

# 情報環境の変化に対応した複数の動的エージェントによるネットワーク内情報共有

研究代表者 菅原真司 名古屋工業大学 大学院 工学研究科 准教授

## 1 はじめに

近年、インターネット上に蓄積される情報は極めて多種多様、かつ大容量となった。しかし、個々のネットワークユーザにとっては不要な情報も大量に存在するため、真に必要とする情報を正確に取得することが困難となることがあり、効率的な情報探索技術が重要となってきた。

また同時に、複数のネットワークユーザが自分の持つ情報をインターネット上で公開し、他のユーザに対して互いに提供し合うことで情報の共有を行い、それらコンテンツの利用効率を高める、或いは、価値の高いコンテンツの流通を促進し、ネットワークの利便性を向上することが一般的になってきている。このような状況では、通常、各ユーザの所有する端末計算機同士が対等な立場で接続される Peer-to-Peer (P2P) と呼ばれるネットワークを用いてコンテンツの共有、交換が行われており、従来のクライアント・サーバ型のシステムと比較して耐故障性、スケーラビリティなどの点で優れていると言われている。

以上のことから、P2P ネットワーク上で提供、共有されている価値の高いコンテンツを効率的に探索、発見し、取得することは非常に重要であり、そのための技術を確立することは急務であることは明らかである。しかし、P2P ネットワーク上の情報探索では、ピアと呼ばれる P2P を構成する個々の計算機が、ネットワークへの加入、離脱を自律的に行う可能性があり、ユーザがあるコンテンツを取得したいと考えた時点で、ネットワークを構成するすべてのピアが持つコンテンツの正確なリストを提供することが困難であるため、必ずしもクライアント・サーバ型の情報共有と同等以上のサービス品質や信頼性を保証できないことが問題点として存在する。これを解決する手段として、P2P ネットワークのトポロジと、それを構成する全ピアの所有コンテンツのリストを確実に把握するための専用のサーバを設けることが考えられるが、ネットワーク全体の規模が大きくなった場合のスケーラビリティの実現や、サーバへの負荷、障害に対応するフォールトトレランスの確立などに問題を残し、現実の実装する場合には高コストになりやすい側面がある。

そこで、本研究では、ユーザがコンテンツの共有を行う P2P ネットワークにおいて、そのネットワーク内を巡回する複数の動的エージェントを常駐させ、各ピアの接続状況からなるネットワークトポロジ情報と、各ピアが所持するコンテンツの情報を常時収集するシステムを提案している。収集された情報は、各ピアによって取得され、ピア上のユーザから必要なコンテンツのリクエストが発生した際に探索先を決定するために利用される。このシステムを用いたコンテンツ共有の有効性または効率に関しては、その探索コスト、および要求コンテンツの取得成功率を、現実のコンテンツ共有を念頭に置いた種々の条件を設定した上で、計算機シミュレーションを用いて試算し、評価している。

これ以降の本報告の内容は以下の通りである。まず 2 において本研究全体を通して共有する仮定条件と、ゴールについて定式化する。次に、3 において最初の検討として行った、複数の管理ノードを用いた動的エージェント間の情報共有による情報探索について述べる。続いて 4 では 3 の手法の問題点を克服するための手法である、単一管理ノードを用いた動的エージェント間の情報共有によるコンテンツ探索に関する要点を説明する。5 では、本研究の最終的な目的に合致した内容としての、動的エージェントによる Peer-to-Peer ネットワーク内コンテンツ共有について述べる。6 では、エージェントが複数の情報源から情報を収集する際の、誤情報の効率的な排除方法について検討を行った結果として、不確かな多数の情報源から信頼できる情報を抽出する手法の検討について述べる。最後に 7 で本研究全体の総括を行う。

## 2 問題の定式化

本研究で仮定する環境を以下に示す。

- 多数のノードとそれらを接続するリンクから構成される物理的なネットワークを仮定し、各ノードをピアとする P2P ネットワークを構成するものとする。

- 各ピアは他のピアに提供するコンテンツを保持している。
- 各ピア上のユーザは、ある頻度で他のピアに存在する可能性のあるコンテンツを取得したいと考える。
- 各ピアは、これらが構成する P2P ネットワーク上を自由に移動できる動的エージェントの動作を許容し、エージェントから要求があれば、自身のネットワークへの接続状況や、保持するコンテンツのリストをこれに与える。
- エージェントの動作には一定のコストが生じる。

以上の条件を仮定した上で、ユーザの要求するコンテンツを発見するために要するコスト総和の期待値を最小とする、あるいは、単位コスト当たりのユーザの要求コンテンツの期待取得数を最大とする情報共有手法を求めるところを目的とする。

### 3 複数の管理ノードを用いた動的エージェント間の情報共有による情報探索

#### 3-1 まえがき

動的エージェントによる効率的なコンテンツの探索・取得を行うためには、エージェントの動作によって発生する通信コスト、計算処理コスト、時間コスト、およびユーザの要求しているコンテンツ（ターゲット）が各ピアに存在する確率を考慮することが重要である。この節では、エージェントが各ピアを巡回して要求コンテンツを発見するための効率的な手法について検討する。

本研究全体の最終目的は、動的エージェントがネットワーク内を常に巡回してネットワークトポロジやコンテンツの位置を効率的に把握し、ユーザのコンテンツ要求に対応するシステムを構築することであるが、ここではまず単純に、事前に与えられている各ピアにおけるターゲットの存在確率を元に、複数のエージェントが協調してターゲットを発見するための効率的な手法を求めよう。

#### 3-2 従来手法の問題点と提案手法の設計方針

文献[1]では、コストのみを考慮した情報探索手法が提案されている。また、文献[2]ではすべてのノードを効率よく探索するための探索方式が提案されている。いずれの手法も各ノードのターゲット存在確率やコストがすべて事前に与えられることを前提とした手法となっているが、ネットワーク上に各ノードに関する情報を集中管理する機能が存在しないという条件下では、これは必ずしも現実的な仮定とは言えないため、効率的な情報探索を行うためには探索手法に各ノードに関する情報を取得する何らかの手順を加える必要がある。

そこでこの節では、フラッディングを用いた問い合わせによって、エージェントの存在するカレントノードから TTL (Time To Live) ホップ内のノードに関する情報を取得することを考える。これにより、フラッディングを行った範囲内のノードに関しては、従来の情報探索手法と同様の手法で探索を行うことができる。

