

令和 5 年 6 月 14 日現在

機関番号：12102

研究種目：基盤研究(B) (一般)

研究期間：2020～2022

課題番号：20H04163

研究課題名(和文) 生成的ロギングに基づいたクラウドシステムの設計・実装手法の確立

研究課題名(英文) Establishment of Cloud System Design and Implementation Method Based on Generative Logging

研究代表者

加藤 和彦 (Kato, Kazuhiko)

筑波大学・システム情報系・副学長

研究者番号：90224493

交付決定額(研究期間全体)：(直接経費) 13,800,000円

研究成果の概要(和文)：本研究は、ネットワーク科学とシステムソフトウェアの2つの研究分野で得られた知見を融合し、「生成的ロギング」と呼ばれる大規模ユーザの挙動を人工的に模倣したデータの生成手法の開発を目的に遂行された。これにより、過去に開発経験を有することができないような巨大なシステムを構築する際に、実際のサービスのログが無くてもシステムの要求分析や設計を可能にすることを目指した。本研究の成果としては、統計的手法とエージェントベースの手法による生成的ロギングのモデルを構築した。さらに、このモデルから生成された人工的なログを入力とするシステムのシミュレータを開発した。

研究成果の学術的意義や社会的意義

本研究の学術的意義としては、これまで独立に研究されてきたネットワーク科学とシステムソフトウェアの2つの分野を融合する点にある。ネットワーク科学は、これまで主にSNSのような個人のミクロなレベルの関係からなるネットワークからマクロなパターンがどのように形成されるのかについて主眼が置かれていたが、本研究は、こうしたネットワークのダイナミズムがプラットフォームとなるシステムに与える影響を分析するという新しい試みである。一方、社会的意義としては、クラウド基盤システムの設計開発に対し新しい手法を提案するとともに、アーキテクチャやアルゴリズムなどの研究を行う上でも利便性の高いツールを提供するものである。

研究成果の概要(英文)：This research was conducted to develop a method of generating data that artificially mimics the behavior of large-scale users, called "generative logging," by combining knowledge obtained in the two research fields of network science and system software. The aim was to enable system requirements analysis and design without actual service logs when constructing huge systems that would not have been possible with previous development experience. As a result of this research, we constructed a model of generative logging using statistical and agent-based methods. Furthermore, we developed a simulator for the system that uses the artificial logs generated from this model as input.

研究分野：システムソフトウェア

キーワード：生成的ロギング クラウドコンピューティング ソーシャルネットワーク アクセスパターン

様式 C-19、F-19-1、Z-19 (共通)

1. 研究開始当初の背景

今日、インターネットを介してサービスを提供するネットワークサービスのシステム開発は、クラウドコンピューティング技術を活用して構築することが多い。こうしたシステムの設計にあたっては、高性能、高信頼、省エネルギーといったさまざまな要求を満たす必要があり、そのためにはクライアントからシステムへのサービスアクセスのパターンを正確に把握し、要求分析を通じて、想定される要求に柔軟に対応できるアーキテクチャとアルゴリズムを選択・設計する必要がある。

クライアントからのサービス要求を知るための最も確実な方法は、類似のサービスを提供するシステムの実際のログ（リアルログ）を利用することである。しかしながらリアルログを使用することは、個人情報保護やセキュリティ・プライバシーの問題から困難を伴う。また、一般にリアルログはビッグデータであり、データ量が巨大であるために、その取り扱いが難しいという場合も多い。

前述のような困難に対処するために、従来用いられてきた有用な手法は、スクレイピングと呼ばれる手法である。人間ユーザのクライアントとしての挙動をコンピュータプログラムにより模倣し、サーバにアクセスすることによって公開データを継続的に取得し、公開データの変化や差分を分析することによってリアルログの代替となるようなデータを入手することができる。この手法はある程度は有効であるが、他のユーザのアクセス操作の推論に限界があったり、誤差を伴うことによる限界を甘受せねばならない。また、継続的な取得に多大な時間を要するため長期的な計画の下でしか実行できなかつたり、そもそもスクレイピング自体が禁止されていることも多くなってきている。さらに、リアルログもスクレイピングも、実際にユーザの挙動が行われる以前には、所望するデータを取得することができないという本質的な制約を抱えている。

