様式 C-19

# 科学研究費補助金研究成果報告書

平成 23 年　4 月　1 日現在

機関番号：33603
研究種目：研究活動スタート支援
研究期間：2009-2010
課題番号：21800063
研究課題名（和文）　クラウドコンピューティングを構成する分散協調ストレージ技術の研究

研究課題名（英文）　Research onthe Distributed Storage Technology Composed by the Cloud Computing Technologies

研究代表者
土屋　健（TSUCHIYA TAKESHI）
諏訪東京理科大学　経営情報学部　助教
研究者番号：90546251

研究成果の概要（和文）：
本研究ではクラウドコンピューティングにおける根幹をなす技術として，この分散するノード間で分散協調によるストレージ技術の研究開発をおこなった．この分散ストレージはクラウドコンピューティングを構成する分散ノード間で広大なストレージ論理空間を構築し，サービスの規模や，分散規模に適応した柔軟なサービススケーラビリティを持つ分散協調ストレージ技術を提案し，その可能性，問題点について明らかにした．また，ソフトウェアとして試作し，多くのサービスとの融合の可能性についても実装レベルにおける評価を行った．

研究成果の概要（英文）：
In this research proposes and carries out distributed storage service composed by the basis of cloud computing. Firstly, to overcome distributed environment, the object management manner, which enables to scale up on the demand of services. It is named Distributed Inverse Tree, and used in this proposal storage service as basis, and clarify the possibility, availability and issues by the experiments. Finally, the implemented software are evaluated the potential as real services, and discuss as the basis of cloud computing technology.

交付決定額

（金額単位：円）

| | 直接経費 | 間接経費 | 合　計 |
|---|---|---|---|
| 21 年度 | 1,060,000 | 318,000 | 1,378,000 |
| 22 年度 | 970,000 | 291,000 | 1,261,000 |
| 年度 | | | |
| 年度 | | | |
| 年度 | | | |
| 総　計 | 2,030,000 | 609,000 | 2,639,000 |

研究分野：分散協調ストレージ，情報通信ネットワーク
科研費の分科・細目：インターネット高度化．ネットワーク，情報通信工学
キーワード：クラウドコンピューティング，分散ストレージ，スケールアウト

## １．研究開始当初の背景

Peer-to-Peer（以下、P2P）技術の普及によって、広域に分散したノード間でオーバレイネットワークを形成し、ノード数の増加に対するスケーラビリティ、そしてノードの一部故障によっても全体に影響を与えないロバストネスなネットワークを構築することが可能になるため、次世代のネットワーキング技術として注目されている。一方でネットワーク上では情報が飽和し、ユーザの端末に存在するストレージにおいて過密・過疎といった問題が存在し、資源の有効活用ができていない。多くのベンダーが提案するNAS(Network Attached Storage)になどのストレージシステムが多くのベンダーより提供されている。これらは、ストレージという

観点において、想定するアプリケーション、ネットワーク形態、そして規模の点において汎用性が低いという課題がある。
また，大規模なストレージシステムとして、OceanStore、CFS、PAST など Distributed Hash Table（以降、DHT と呼ぶ）技術を用いてネットワークのスケーラビリティを実現しているが、あくまでモデルの提案であり、実システムとして動作するものは存在していない。また、分散ストレージシステムとして提案されている技術であるが、次世代のサーバインフラとして利用が想定されるクラウド環境を想定しておらず、サーバへの要求へ適応可能なスケールアウト技術は検討されていない。従って、次世代サーバ環境のインフラとして活用可能な技術の検討が必要となる。