エージェント間におけるターゲットの探索状況の共有は重要であるが、一般には各エージェントにとってネットワークトポロジの全体像と他のエージェントの位置情報が未知であるため困難である。この問題を解決するために、この節では、エージェントが最初に配置されたノードをそのエージェントの管理ノードとし、管理ノード間の通信によってエージェント間の情報共有を可能とする。

#### 3-3 提案手法の動作概要

この節の提案手法では、各エージェントがカレントノードでの探索、次の探索ノードの決定、次の探索ノードへの移動を繰り返してターゲットを発見するまで探索を継続する。また、これらの動作を繰り返す際に、フラッディングによりカレントノードから一定範囲のノードの情報を取得する。フラッディングを行うのは 1 回目のカレントノードでの探索後、前回のフラッディング時のノードから一定ホップ数以上移動している場合である。フラッディングを行う範囲は、パラメータ TTL で設定する。また、一定時間が経過する毎に管理ノード間での通信、管理ノードとエージェント間の通信を行い、エージェント同士での探索状況に関する情報共有を行う。

#### 3-4 評価

計算機シミュレーションにより提案手法を評価した。評価には、通信コスト、処理コスト、時間コストを用いている。

##### (1) 比較方式

提案方式と比較するための方式として、独立方式、スタンプ方式[3]を用いた。ただし、次の探索ノードを決定するアルゴリズムは提案手法と同一のものを用いている。

- 独立方式

すべてのエージェントが、エージェント間で探索状況の情報交換を行わずに、各々独立に情報探索を行う方式である。

- スタンプ方式

エージェントがあるノードでの探索を終えて次の探索ノードへ移動する前に、カレントノードにスタンプを残してから移動を行う方式である。スタンプには移動先ノード番号が記されており、エージェントはカレントノードでの探索を開始する前にスタンプの有無を確認する。もしカレントノードにスタンプが存在するならば、カレントノードでの探索を行わず、スタンプに記されている移動先ノード以外を次回探索ノードとして選択し、移動する。

## (2) シミュレーション条件

シミュレーションではネットワークトポロジに BA モデルトポロジを用いた。共通のパラメータは以下の表 1 に示す通りである。また、2 つの条件を設定し、それぞれ異なる環境で各方式を評価した。条件 1 では、エージェント数を 3 に固定し、ターゲットの各ピアにおける存在確率は Zipf 分布として、その最大確率を 0.2 ~ 0.9 の変数としている。また、条件 2 では、エージェント数を 1~9 の変数とし、ターゲットの存在確率は最大確率 0.5 の Zipf 分布に固定している。

表 1：共通のパラメータ

ネットワーク形態	BA モデルトポロジ
シミュレーション回数	1,000
総ノード数	100
リンクの通信コスト (エージェントの移動時)	300
リンクの通信コスト (1 ホップ間のフラッディング)	0.02
エージェントの移動時間 (1 ホップ)	300
ノードでの探索処理コスト	500~1,000 の一様分布
ノードでの探索処理時間	500~1,000 の一様分布
TTL 値	3
エージェントから管理ノードへの情報送信間隔	1,000
管理ノードからエージェントへの情報送信間隔	1,000
管理ノードから他管理ノードへの情報送信間隔	1,000

## (3) 結果と考察

図 1~3 に条件 1 のシミュレーション結果を示す。なお、値は 1,000 回のシミュレーションの平均値である。

図 1 に条件 1 における通信コストを示す。すべての手法において、ターゲットの存在確率の最大値が大きくなるにつれて通信コストが小さくなる傾向にある。これは、ターゲット存在確率の最大値が大きいほど、ネットワーク上にターゲットが存在する確率も高くなり、より少ない探索ノード数でターゲットの発見に至る傾向があることが原因である。また、通信コストは、提案手法、独立方式、スタンプ方式の順に抑制されている。スタンプ方式では、複数のエージェント間で他のエージェントのターゲット発見情報を共有できないため、すべてのエージェントがターゲットを発見するまで探索を終了せず、非常に効率の悪い探索になっている。さらにスタンプ方式では、探索したノードにスタンプとして移動先ノードを記録し、その移動先へは探索に行かないような制御をするため、余計にターゲットの発見が遅れてしまう。

図 2 に条件 1 における計算処理コストを示す。これは前述の通信コストと同様の理由でターゲットの存在確率が高くなるにつれて減少する傾向にある。計算処理コストはスタンプ方式が最も抑制できているが、この方式では各ノードでエージェントが探索を開始する前に、そのノードでのスタンプの有無を確認し、無駄な探索を行わないことがその主たる原因である。

