-Stub モデルを使用し，Scamp で構築されたオーバーレイネットワークに対して Localiser を適用し，パラメータ T を変化させどのような影響があるかを調べる．評価する特性は (1) 通信コスト (2) 次数分布 (3) 耐故障性の 3 つである．1 サイクルで各々のノードで Localiser の処理を行っている．

5.1 通信コスト

オーバーレイ上の隣接ノード同士の距離を測った．図 4 と図 5 はノード数が 5000 のシステムを図 6 と図 7 はノード数が 10000 のシステムを表している．図は横軸が実験を行ったサイクル数を表し，縦軸が隣接ノード同士の通信コストの平均を表している．

図から温度パラメータ T が小さいほどより良い値に収束していき，T を小さくしても局所最適解にはおちいらないことが分かる．

5.2 次数分布

Localiser を適用していない場合と 300 サイクル繰り返した場合のオーバーレイネットワークの次数を調べた．

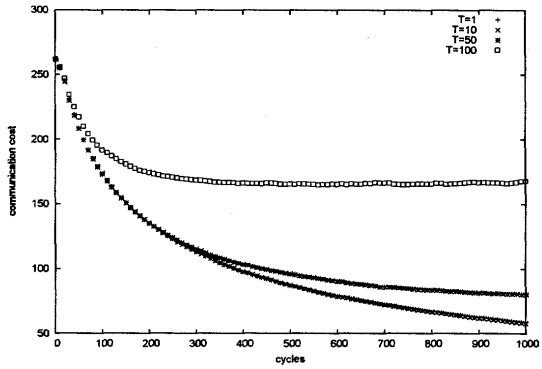

図 6　通信コスト (Transit-Stub モデル，ノード数 5000)

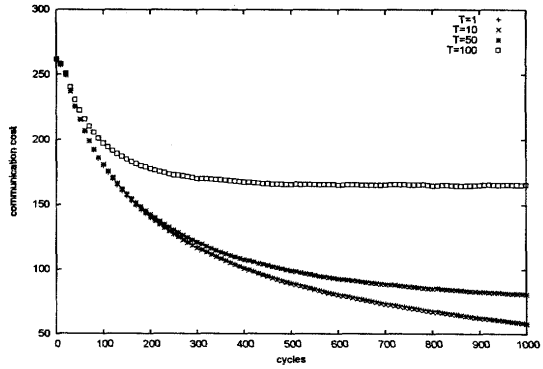

図 7　通信コスト (Transit-Stub モデル，ノード数 10000)

図 8 は Waxman モデルでの次数分布を表し，図 9 は Transit-Stub モデルでの次数分布を表している．ネットワーク全体のノードの数はどちらも 5000 である．図は横軸が次数を表し，縦軸がその次数を持つノードの数を表している．300 サイクル後はどちらのモデルでもノードの次数が $(a+1)\log(N)$ に均一化されている．パラメータ T を変化させてもほとんどかわりは無かった．

5.3 耐故障性

Localiser を適用していない場合と 300 サイクル繰り返した場合のオーバーレイネットワークの耐故障性を調べた．耐故障性を評価するために，故障したノードの数を増やし，連結しているネットワークの数を調べる．

図 10 は Waxman モデルでの耐故障性を表し，図 11 は

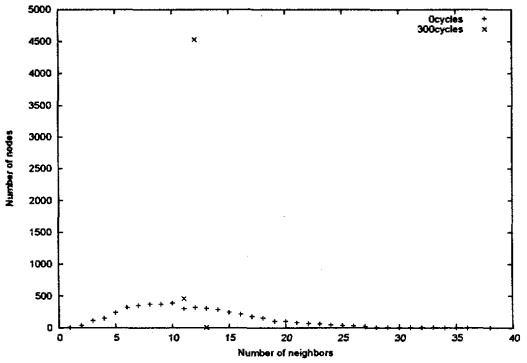

図 8　次数分布 (Waxman モデル，ノード数 5000)

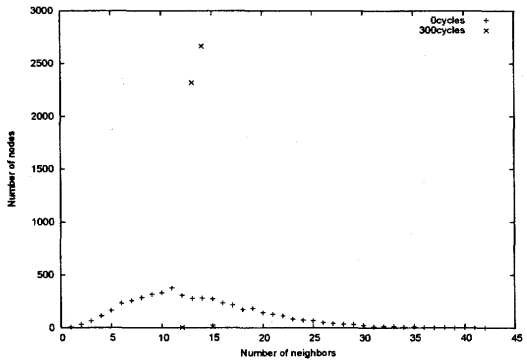

図 9　次数分布 (Transit-Stub モデル，ノード数 5000)