２．研究の目的

次世代のサーバ環境として想定されるクラウドコンピューティング環境上において，ダイナミックに Web インフラを構成すると同時に，その分散コンピューティング環境にスケールアウト可能な分散協調ストレージサービスの研究開発を目的としている。つまり，利用するユーザ数や、アプリケーション規模に適応してアプリケーションが利用する資源を動的に増加・解放することで、提供するサービス規模に適応する技術を指す。次世代の web 系サービスを提供するサーバ環境では、高機能・高価であるサーバよりも、安価である PC をクラスタリングし、そのスケールによって性能・資源を確保することが見通すことができ、本研究でもこの指向を持つ基本技術を目指すこととなる．このような次世代サーバ環境を構成するサーバインフラの基幹技術として利用を想定した分散協調ストレージ技術を明確にする。本研究で提案する分散協調ストレージ技術は、P2P 技術をベースとしたスケールアウト可能な分散ストレージ技術、具体的にはストレージを構成する分散協調ストレージのノードが状況に適応して増加・減少、そして変動へサービスを停止することなく対応可能とする分散アルゴリズムと、プロトコルを明確にし、これら機能を提供するソフトウェアプラットフォームの研究開発を行う。

３．研究の方法

以下に示す課題を解決することで，本研究の目的を果たすと考え，研究期間を 3 つのフェーズに分けて研究を推進した．

1. コンテンツの分散管理手法の明確化: 様々な種類のノードが構成するオーバレイネットワークが構築する分散協調ストレージ上においてデータを分散配置するための手法を明確化．このとき、各ノードにおいて同一のデータ分散アルゴリズムを採用することで、各ノードから一意に目的のデータへ到達することを可能とする。

2. ノード間でのコンテンツの負荷分散・障害への対応: 特定のノードへのオブジェクトの集中による予期しない集中することなく、各ノードの提供可能なディスクサイズ、ノードのパフォーマンスに適応してコンテンツを配置する。

3. ファイルおよびディレクトリの追加、更新、削除などのオペレーションと分散管理される情報の検索のためのプロトコルの検討: 既存の P2P ファイル共有と異なり、コンテンツの明確な管理が可能であることを意味する。つまり、更新操作による一貫性の保証、削除操作によるノードに分散するデータの削除、そして各ノードにおいて不要パケット・通信データを高効率でストレージサービスを提供するためのオペレーションプロトコルの検討を行う。また、分散管理されている情報に対して、検索を可能とする情報を検討することで、実用性の高いストレージサービスの提供が可能となる。各ノードは追加や取得などオペレーションされる情報をアプリケーションレベルで管理を行い、複数のユーザからの利用に対しても、ストレージの一貫性を保持する手法を明確化する必要がある。

4. ソフトウェアとして実装: 上記に示した各要素技術をプラットフォームとして利用可能とするため、API (Application Programmable Interface)，ソフトウェアとして利用ができるように実装を行い，本研究の実用化を目指す

４．研究成果

ストレージに投入されるオブジェクトの分散管理手法として，分散区分木アルゴリズムを提案した．そして，この分散アルゴリズムにおける特定のノードへの負荷、ノード消失などの障害への対応と言った各項目について明らかにしている。
図 1 に示すように、ネットワークに参加するノードが管理ノードとエッジノードに自律的に分類され，管理ノードに接続するエッジノードの数が閾値を超えた場合に新たな管理ノード（サブ管理ノード）を新規に選定する手法を明らかにすることにより，ネットワークを自律的に拡張可能とする技術を示している．これにより 管理ノードに接続するエッジノード数 (N) を分岐数とする N 分岐の木構造ネットワークを構成し，通信コストが O(logN)になる低コストの通信を実