図 3 に条件 1 における時間コストを示す。これも二つのコストと同様にすべての手法においてターゲット

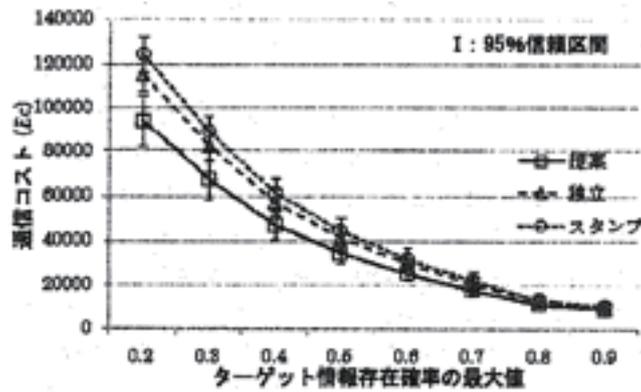


図1 通信コストとターゲット情報存在確率の最大値との関係

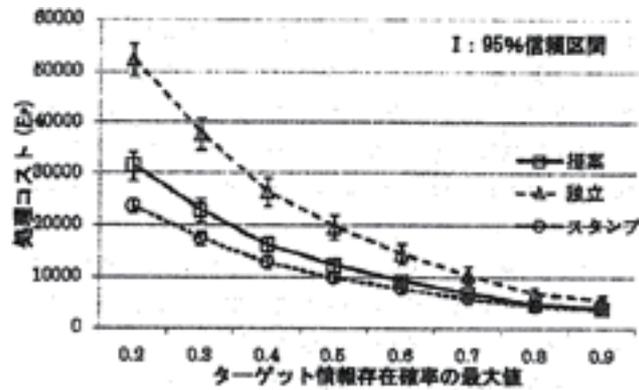


図2 処理コストとターゲット情報存在確率の最大値との関係

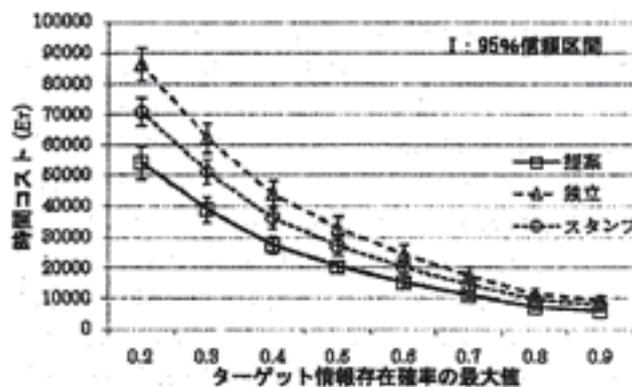


図3 時間コストとターゲット情報存在確率の最大値との関係

の存在確率が高まるにつれて減少している。最も時間コストを抑制できるのは提案手法であるが、これは重複探索を避けられない代わりにスタンプ方式のような遠回りをするのがなく、ターゲット発見までの時間が抑制されることがその理由である。

図4～6に条件2のシミュレーション結果を示す。

図4は通信コストの結果であり、すべての手法においてエージェント数に比例してコストが増加している。理由はどの方式でも各エージェントが等しく一定の通信コストを要するためであるが、提案手法が最もコストを抑制できている。

図5は計算処理コストで、すべての手法において通信コストと同様にエージェント数の増加に伴いコストが増加している。提案手法とスタンプ方式については、コストの増加傾向が鈍化している。これは両手法と

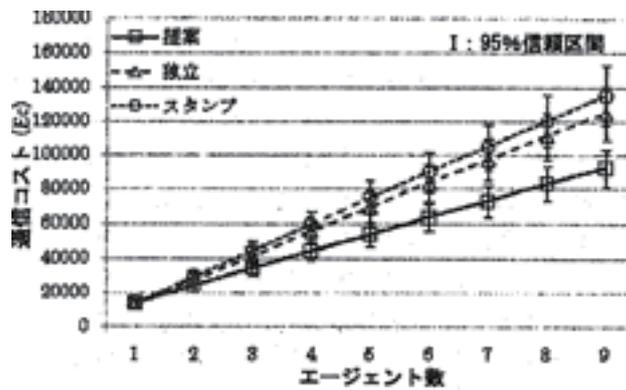


図4 通信コストとエージェント数との関係

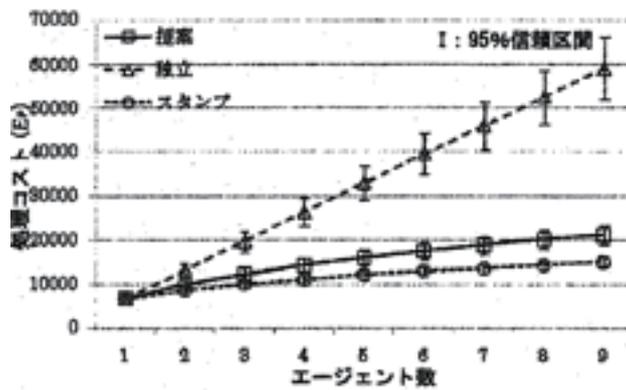


図5 処理コストとエージェント数との関係

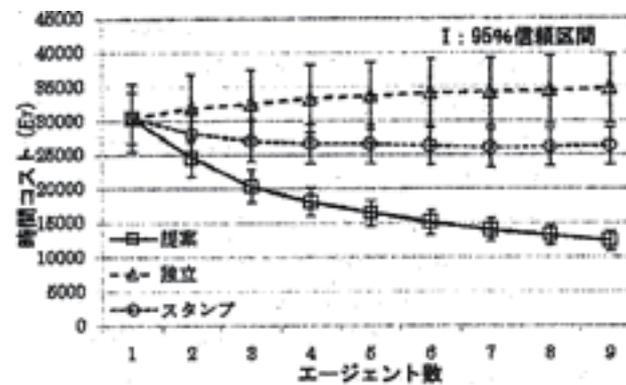


図6 時間コストとエージェント数との関係

も、エージェント間の情報共有を何らかの方法で行っているため、重複探索を抑制できているためだと考えられる。

図6に条件2における時間コストを示す。エージェント数が増加するに伴い、提案手法はコストが大きく減少している。これは、エージェント間の情報共有が他の方式よりもうまくできているためで、エージェントが多いほど短い時間でターゲットを発見できることがわかる。

### 3-5 むすび

提案手法はこの節で行ったシミュレーションの範囲では、他の手法と比較して通信コスト、時間コストを最も抑制することができた。特に、ターゲットの存在確率が小さいほど、また、一定数まではエージェント数が多いほど有効であることが示された。しかし一方で、処理コストについてはまだ改善の余地がある。す

べてのコストをトータルに評価する場合には、この節での提案手法が適用可能な条件が存在すると考えられる。

#### 4 単一管理ノードを用いた動的エージェント間の情報共有によるコンテンツ探索

##### 4-1 まえがき

前節の結果をもとにして、複数のエージェントによるターゲットコンテンツの効率的な探索手法について検討を続けた。これについて下記の結果を得た。

##### 4-2 従来手法の問題点

前節の手法では、あるエージェントの探索状況を他のエージェントに伝えるために、自身の管理ノードと通信相手の管理ノードを介して通信を行う必要があり、探索状況の伝搬に無視できない遅延が生じ、重複探索が発生してしまう可能性があった。そこで、この節に示す探索方式では、従来はエージェントの数だけ存在していた管理ノードを一つに集約し、各エージェントに関する情報はすべてその単一の管理ノードが管理する方式に改めた。

##### 4-3 提案手法の動作概要

この手法では、各エージェントがカレントノードでの探索、次回探索ノードの決定、次回探索ノードへの移動を繰り返し、これを、ターゲットを発見するか、全ノードを探索し終えるまで継続する。またこれらの動作の間に、フラディングによりカレントノードから一定の範囲のノードで把握できる情報を取得する。なお、フラディングで得られる情報は、当該ノードにおけるターゲット情報存在確率、探索時の処理コストと時間コスト、隣接するノードとの通信コストと時間コストおよび隣接するノード番号とし、以降ノード情報と称する。

フラディングを行うのは、探索開始直後のカレントノードでの探索後、最後にフラディングを行ったノードから一定ホップ数以上移動している場合とする。フラディングを行う範囲は、パラメータ TTL で設定する。また、エージェントが既に探索を行ったノードを示す探索済ノードリスト、および次回移動予定先ノードを随時管理ノードに送信し、その際、他のエージェントの探索状況の情報共有を行う。

