様式 C-19

科学研究費補助金研究成果報告書

平成 22 年　6 月　11 日現在

研究種目：基盤研究(C)
研究期間：2007～2009
課題番号：19500058
研究課題名（和文）　実ネットワークとの親和性が高い仮想オーバーレイネットワーク構成方式に関する研究
研究課題名（英文）　A Study on Constructing Virtual Overlay Networks Being Aware of Underlying Physical Networks
研究代表者
石橋　勇人　（ISHIBASHI HAYATO）
大阪市立大学・大学院創造都市研究科・教授
研究者番号：70212925

研究成果の概要（和文）：
実ネットワークにおける RTT (Round Trip Time)やホップ数などの値に基づいてノードを階層的にクラスタリングし，その構造に対応した形で階層的な分散ハッシュテーブルを構成する方式を提案した．これによって，実ネットワークの局所性を活かした通信が可能となり，仮想オーバーレイネットワーク上での通信のオーバーヘッドを減少させることができる．

研究成果の概要（英文）：
This study proposes a method to construct a hierarchical distributed hash table based on node distance metrics.　The hierarchy is determined by a clustering method that classifies nodes on the Internet by some distance metric like RTT (Round Trip Time) or hop counts between nodes.　The proposed method reduces communication overhead between nodes on an overlay network making use of communication locality on the Internet.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2007 年度	1,100,000	330,000	1,430,000
2008 年度	1,200,000	360,000	1,560,000
2009 年度	700,000	210,000	910,000
年度			
年度			
総　計	3,000,000	900,000	3,900,000

研究分野：情報学
科研費の分科・細目：情報学・計算機システム・ネットワーク
キーワード：オーバーレイネットワーク，階層的分散ハッシュテーブル，P2P

１．研究開始当初の背景

近年，インターネットの一般ユーザへの普及は目覚ましく，扱う情報も極めて多岐にわたるため，インターネット上に流通する情報の量は増加の一歩をたどっている．これに加えて，流通する情報が音声や動画像などのマルチメディアデータへとシフトしつつあるために，従来のサーバ・クライアント型モデルによる情報の供給は，サーバの処理能力ならびにサーバの接続されるネットワーク帯

域の双方に関して物理的あるいはコスト的な限界に達しつつある．このため，情報処理を極めて高度に分散化可能なP2P(peer-to-peer) 型モデル，すなわち，一般のクライアントノードが同時にサーバとしても動作し，中央集権的なサーバを持たずに各ノードが自律的に動作する非集中型(de-centralized)モデルの重要性は高まる一方である．

特別な機能を有するサーバノードを持たず，各ノードが対等な機能を持って動作する純粋な P2P 型モデルでは，実際のネットワーク(インターネット)上にオーバーレイネットワーク(あるサービスを享受したいノードの集合によって構成される仮想的なサービスネットワーク)を構成することになる．

オーバーレイネットワークの構成方式としては，ノード間の接続形態に制約のない非構造化オーバレイと，何らかの制約に基づいてノード間の接続が決定される構造化オーバレイの 2 つのタイプがある．

非構造化オーバレイを用いた方式では，オーバーレイネットワーク上の資源を発見するための探索(lookup)において，ブロードキャスト的に探索メッセージを配布する必要があるために無駄なネットワークトラフィックが発生しやすく，また，探索の不確実性や探索に要する時間コストが不明確であるといった問題が指摘されている．

これらの問題を解決するために，分散ハッシュテーブル(DHT: Distributed Hash Table)に代表される構造化オーバレイを用いた P2P ネットワークが提案されており([Rat01], [Sto01], [Zha01]など)，現在の主要な研究対象となっている．

[Rat01] S. Ratnasamy, P. Francis, M. Handley, R. Karp, S. Shenker: A Scalable Content-Addressable Network, Proc. of ACM SIGCOMM 2001, pp.161-172, 2001.
[Sto01] I. Stoica, R. Morris, D. Karger, F. Kaashoek, H. Balakrishnan: Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications, Proc. of ACM SIGCOMM 2001, pp.149-160, 2001.
[Zha01] B. Y. Zhao, J. Kubiatowicz, A. D. Joseph: Tapestry: An Infrastructure for Fault-tolerant Wide-area Location and Routing, Technical Report UCB/CSD-01-1141, Computer Science Division, U.C. Berkeley, 2001.