以上の問題から、本研究の開始当初においては、大規模ユーザを対象とするようなシステムの開発は、一部の巨大プラットフォームを除けば十分に確立しているとは言い難い状況であった。多くの情報システムは、大規模ユーザまでスケールする訳ではなく、十分なユーザ数を獲得するまでは大規模システム開発経験を有することができないというジレンマを抱えていた。

2. 研究の目的

近年、Watts-Strogatz や Barabási-Albert らの先駆的な研究によって端緒が開かれたネットワーク科学や複雑ネットワークと呼ばれる研究分野が進展し、SNS を代表とするさまざまなネットワークにおいて頻繁に観察される現象が数理科学的手法により分析されるようになってきた。本研究は、こうしたネットワーク科学の分野で得られている知見と、近年のシステムソフトウェアの研究開発で得られている知見とを融合し、数十万から数百万規模の大規模ユーザに対応し得るシステムの開発技術の確立を目的に遂行された。

ここでは特に、ネットワーク科学の有力な研究成果の一つであるスケールフリーネットワークとその変種に注目し、大規模ユーザ間の関係や、データに付与されるタグ情報をモデル化することを目指した。さらに、適切なパラメータ設定およびパラメータ変動によってユーザの挙動を模倣し、それをリアルログの代替としながら、ネットワークサービスを実現するシステムの開発（設計・実装・テストを含む）の各フェーズで有効利用できることを実証することを目指して研究を行った。ネットワーク科学で得られている知見に基づくアルゴリズムに基づいてログを生成することを、ここでは「生成的ロギング（generative logging）」と呼ぶ。本研究の最大の特色は、リアルログでなく、スクレイピングによって得られる擬似的なログでもなく、ネットワーク科学に基づくアルゴリズムによって生成されたログを活用することによって、実際の有用なシステム開発を行う新たな方法論の開拓を最終的な目標とした。

本研究は、これまで独立に研究されてきたネットワーク科学とシステムソフトウェアの 2 つの分野を融合する新しい試みである点に独自性があると言える。クラウドコンピューティングの設計開発に対し新しい手法を提供するものであり、実際のシステム構築のみならず、アーキテクチャやアルゴリズムなどの研究を行う上でも利便性の高いツールを提供するものである。また一方で、ネットワーク科学の分野に対しては、SNS を提供するサービス基盤という新しい観点を与え、工学的応用という側面からも意義を有すると言える。

3. 研究の方法

本研究は、生成的ロギングの理論的基礎となるモデル構築と、それをもとにしたシステムのシミュレータの開発の 2 つを目指して遂行された。その具体的な方法は以下のとおりである。

生成的ロギングは、ソーシャルネットワーク上で生成されるクライアントからのサービス要求の一般化されたモデルを与えるものである。このモデルでは、ユーザ数やユーザ間のフォロワー・フォロワー関係などのさまざまな属性がパラメータ化されており、それらの値を設定することで現実に近いサービス要求の具体的なパターンを生成することができる。また一方で、クラウド基盤のシミュレータを作成した。これは、開発の対象とするシステムのアーキテクチャやアルゴリズムがパラメータの設定やモジュールの選択により容易に変更できるようになったもので

あり、クライアントからのサービス要求を受け取ると、応答時間やシステムの電力消費量を計測することができるものである。以上で得られた2つの成果により、実際のサービスのログがなくても手軽にクラウド基盤の要求分析や設計を行えるようにすることを目指して本研究を遂行した。さらに、上記で得られた設計をもとに、実機を用いてクラウド基盤のプロトタイプを実装しながら、開発時のテストやシステム構築後のチューニングで本提案手法の有用性の実証を目指した。

4. 研究成果

生成的ロギングのモデルの構築のために、本研究では数理的手法をもとにしたトップダウンによるアプローチと、エージェントシミュレーションの手法をもとにしたボトムアップによるアプローチの2つを採用した。