以下の図 7 に、提案手法の動作アルゴリズムのフローチャートを示す。

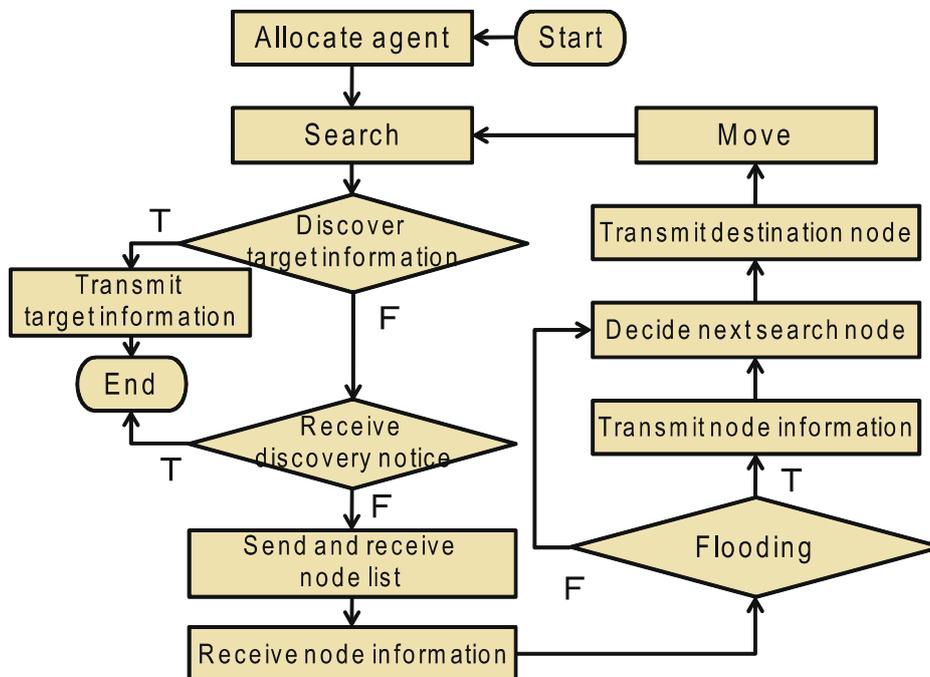


図 7 エージェントの動作フロー

なお、次回探索ノード決定アルゴリズムは、カレントノードから近く、かつターゲット情報存在確率が高いノードを優先して選択する貪欲戦略を用いる。具体的には、各ノードのターゲット存在確率をカレントノードからそのノードまでのホップ数で除した値を計算し、その値が最も大きいノードを選択している。

#### 4-4 評価

前節と同様に、計算機シミュレーションを用いて提案方式の有効性を議論する。

##### (1) 比較方式

提案方式と比較するための方式として、前節で用いた独立方式、スタンプ方式に加えて、ロック方式[3]と複数管理ノード方式（前節の提案方式）を用いた。次の探索ノードを決定するアルゴリズムは提案手法と同一のものを用いている点も前節と同様である。下記にロック方式について追加で説明を加える。

##### ● ロック方式

エージェントがノードでの探索を終えて次回探索ノードへ移動する際に、すべての他エージェントに対して移動しないようにロックをかけてから自らの移動処理を行う。ロックをかけられたエージェントは、かけたエージェントの移動処理後にロックが解除される。解除後は、予約されたエージェントから順に移動処理を実行する。ただし、エージェント間でロックに関する情報を直接通信する必要があるため、探索開始後に一つのエージェントがネットワーク全体に対してフラッディングを行い、その結果を他エージェントに送信する方法をとる。

##### (2) シミュレーション条件

シミュレーションで用いた条件は、前節のものと同じである。

##### (3) 結果と考察

図 8～10 に条件 1 のシミュレーション結果を示す。

図 8 は通信コストに関する結果であるが、結果は前節と同様である。前節の提案方式よりもこの節の提案方式の結果が改善されている理由は、管理ノードの持つ情報が一元化され、前節のように正確な情報がエージェントに伝達されるまでのタイムラグが解消されているからである。

図 9 は処理コストに関する結果である。ここでもエージェントが管理ノードから受け取る情報にタイムラグがなくなったことで状況が改善し、スタンプ方式よりも小さい処理コストとなり、比較している方式の中では最も効果が大きいと考えられるロック方式と同程度まで抑制できている。

図 10 は時間コストであるが、これも提案方式が最も抑制できている。新たに加えたロック方式は時間的にロスが大きいことは自明であるが、前節の提案方式との比較においても、やはり管理ノードが持つ情報のタイムラグの解消により優越していることが明らかになった。

図 11～13 に条件 2 のシミュレーション結果を示す。

図 11 は通信コストに関する結果であり、ロック方式はエージェントの数によらず、一貫して高いコストを必要とする。これは、この方式では、探索開始後に一つのエージェントがネットワーク全体に対してロックをかけるためのフラッディングを行う必要があるためである。それ以外は前節と同様の結果であるが、この節の提案方式は前節のそれと比較してコスト的に大きく改善している。

図 12 は処理コストであるが、この結果も前節とほぼ同様である。ロック方式も処理コストに無駄が少ないため、スタンプ方式と同様にコストを抑制できており、提案方式を加えた三者はほぼ同程度に良い結果となっている。ここにおいても、提案方式は前節のそれに対して優越している。

図 13 は時間コストとエージェント数の関係である。独立方式に加えてロック方式もエージェント数が増加するに伴い、コストが増大している。ロック方式では複数のエージェントが同時に移動できないため、各エージェントの拘束される時間が長くなることがその原因である。提案方式は前節同様エージェントが増えるに従って時間コストを減少させる傾向にあり、コスト抑制効果も前節の提案方式を上回っていることが確認できる。

#### 4-5 むすび

この節では、前節で提案した方式を改善し、複数の動的エージェントを用いたネットワーク内情報探索に

において、全体のネットワークポロジが未知である環境にあっても、エージェント間の探索状況の共有を効率的に行う方式を求めた。この節で用いた条件下では、提案方式は、比較しているどの方式よりも通信コスト、時間コストを抑制できており、処理コストについてもロック方式やスタンプ方式と同程度に抑制できている。また、前節の提案方式に対してすべての条件において優越しており、改善の効果が見られた。