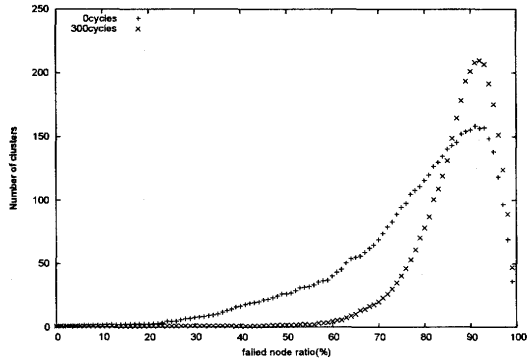

図 10　耐故障性 (Waxman モデル，ノード数 5000，T=1)

Transit-Stub モデルでの耐故障性を表している．図は横軸が故障しているノードの割合を表し，縦軸がその連結しているネットワークの数を表している．300 サイクル後はどちらのモデルでも約 50 パーセントのノード障害に耐えることができ，故障耐性が向上していることがわかる．パラメータ T を変化させてもほとんど変化は無かった．

6. ま と め

本研究では，メトロポリス法の特徴である，一時的に改悪となる暫定解への更新を認めることで局所解への収束を避ける戦略が，非構造化オーバーレイネットワークの構築に有効かどうかを調べるために，メトロポリス法を用いたアルゴリズムである Localiser を対象とし，シミュレーション実験を行った．シミュレーション実験の結果，現在の状態より良い解のときにだけ暫定解を更新する場合でも，局所解に収束しないことが分

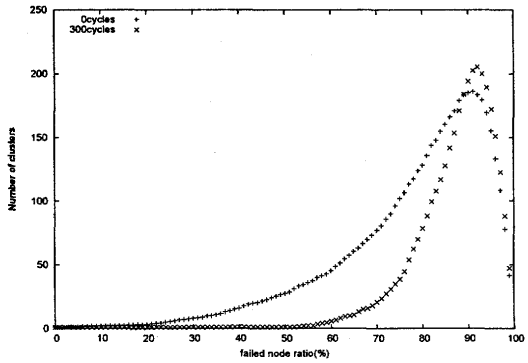

図 11　耐故障性 (Transit-Stub モデル，ノード数 5000，T=1)

かった．今後の課題として，その他のメトロポリス法を適用するアルゴリズムでもメトロポリス法が有効かどうかを調べることが挙げられる．

謝　　辞

本研究の一部は，グローバル COE プログラム「アンビエント情報社会基盤創成拠点」の支援を受けている．

文　　献

[1] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Schenker, "A scalable content-addressable network," Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM '01), pp.161–172, New York, NY, USA, ACM, 2001.

[2] I. Stoica, R. Morris, D. Karger, M.F. Kaashoek, and H. Balakrishnan, "Chord: A scalable peer-to-peer lookup service for internet applications," Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM '01), pp.149–160, New York, NY, USA, ACM, 2001.

[3] P. Maymounkov and D. Mazières, "Kademlia: A peer-to-peer information system based on the xor metric," First International Workshop on Peer-to-Peer Systems (IPTPS '01), pp.53–65, London, UK, Springer-Verlag, 2002.

[4] A.I.T. Rowstron and P. Druschel, "Pastry: Scalable, decentralized object location, and routing for large-scale peer-to-peer systems," Proceedings of the IFIP/ACM International Conference on Distributed Systems Platforms Heidelberg (Middleware '01), pp.329–350, London, UK, Springer-Verlag, 2001.

[5] L. Massoulie, A.M. Kermarrec, and A.J. Ganesh, "Network awareness and failure resilience in self-organising overlay networks," In Proceedings of the 22nd Symposium on Reliable Distributed Systems (SRDS 2003), pp.47–55, 2003.

[6] Z. Li, Z. Zhu, Z. Li, and G. Xie, "SAP2P: Self-adaptive and locality-aware p2p membership protocol for heterogeneous systems," Parallel, Distributed, and Network-Based Processing, Euromicro Conference on, vol.0, pp.229–236, 2008.

[7] B.M. Waxman, "Routing of multipoint connections," IEEE Journal on Selected Areas in Communications, vol.6, no.9, pp.1617–1622, 1988.

[8] E.W. Zegura, K.L. Calvert, and S. Bhattacharjee, "How to model an internetwork," In Proceedings of IEEE INFOCOM, pp.594–602, 1996.

[9] A.J. Ganesh, A.-M. Kermarrec, and L. Massoulié, "Peer-to-peer membership management for gossip-based protocols," IEEE Trans. Comput., vol.52, no.2, pp.139–149, 2003.

[10] "Peersim:a peer-to-peer simulator," http://peersim.sourceforge.net/.