２．研究の目的

DHT を用いた P2P 型ネットワークでは，探索が効率的であるばかりではなく，探索に必要となる時間コストが解析的に与えられるという特徴があり，その点では極めて好ましい特性を持つと言える．ただし，通常の DHT 方式の場合，オーバーレイネットワークにおけるノード間の近接性と実ネットワーク（インターネット）におけるノード間の近接性との間に関係がないことが多いため，オーバーレイネットワーク上では隣接するノードに対するアクセスが，極端な場合にはインターネット上では地球を半周する可能性があるという点において，ネットワークの利用効率に問題がある．

このため，実ネットワーク上における近接性を考慮したオーバレイ構成方式の提案として，[Str02], [Xu03]などの研究がある．これらでは，近接性の指標として RTT (Round Trip Time)を用いることによってトラフィックの局所化を図ろうと試みている．

しかし，オーバーレイネットワーク上の通信が実ネットワークに与える影響を考慮すると，RTT による近接性だけではなく，実ネットワークの構成形態に対する配慮を行うことが，ネットワーク帯域という資源を効率的に利用するためには重要であると言える．

ここで，実ネットワークの構成形態に対する配慮が重要である理由は，現在のインターネットの構造は，ノード間が一様に結ばれているわけではなく，多数のノードを収容する ISP (Internet Service Provider)同士が少数の結節点(IX: Internet eXchange)を介して接続される形態を持つためである．これによって，異なる ISP 間を跨がるトラフィックは，IX と ISP を結ぶ特定のリンクに集中することになる．したがって，そのようなトラフィックが増大することは，ISP 内部のトラフィックが増大することに比較して実ネットワークに対する影響が大きい．昨今急速に利用の拡大しつつある動画像配信のようなアプリケーションにおいては，トラフィック量が特に大きくなるため，この影響は無視できないものがある．

このように，現在のオーバレイ構築方式では，その下位に存在する実ネットワーク，すなわちインターネットの利用効率と言う点では非効率的となっている面が否めない．オーバレイネットワークが広く一般的に普及し，P2P 型ネットワークの真価を発揮するためには，この点を改善することが必要となる．

そこで，本研究課題では，実ネットワークの利用効率に配慮した新たなオーバーレイネットワーク構成方式を提案しようとするものである．

[Str03] J. Stribling, K. Hildrum, J. D. Kubiatowicz: Optimizations for Locality-Aware Structured Peer-to-Peer Overlays, Technical Report

UCB/CSD-03-1266, EECS, U.C. Berkeley, 2003.
[Xu02] Z. Xu, C. Tang, Z. Zhang: Building Topology-Aware Overlays Using Global Soft-State, Technical Report HPL-2002-281, HP Labs, 2002.

３．研究の方法

本研究計画の遂行に当たっては，大きく２つの課題を解決する必要がある．１つは実ネットワークの構造を推定することであり，もう１つは仮想ネットワークを構築する方式を考案することである．

前者を実現するためには，直接的に物理ネットワークから構成情報が得られれば最も確実であるが，実際にはそれは困難であるため，計測可能なパラメータを利用して推測することが必要となる．ここでいう推測とは，物理ネットワークを完全に同定することではなく，効率の良いオーバーレイネットワーク構築の手助けとなる情報を得ることを意味している．

後者の実現のためには，我々のグループにおいて研究を進めている階層的分散ハッシュテーブル構築方式を採用することを検討し，その改良を行った．この際，実ネットワーク上の通信における“距離”によってノードをクラスタリングする方式を利用し，クラスタリングによって得られた近接性の情報を利用して分散ハッシュテーブルを構成する方式について検討した．

４．研究成果

（１）実ネットワークの構造を反映するパラメータ

実ネットワークの構造を反映するパラメータを検討し，トポロジ情報，ノード間の RTT (Round Trip Time)，ノード間のホップ数，ノードの属するネットワークの AS 情報などが候補として考えられた．

一般に，下位ネットワークのトポロジ情報を直接得ることは困難である．AS 情報は，AS 間で経路情報を交換しているルータであれば取得が可能であるが，一般のノードが直接得ることは難しいとも言える．ただし，外部の経路情報データベースを利用すれば取得が可能である．

AS 情報を利用することによって，最上位のレベルでインターネットをクラスタリングすることが可能となるが，AS 内部のクラスタリングにはより詳細な情報が必要となる．このために，RTT やホップ数の情報を利用する．