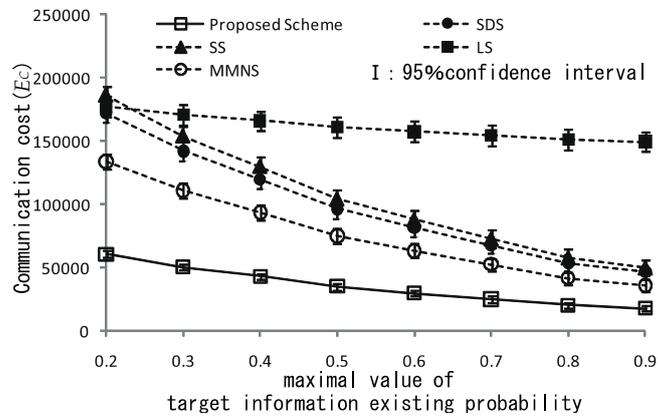


図8 通信コストとターゲット存在確率の最大値との関係

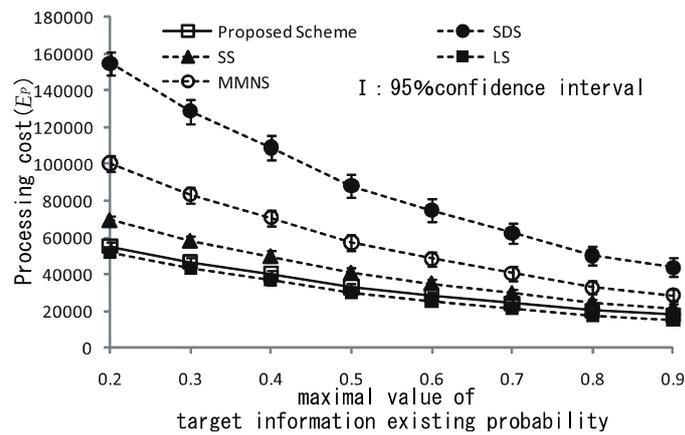


図9 処理コストとターゲット存在確率の最大値との関係

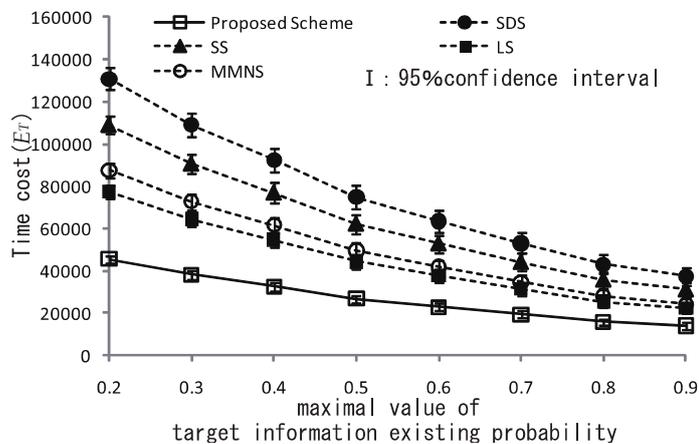


図10 時間コストとターゲット存在確率の最大値との関係

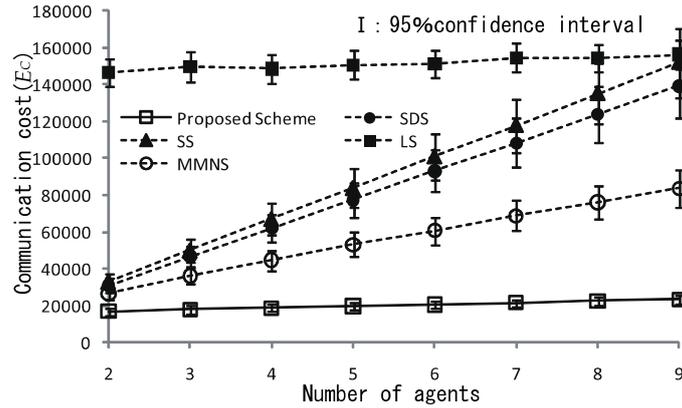


図 11 通信コストとエージェント数との関係

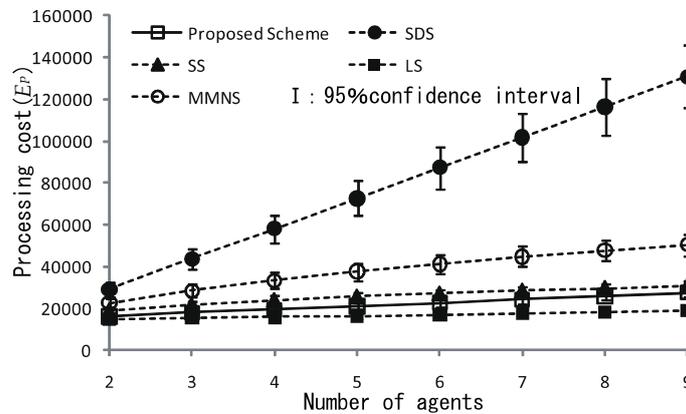


図 12 処理コストとエージェント数との関係

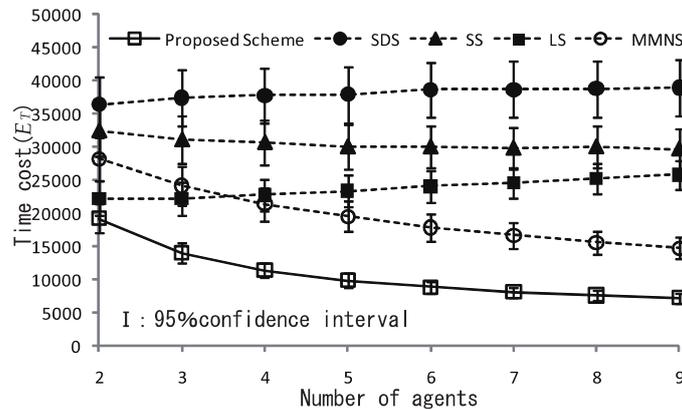


図 13 時間コストとエージェント数との関係

## 5 動的エージェントによる Peer-to-Peer ネットワーク内コンテンツ共有

### 5-1 まえがき

前節までの検討では、複数の動的エージェントが、ユーザからのコンテンツ要求に従って、各ピアにおけるそのコンテンツ（ターゲットコンテンツ）の存在確率をもとに探索を行う際のコスト効率を追求した。本研究の最終的な目的は、複数の動的エージェントが協調してネットワーク内を巡回し、時間的に変化するネットワークトポロジを把握しながら流動的なコンテンツの所在に関するリストを作成し、その精度を高いまま維持することで、ユーザからのコンテンツ要求に正確に答えられるシステムを確立するところにある。そ