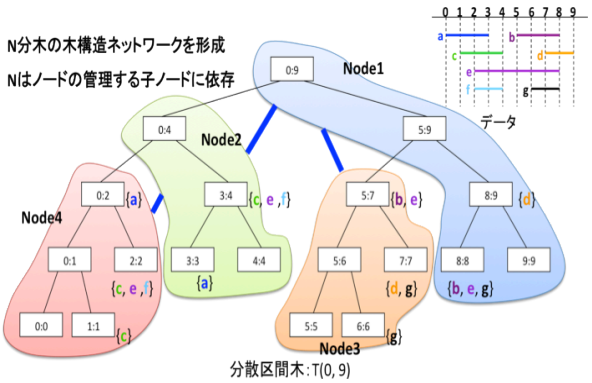


図１　分散区間木アルゴリズム

現している．この木構造ではストレージに格納するオブジェクトデータをすべてチャンク単位に分割し、ハッシュ化された木構造へ配置される。このチャンクのアドレスに基づくハッシュ値を分散に利用することで，この区間木アルゴリズムの特徴として、格納されたチャンクは右上の節点に複製されたデータチャンクが配置される．以上のことから、分散協調ストレージを構成するノードへ木構造の分割手法としては複製をもつ節点は別ノードとなるように分散管理する(一例として、図１では塗り分けられた色ごとにノードが分散して管理)ことで、ノードの消失による障害時にも，複製されたチャンクを利用することで障害へ対応できる。また、前述したようにアルゴリズムの特徴から右上の節点に複製が存在することから、木構造ネットワークのボトルネックであるルートノードの単一障害に対しても，各ノードが兄弟ノードを代替経路として認識することにより，ルートノードの障害や管理ノードの親子いずれかの障害に対しても代替経路を利用してネットワークを維持できることを明らかにしている．

クラウドコンピューティング環境における実現性を評価するため，提案手法をソフトウェアで実装し，シミュレーションによりノード障害時などの異常時でもサービスを問題なく継続し提供できることを示している．また，ヘテロジニアスなネットワーク環境に対応するためにノードを処理能力などに応じて自律的に２種類に分類する手法を提案し，これにより安定したオーバレイネットワークを低い通信コストで実現できることを明らかにしている．

サーバだけでなく，ユーザ環境なども含まれることになるクラウドコンピューティング環境では不安定な特性を持つオーバレイネットワーク上で安定した仮想分散ストレージを上位のアプリケーションに提供するための新たなレプリケーション手法を提案した．

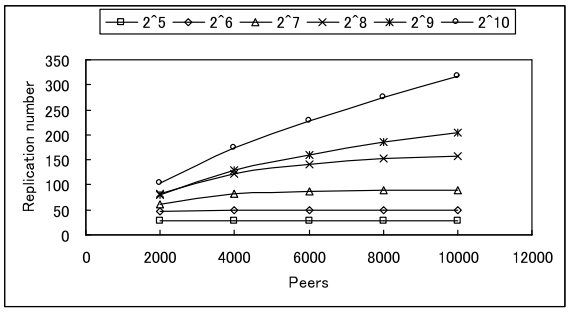


図２　ノード数とオブジェクト複製数

具体的には，まず，ミラーグループ手法によってオーバレイネットワーク上の管理ノードを 160 ビットのハッシュ空間にマッピングし，複数の管理ノードで同一のハッシュ値の範囲を管理する方式を述べている．オブジェクトから導かれたハッシュ値によって該当する管理ノードにそのハッシュ値を示すオブジェクトを転送することにより同一オブジェクトの複製を複数の管理ノードが管理することを可能にしていた．このミラーグループ手法の評価のため，提案手法を実装したシミュレーションを行い，オブジェクト分散管理システムの高いアベイラビリティを保証できることを明らかにした．

一方，ミラーグループ手法では，グループサイズが固定になってしまうため，ネットワークの規模が拡大した場合には特定のオブジェクトを利用する頻度がより多くなるとオブジェクト取得までのレイテンシが劣化することを明らかにしている．そこで，この問題に対処するために，区間木手法を拡張した新たな分散区間木手法によるオブジェクトの管理技術を提案した．分散区間木手法は，線分データを保存する区間木を応用してオーバレイネットワーク上の管理ノードが区間木の節点を管理し，オブジェクトに割り当てられた線分値から該当する管理ノードにオブジェクトを転送する手法である．分散区間木を実装したシミュレーションでは，数万ノードの大規模オーバレイネットワーク上のオブジェクトの分散性を実現している．これによりオブジェクトの複製数が O(logN)となり，ネットワークの規模に応じて動的に複製数が制御可能であることを述べている．また，従来の複製手法である，ネットワークの通信経路上に複製を配置する手法やランダムに複製を配置する手法と提案手法のヒット率を比較した結果，経路上に配置した場合と同程度のヒット率であることを述べている．一方で，経路上に配置する場合は通信の度に複製が保存されるため，複製数が膨大になりストレージ資源の利用率の低下となるが，分散区間木手法の場合には，ネットワークの規模に対して O(logN)の複製数となるためストレ