（２）インターネットノードのクラスタリング方式

RTT やホップ数の情報をオーバーレイネットワークの構築において利用するために，[Ued06]に基づく方式を用いることを検討した．[Ued06]では，RTT やホップ数のような，任意のノード間の距離として扱えるメトリックに基づいてノードを階層的にクラスタリングしている．これによって得られたクラスタを次で述べる方式においてオーバーレイネットワークの構成に用いることによって，ノード間の距離を意識したオーバレイの構築が可能となった．

また，これに加えて，インターネット上の各ノードに対して絶対座標を与える方式[Deb04]をクラスタの構築に利用する手法を提案した．シミュレーションによって，この方式においても良好なクラスタリング結果が得られることを確認した．この場合には，RTT の情報を元にノードに絶対座標を与えておき，その絶対座標を利用してノード間の距離を算出するため，物理ネットワークから得られる情報である RTT を，直接ではなく間接的に使用している形になるが，直接の計測結果を持たない 2 ノード間においても距離を算出できるという利点があり，システム全体として計測に必要なオーバーヘッドが少なくなる．

（３）オーバーレイネットワーク構成方式

ここでは，新たに提案するオーバーレイネットワーク構成方式の概要について述べる．この方式は，我々が研究してきた階層的分散ハッシュテーブル構成方式をベースに改良を加えたものである．従来の方式では，DHT リングを構成するノード数にばらつきがあった場合に検索に要するホップ数が増大するという問題があったが，その問題を解決している．ノードの近接性のようなオーバーレイネットワーク外部の要因によって決定されるパラメータを持ち込んだ場合には，リングの構成ノード数が均一になるとは限らないため，このような改良が重要である．

① データ構造

本方式は，Chord[Sto01]をベースとしている．（２）で述べた方式を用いることによってネットワーク距離に基づいて構成したクラスタの各々に対して，Chord リング相当の仮想リングをそれぞれ１つずつ生成し，そのクラスタに所属するノードをリング上に配置する．各ノードは，一意なノード ID を保持し，所属するすべてのクラスタのリングに同時に参加している．最下層のリングでは Chord と同様に隣接ノードへのポインタである successor と predecessor，ならびに検索を高速化するための finger table を保持しているが，より上位の階層では，それらの代わりに次に述べる Sibling Predecessor Set と Sibling Successor を保持している．

①-1 Sibling Predecessor Set

Sibling Predecessor Set(SPS)は兄弟クラスタに所属するノードへのポインタ集合である．図 1 のように，Level i のノード (N1),

Level (i - 1) での predecessor (N1.pred) との間にある兄弟クラスタ上のすべてのノードへのポインタを保持する．これにより兄弟クラスタ間でノード数に偏りがあっても 1 ホップでそのレベルにおけるルートノードへ到達できる．
なお，SPS の大きさは有限であり，その適切な大きさについては ③-1 で述べる．

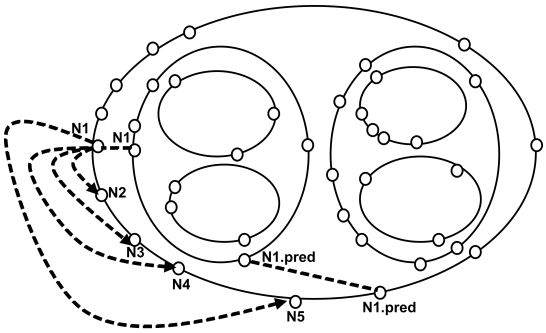

図 1 Sibling Predecessor Set

①-2 Sibling Successor
Sibling Successor(SS)は，SPS とは逆向きのポインタである．SS は，SPS の維持のために使用する．

② アルゴリズム
ここでは，lookup と join のアルゴリズムについて述べる．leave や stabilize など，その他のオペレーションのアルゴリズムについては省略する．
②-1 lookup
lookup は以下のように行う．まず検索するノードが所属する最下層のリングで Chord と同様のルーティングを行い，最下層でのルートノードを探す．ルートノード上に目的のデータが見つからない場合，階層を 1 段上がって SPS を用いてルートノードへ 1 ホップで到達する (図 2)．データが見つかるか，最上層のルートノードまで到達するまでこれを繰り返す．