のための準備として、前述の検討結果は非常に有益である。

この節で述べる検討は、最終的な目的に直接取り組むものと位置づけている。ここで求められるエージェントは、ネットワーク管理を行うことができる自律的な主体[4]であり、これまでにいくつかの研究が行われている。例えば文献[5]では、単一のモバイルエージェントを用いてネットワーク上のピアが所有する情報を収集することを目的として、単位時間当たりの平均費用を最小化する最適エージェント巡回政策を求める手法を提案している。この手法では、エージェントが移動することで生じる単位時間当たりのコストを既知とし、巡回政策を決定するために、マルコフ決定過程における政策反復法を用いている。また、単一のモバイルエージェントではなく、複数のエージェントを用いたネットワーク管理の例として文献[6]がある。これは、エージェントを用いた情報収集において、情報統合エージェントと情報収集エージェントという二種類の役割を与えることで、効率的な管理システムを構築することを意図している。

ここで提案するシステムは、上記のネットワーク管理とは異なり、ネットワーク上でピアの加入・脱退、コンテンツの新規加入・削除・更新が生じるような動的なネットワーク環境において、複数の動的エージェントを用いることで効率的なコンテンツ共有を行うことを可能とするものである。

## 5-2 仮定する環境

この節では、以下の仮定を置く。

- ピア上には複数のコンテンツが存在し、コンテンツの追加、削除が生じる。
- 各ピアは、ネットワークトポロジとコンテンツのメタ情報（どのピアにどのコンテンツが存在するか）を記述したリストを保持し、初期状態では隣接ピアのリンク情報のみを知るものとする。
- 各ピアはエージェントが最後に移動した時刻と現在の時刻との差である探索間隔を保持する。
- エージェントはピアが保持するリストと同様のリストを保持するが、初期状態ではそのリストは空であるものとする。
- ネットワーク上のクエリ伝播はTTLにより上限回数を制限されるものとする。

## 5-3 提案手法

従来のフラッディング方式と異なり、エージェントがネットワークを巡回することで効率的なコンテンツ検索を実現する。以下にコンテンツ要求時のアルゴリズム、エージェント巡回アルゴリズムの要約を示す。

### (1) コンテンツ要求時のアルゴリズム

1. あるピアにおいて、コンテンツ要求が生じるか、他のピアからのクエリを受信した場合、ピアが保持するリストを参照し、コンテンツを保持しているピアが記されているなら、そのピアにクエリを転送する。そうでない場合は、隣接するピアのうち、探索間隔の小さいものにクエリを送信する。クエリ転送後、 $TTL \leftarrow TTL - 1$  とする。
2. クエリを受信したピアは、要求されたコンテンツを保持しているかどうかを確認する。コンテンツを保持している場合、その結果を通知し、終了する。コンテンツを保持していない場合、 $TTL \neq 0$  ならば1に戻る。  $TTL = 0$  ならば、検索失敗の結果を通知し、終了する。

### (2) エージェント巡回アルゴリズム

エージェントの巡回方式を、情報管理フェーズと情報共有フェーズに分割し、変数  $t$  (初期値 0) がエージェントに定めた閾値  $T$  に達した場合、一方のフェーズから他方へ切り替える。

1. エージェントは自身のリストを参照し、隣接ピアの中から移動するピアを決定する。情報管理フェーズである場合、探索間隔が最大のピアを選択する。情報共有フェーズである場合、他エージェントが記した探索間隔が最小のピアを選択する。該当するピアがない場合、情報管理フェーズと同様の方法でピアを選択する。
2. エージェントは移動する度に、自身のリストとピアが保持するリストを更新し、 $t \leftarrow t + 1$  とする。
3.  $t = T$  ならフェーズを切り替え、 $t \leftarrow 0$  として手順1に戻る。  $t \neq T$  なら現在のフェーズを維持し、手順1に戻る。

## 5-4 評価

### (1) 評価方法

計算機シミュレーションを用いて評価を行う。評価尺度として、検索成功率とクエリ数を用いる。検索成功率は、コンテンツの要求回数に対する成功した回数の割合とし、クエリ数はクエリ転送回数の合計とする。

また、比較手法としてフラッディング方式、ランダムフラッディング方式を用いる。

### (2) シミュレーション条件

ネットワークトポロジは、ピア数 300 の BA モデルトポロジとし、コンテンツは 10 種類とする。各ピアは初期状態として無作為に 2 種類のコンテンツを保持する。コンテンツ要求は  $\lambda_{req} = 5.0$  のポアソン分布に従う回数、コンテンツの追加・削除は  $\lambda_{mov} = 0.1$  のポアソン分布に従う回数がそれぞれ単位時間毎に発生する。また、ネットワーク上に存在するエージェント数を 5 とし、閾値  $T = 10$  でフェーズを切り替える。

### (3) 結果及び考察

シミュレーション結果を以下の図 14 に示す。(a)は各手法における検索成功率、(b)はクエリ数の合計である。提案手法はフラッディング方式とほぼ同等の検索成功率を維持しつつ、クエリ数を抑制している。提案手法では、エージェントがネットワークを巡回し、トポロジとコンテンツの把握を行うため、クエリ数を抑制できていると考えられる。

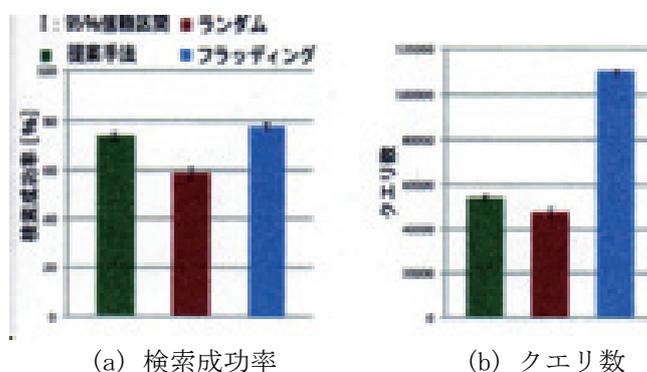


図 14 各方式の検索成功率とクエリ数

## 5-5 むすび

この節では、非構造化オーバーレイネットワークにおける Peer-to-Peer コンテンツ共有を目的として、複数の動的エージェントがネットワーク内を巡回することで、動的に変化するネットワークトポロジとコンテンツの配置状況を継続的に把握し、ユーザの要求に応じて必要なコンテンツの位置情報を提供する機能を実現する手法を提案した。また、现阶段では十分ではないが、基本的な条件を設定して計算機シミュレーションを行い、提案方式がクエリ数を抑制しながらユーザの要求するコンテンツを高い確率で取得できる機能を実現させるための基本的な性能を有することを確認した。