まず、数理的手法をもとにしたモデルの基本的なアイデアは、イベント発生の時系列に着目し、その分布を模倣した時系列を生成するというものである。ユーザやシステムに関する深い知識を要しないため、幅広い対象に適用が可能であるという利点がある。詳細について説明するために、ここでは以下のようなSNSサービスを仮定して議論を進める。まずSNSのユーザ間にはフォロー・フォロワー関係があり、あるユーザがファイル（写真や記事）をポストすると、単位時間ごとにフォロワーのうちの n 人が確率 p でそのファイルにアクセスをする。この設定のもとで d 単位時間内に k 回のアクセスがあるとすると、このファイルに対するアクセス頻度は、これらの変数をパラメータとする方程式によってモデル化することができ、パラメータの設定によってさまざまなシステムログ（この場合では、ファイルシステムに対するユーザからのアクセスログ）を人工的に作り出すことが可能となる。本研究では、このような手法によって生成されたログを、研究代表者らが先行研究においてスクレイピングで取得した Flickr（写真共有サイトの一つ）のアクセスログと比較し、実際のシステムのアクセスパターンを模倣していることを確認した。

一方、エージェントシミュレーションの手法をもとにしたモデルの基本的なアイデアは、想定する SNS の背景にあるユーザ間の関係をモデル化し、それに基づいて時系列の模倣を行うものである。SNS のユーザ間ネットワーク（フォロー・フォロワー関係）にはバラバシアルバートモデルを、またタグの発生にはユールサイモンモデルを適用することでモデル化を行った。ユーザは、フォローしているユーザとタグの両方の投稿を参照する。ユーザは、優先的選択モデルに基づいてフォローフォロワー関係を構築し、さらに、新規投稿時にはユールサイモンモデルに基づいて既存タグもしくは新規生成タグを投稿に付与する。投稿を閲覧する際には、フォローしているユーザによる投稿と、自分の購読しているタグの付与された投稿がタイムラインに表示されるとする。以上で述べた手法によって生成されたログを分析した結果、上記の数理的手法によるアクセスパターンと類似のログが得られることが確認された。

本研究では、以上で述べた生成的ロギングのモデルをもとに、システムの性能評価のためのシミュレータを開発した。ここでは特に、複数階層からなるストレージシステムに焦点を当てて研究を行った。複数階層からなるストレージシステムでは、アクセスの発生に応じてオブジェクトの格納場所が変更される。それによりアクセス遅延の低減やスループットの向上が見込まれる。一方で、ファイルのマイグレーション（システム内部での転送）が発生するため、それによる性能低下やエネルギー消費が発生し得る。そのため、各階層の適切な容量配分や、適切な入れ替えアルゴリズムを設計することは非常に重要となる。しかし、高々2層程度であればある程度の理論的な分析は可能であるが、層が増えれば増えるほど分析は困難になるため、シミュレーションを用いた評価が重要になる。

本研究では、多種かつ多数のストレージデバイスに非常に多数のオブジェクトが分散格納されているストレージシステムを効率的にシミュレーション評価することを目的として、複数のCPUコアによる並列シミュレーションを可能とする枠組みを実現した。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Fumio Machida, Koji Hasebe, Hirotake Abe, Kazuhiko Kato	4. 巻 accepted
2. 論文標題 Analysis of optimal file placement for energy-efficient file-sharing cloud storage system	5. 発行年 2020年
3. 雑誌名 IEEE Transactions on Sustainable Computing	6. 最初と最後の頁 12 pages
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TSUSC.2020.3037260	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計1件（うち招待講演 0件/うち国際学会 1件）

1. 発表者名 Horleang Choeng, Koji Hasebe, Hirotake Abe, and Kazuhiko Kato
2. 発表標題 ulti-Tier Power-Saving Method in Cloud Storage Systems for Content Sharing Services
3. 学会等名 17th International Conference on the Economics of Grids, Clouds, Systems and Services (国際学会)
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	阿部 洋丈 (Abe Hirotake) (00456716)	筑波大学・システム情報系・准教授 (12102)	
研究分担者	岡 瑞起 (Oka Mizuki) (10512105)	筑波大学・システム情報系・准教授 (12102)	
研究分担者	町田 文雄 (Machida Fumio) (50842209)	筑波大学・システム情報系・准教授 (12102)	

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	長谷部 浩二 (Hasebe Koji) (80470045)	筑波大学・システム情報系・准教授 (12102)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関