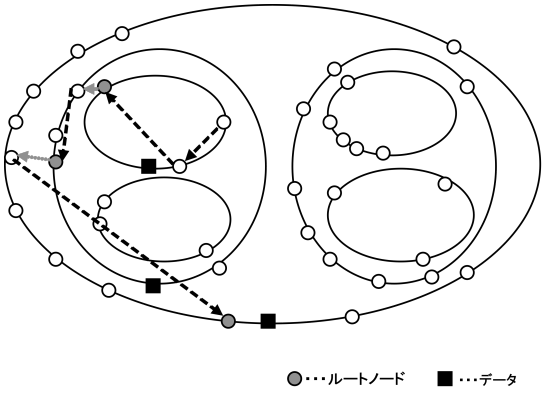

図 2 lookup

②-2 join
ノード(N)が DHT に参加する場合，最下層は Chord のアルゴリズムで参加する．次に，successor (N.suc)は，自身の SPS から N の両隣に位置する ノード(図 3 の A,B)に N が参加したことを伝える．これにより，N と N.suc の間にあるノード(A) は，SPS に N を追加し(①)，間にないノード(B) は SS を N.suc から N に付け替える (②)．同時に N は，N.suc の SPS から自身の SPS,SS を作成する(③)．最後に，N.suc の SPS から N 以降のノードを削除する．これを最上層まで繰り返す．

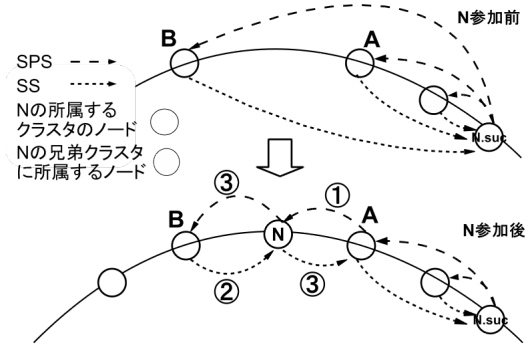

図 3 ノード N の join 操作

③ 考察
ネームサーバ間でのフルメッシュのネットワーク距離を測定した King Data Set[Gum02] を用いて，文献 [Ued06] のクラスタリングを乱数の種を変えながら 10 回行った．ノード数は 1740 であり，深さは 10，子クラスタの最大数は 2 とした．
③-1 SPS の適切な大きさ
実験の結果，兄弟クラスタ間のノード数の比は最大 278 であった．ノード ID が Chord リング内で必ずしも一様に分布していないことを考慮しても，SPS の大きさはこの比の数倍程度でよいと考えられる．
③-2 Crescendo との比較
上の方法で生成したクラスタ上で，既存の類似研究である Crescendo[Gan04]と提案手法のそれぞれについて，最上位クラスタのルートノードまでの lookup のホップ数をシミュレーションによって測定した．その結果，Crescendo では最大ホップ数が 26.1，平均ホップ数が 13.8 であったのに対して，提案手法では最大ホップ数が 7.9，平均ホップ数が 7.0 となり，提案手法の有効性が確認された．

[Ued06]上田達也，安倍広多，石橋勇人，松浦敏雄：P2P 手法によるインターネットノードの階層的クラスタリング，情報処理学会論文誌，Vol. 47，No. 4，pp. 1063-1076，

2006.
[Dab04] Dabek, F., Cox, R. and Morris, R.: Vivaldi: A Decentralized Network Coordinate System, Proceedings of the ACM SIGCOMM 2004, pp.15-26, 2004.
[Gum02] K. Gummadi, et al.: King: Estimating Latency between Arbitrary Internet End Hosts, SIGCOMM Internet Measurement Workshop, 2002.
[Gan04] P. Ganesan, et al.: Canon in G major: Designing DHTs with Hierarchical Structure, 24th International Conference on Distributed Computer Systems, pp. 263-272, 2004.

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔学会発表〕（計１件）
① 尾崎永径, 上田達也, 安倍広多, 石橋勇人, 松浦敏雄: 階層的クラスタを用いた DHT におけるノード数の偏りによる影響の排除, 情報処理学会第 70 回(平成 20 年)全国大会第 3 分冊, pp. 177-178, 情報処理学会, 2008.

６．研究組織
(1)研究代表者
石橋　勇人　(ISHIBASHI HAYATO)
大阪市立大学・大学院創造都市研究科・教授
研究者番号：70212925

(2)研究分担者
なし

(3)連携研究者
松浦　敏雄　(MATSUURA TOSHIO)
大阪市立大学・大学院創造都市研究科・教授
研究者番号：40127296
安倍　広多 (ABE KOTA)
大阪市立大学・大学院創造都市研究科・准教授
研究者番号：40291603

(4)研究協力者
上田達也 (UEDA TATSUYA)
大阪市立大学・大学院創造都市研究科・博士（後期）課程
尾崎永径 (OZAKI HISAMICHI)
大阪市立大学・大学院創造都市研究科・修士課程（平成 19 年度）