## 6 不確かな多数の情報源から信頼できる情報を抽出する手法の検討

### 6-1 まえがき

前述のように、ネットワーク上の動的エージェントは、多数のピアからネットワークトポロジ、および所有コンテンツに関する情報を収集する。それらの情報は、同一のコンテンツやピアの接続情報であっても、複数のエージェントがそれぞれ別のタイミングで、異なる文脈による情報収集を経て集積されることになる。更新の前後で収集された場合には両者間に矛盾が発生することが考えられ、何らかの別の雑音が乗ることにより、本来得られるべき情報とは別の内容を取得することもあり得る。このような状況では、エージェントがユーザに提供する情報の中から、できるだけ誤情報を排除することが望ましい[7]。

この節では、多数の情報源（ピア）から収集した大量の情報を集積し、同一のコンテンツについて矛盾した情報がある場合に、それをできるかぎり排除する手法について検討する。

### 6-2 仮定する環境

この節では以下の仮定を置く。

- 多数の情報源（ピア）がネットワーク上に存在し、それらがエージェントに対して多様な情報を公開している。
- エージェントはそれらの情報を自由に収集することができる。
- 各情報源は、それぞれ独立に、常に提供する情報内容を書き換え、消去する可能性がある。

このような環境で、エージェントが収集した多種かつ大量の情報を整理し、同一の情報について矛盾する複数の内容が得られた場合に、正しい内容は一つであると仮定して、それ以外の情報を排除するための効率的な手法を確立する。

この状況は図 15 のように表現することができる。この場合、図中の Searcher がエージェントの収集した情報を整理する主体に相当する。

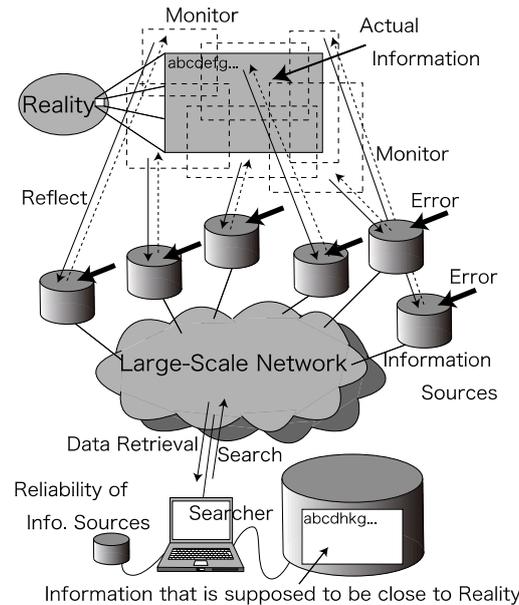


図 15 仮定される環境

### 6-3 提案手法

この節で提案している手法では、ネットワーク上に提供されている情報の内容（コンテンツ）は、コンテンツの新しさ、コンテンツを提供する情報源の信頼度、および同一のコンテンツを提供する情報源の数の三つの観点からその信頼性を評価する。

ある同一のコンテンツに関して複数（ $N$  個）の情報源が矛盾する内容を含んだ状態で情報公開をしている場合を考える。そのコンテンツ  $i$  を  $M$  個（ $1 \leq y \leq M$ ）の断片  $i_y$  に分割し、それぞれについて、その内容が  $p$  である確からしさを上記の三つの観点から評価する下記の関数  $F(i_y, p)$  の値を求め、この値を最も大きくする  $p$  を、 $i_y$  の正しい内容であると推定する。

$$F(i_y, p) = \sum_{x=1}^N \left( \frac{A}{T - t_x} \delta_{d(x, i_y), p} \cdot R_x \right)$$

各断片について同様のことを行って、正しいと推測された  $M$  個の断片を集めて並べたものが、最終的に正しいコンテンツであると推定する。手法の詳細については、本報告の最後にある発表資料に挙げた国際会議発表論文 "A Reliable Information Retrieval from Plural Partially Unreliable Databases" を参照されたい。

#### 6-4 評価

上記の方式の有効性について、計算機シミュレーションを用いて評価した結果の一部を下の図 16, 17 に示す。シミュレーション上で擬似的に作成した現実の情報と、それを反映した複数の情報源のコンテンツから現実の情報を推定した結果との一致率に関して、図 16 では、情報源の提供するコンテンツの更新頻度が、図 17 では情報源の提供するコンテンツに対する誤情報の混入率が、それぞれ与える影響をまとめている。

提案方式との比較のために、正しいコンテンツの断片を決定する方法として、情報源の多数決を用いる手法 (Majority Oriented Scheme: MOS)、過去の提供情報の確かさに基づいた情報源の信頼度を用いる手法 (Reliability Oriented Scheme: ROS)、コンテンツの提供された日時に基づき情報の新鮮なものが最も正しいと判断する手法 (Freshness Oriented Scheme: FOS) の三つも併せて評価した。

両図から、情報源のコンテンツの更新レートが高くなる程、あるいは、情報源の提供コンテンツへの誤情報混入率が高くなるほど、一般には誤情報の排除は困難になっている。しかし、提案手法は他の手法と比較して誤情報を適切に排除することが可能であることが明らかになった。