ージ資源の有効利用が可能になることを明らかにしている．ミラーグループ手法および分散区間木手法では，大規模なネットワークにおいても管理ノードとオブジェクトが同一のアドレス空間に存在するためオブジェクトの所在が明らかになり，オブジェクトの更新や削除ができる点も評価できる．

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）

〔雑誌論文〕（全て査読付き）（計 10 件）
吉永浩和，**土屋健，** 小柳恵一，“分散協調ストレージシステムにおけるネットワーク規模および一貫性を考慮した複製配置手法の提案”，日本知能情報ファジィ学会学会誌”知能と情報” ，Vol.22(2)， pp.246-256, 2010

2. T.wei, C.Madrazo, **T.Tsuchiya**, and K.Koyanagi, “Unbalanced Replication by Evaluating Popularity over Web Scalability Strage,” Proc’ of IEEE Inte’l Conf. on Computer Science and Automation Engineering, Shanghai China, June, 2011 (to be published.)

3. **T.TSUCHIYA**, H.YOSHINAGA, and K.KOYANAGI, “Streaming Cloud Service Concept by Peer-to-Peer Distributed Technology”, Proc. Of Computation World 2010, pp. 196 -201, Lisbon, Portugal Nov, 2010

4. G.JING, **T.TSUCHIYA**, and K.KOYANAGI et. al：” Tree Based Structured Management of Delivery Path for the Multi-Path Streaming”, Proc. of cicsyn 2010, pp.334 - 339, Liverpool, UK, 2010 （第 2 著者）

5. **T.TSUCHIYA**, H.SAWANO, H.YOSHINAGA, and K.KOYANAGI, “Streaming Management Platform for Distributed Camera Systems” Proc. of Int’l Conf. of Computation World 2009, pp.392-397, Athens/Glyfada, Greece、

6. TSUCHIYA.T, Yoshinaga.H, Lihan.M, Koyanagi.K, “Localization algorithms for distributed platform among vehicles”, Proc. of Int’l Conf. Ultra Modern Telecommunications, 2009, 12-14 Oct. 2009 pp. 1 - 6, Saint-Petersburg, Russia

7. **T.TSUCHIYA**, H.YOSHINAGA, K.KOYANAGI,”Distributed Multimedia Information Retrieval”, Proc. of IEEE Int’l Conf. on Semantic Computing, 2009, pp.642-647, Barkley, CA, Sep.2009

8. H.YOSHINAGA, **T.TSUCHIYA** and K.KOYANAGI, “A Study on Scalable Object Replication Method for the Distributed Cooperative Storage System”, Proc. of 4th Int’l Conf. on Digital Telecommunications, Pp.96-101, Jul. 2009 Colmar, France

9. H.Yoshinaga, **T.Tsuchiya**, K.Koyanagi, “Scalable and Persistent Multimedia Data Management System using The Distributed Interval Trees”, Proc. Of Int’l Conf. EuroIMSA 2009 pp. 86-92, 2009, Cambridge

10. G.JING, **T.TSUCHIYA**, and K.KOYANAGI, et. At al. ” Study on Multi-Path Transmission Model based on P2P Overlay Networks for Streaming Data”, Proc. of IEEE 9th Int’l Sym’ on Applications and the Internet, 2009.Jul Pp.271 -274, Seattle , （第 2 著者）

〔学会発表〕（計 2 件）
**1.土屋 健**, 吉永 浩和, 小柳 恵一, et at al.″クラウドコンピューティング環境における既存サービスを活用したDBサービスの高度化に関する研究″ IEICE-NS2010-204 ，信学技報Vol. 110 ，No.448, pp.235-240 ，2011年3月 （第1著者）

**2.土屋 健**, 吉永 浩和, 小柳 恵一 ″ユビキタスサービス環境を対象とした分散検索におけるユーザプライバシ動的導出手法に関する

研究″ IEICE-NS2009-46 , 信学技報Vol. 109, No.129, pp.19-24 ., 2009年7月

６．研究組織

(1)研究代表者

土屋　健（TSUCHIYA TAKESHI）
諏訪東京理科大学経営情報学部　助教
研究者番号：90546251