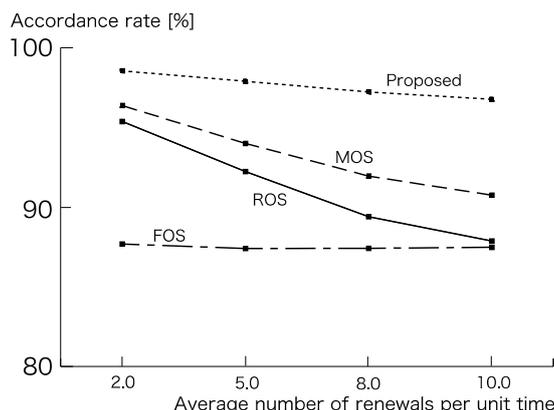


図 16 情報の更新レートと一致率の関係

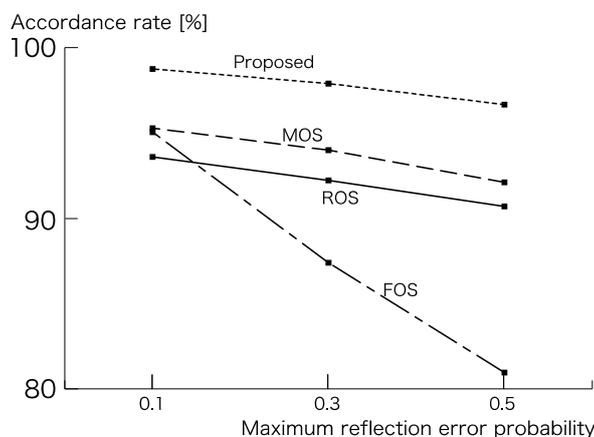


図 17 誤情報混入確率と一致率の関係

#### 6-5 むすび

この節では、複数の情報源から情報を収集する際に、同一内容であるはずの情報、情報源によって矛盾した内容を含む場合に、誤情報を排除し、エージェントに対して適切な情報を与えるための基礎的な研究について報告した。この手法は、本研究におけるエージェントの情報収集に適用できるだけでなく、広く一般的な情報探索にも適用できると考えられ、応用範囲が広い。今後もこの内容に関しては研究を継続してゆく予定である。

#### 7 おわりに

本研究では、ユーザがコンテンツの共有を行う P2P ネットワークにおいて、そのネットワーク内を巡回する複数の動的エージェントを常駐させ、各ピアの接続状況からなるネットワークトポロジ情報と、各ピアが所持するコンテンツの情報を常時収集するシステムについて検討を行った。

具体的には、まず研究の最初の段階として、複数の管理ノードを用いた動的エージェント間の情報共有による情報探索について検討した。提案方式では、通信コスト、時間コストを最も抑制することができ、特にターゲットの存在確率が小さいほど、また、エージェント数が多いほど有効であることが示された。しかし一方で、処理コストについてはまだ改善の余地があることが判明した。

次に、前節で提案した方式を改善し、複数の動的エージェントを用いたネットワーク内情報探索において、全体のネットワークトポロジが未知である環境にあっても、エージェント間の探索状況の共有を効率的に行う方式を求めた。この節で用いた条件下では、提案方式は、比較しているどの方式よりも通信コスト、時間コストを抑制できており、処理コストについてもロック方式やスタンプ方式と同程度に抑制できている。また、前節の提案方式に対してすべての条件において優越しており、改善の効果が見られた。

第三に、非構造化オーバーレイネットワークにおける Peer-to-Peer コンテンツ共有を目的として、複数の動的エージェントがネットワーク内を巡回することで、動的に変化するネットワークトポロジとコンテンツの配置状況を継続的に把握し、ユーザの要求に応じて必要なコンテンツの位置情報を提供する機能を実現する手法を提案した。現段階では十分ではないが、基本的な条件を設定して計算機シミュレーションを行い、提案方式がクエリ数を抑制しながらユーザの要求するコンテンツを高い確率で取得できる機能を実現させるための基本的な性能を有することを確認した。

さらに、エージェントが複数の情報源から情報を収集する際の、誤情報の効率的な排除方法を求めるために、不確かな多数の情報源から信頼できる情報を抽出する手法を検討したことについても述べた。この手法は、本研究におけるエージェントの情報収集に適用できるだけでなく、広く一般的な情報探索にも応用できると考えられる。

一年間の研究により、当初目的としていた研究内容はほぼ完了することができた。しかし、本研究を行った結果、さらに踏み込んだ内容として、より自律性の高い P2P システムを実現するためには、中央サーバに依存しないシステムへの拡張を行うことで、スケーラビリティとフォールトトレランスを実現することが重要であると考えに至った。これ以降は上記のような方向性でさらに研究を発展させてゆく予定である。

## 【参考文献】

- [1] 菅原真司, 三木哲也, “複数の動的エージェントによるネットワーク内情報探索に関する検討,” 信学技報, IN2000-194, pp. 17-23, Mar. 2001.
- [2] 鎌田英朗, 木下和彦, 山井成良, 戸出英樹, 村上孝三, “時間制約下での効率的エージェント制御方式,” 信学技報, IN2004-56, pp. 65-68, Sep. 2004.
- [3] 堀口晃佑, 川越恭二, “マルチモバイルエージェントによる情報探索行動における移動制御方式,” 情処学研報, DSM 28-2, pp. 7-12, Dec. 2002.
- [4] 大内東, 山本雅人, 川村秀憲, “マルチエージェントシステムの基礎と応用,” コロナ社, Apr. 2002.
- [5] 井家敦, 石坂充弘, “情報収集型モバイル・エージェントの効果的な巡回方法,” 信学技報, NS2006-219, pp. 317-321, Mar. 2007.
- [6] 平山勝敏, 北村康彦, “情報収集のための分散タスク割り当て,” 信学技報, ICS132-12, pp. 63-68, Mar. 2003.
- [7] S. Sugawara, N. Sonehara, “An Efficient Information Retrieval from Plural Independent Databases Partially Unreliable,” Proc. IASTED Euro IMSA 2007, PP. 148-153, Mar. 2007.

〈発表資料〉

題名	掲載誌・学会名等	発表年月
管理ノードを用いたモバイルエージェント間の情報共有による効率的ネットワーク内情報探索手法の検討	電子情報通信学会技術研究報告 (インターネットアーキテクチャ研究会)	2010年8月6日
マルチモバイルエージェントを用いた役割分担による効率的なPeer-to-Peerコンテンツ共有ネットワーク管理手法	電子情報通信学会技術研究報告 (ネットワークシステム研究会)	2010年10月21日
複数のモバイルエージェントを用いたネットワーク内情報探索における単一管理ノードによる探索状況の共有	電子情報通信学会技術研究報告 (インターネットアーキテクチャ研究会)	2011年2月18日
マルチモバイルエージェントを用いたPeer-to-Peerネットワークにおける効率的なコンテンツ共有	電子情報通信学会 2011年総合大会講演論文集	2011年2月28日発行 (講演は東日本大震災の影響により中止)
A Reliable Information Retrieval from Plural Partially Unreliable Databases	Proceedings of 2011 International Conference on Data Engineering and Internet Technology	2011年3月15日