様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成２４年　４月１５日現在

機関番号：１２６０８
研究種目：若手研究（A）
研究期間：2010～2011
課題番号：２２６８０００５
研究課題名（和文）　数百万ノードからなる自律分散システムの実験環境構成法

研究課題名（英文）　Simulating distributed systems composed of millions of nodes

研究代表者
首藤　一幸（SHUDO KAZUYUKI）
東京工業大学・大学院情報理工学研究科・准教授
研究者番号：90308271

研究成果の概要（和文）：今後我々は数百億～兆という規模の分散システムを研究の対象としていかなければならない。これまで、我々研究者が実験可能な分散システムの規模は 10 万～100 万にとどまっていた。本研究ではそれを数百万まで向上させた。成果は、各国の研究者が研究に用いているオープンソースソフトウェアの一部として公開・配布されている。また、汎用分散処理システムの上でシミュレーションを行うという新しいアプローチでの研究を開始した。

研究成果の概要（英文）：A research community in distributed systems has to deal with the scale from tens of billion to thousands of billion. We have improved the scale that we can simulate from sub-million to millions in this research. Our achievements include the new approach to large-scale simulation in which a simulator is based on a general purpose distributed processing systems such as graph processing systems.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
２０１０年度	5,200,000	1,560,000	6,760,000
２０１１年度	3,200,000	960,000	4,160,000
年度			
年度			
年度			
総　計	8,400,000	2,520,000	10,920,000

研究分野：総合領域
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：大規模ネットワークシミュレーション

１．研究開始当初の背景

インターネット上には100万台を超える規模の分散システムが出現している。例えばBitTorrent mainline DHTを構成するノードの数は、1,000 万に迫っている。また、インターネットに接続される機器は爆発的に増加しており、センサネットワークやInternet of Thingsの規模を想定すると､我々は今後、数百億～兆の規模を研究対象としなければならない。しかし、2009 年時点で手に入る実験手法・手段では、10 万～100 万ノードという規模が限界となっていた。手段がないため、今後現れる規模は言うに及ばず、現実に稼働している規模すら我々は実験できないという危機的な状況を迎えている。

２．研究の目的

そこで、実験可能な規模をまずは数百万に向上させることを狙った。

３．研究の方法

(1) まず、実験プラットフォームに求められ

る性能を見積もった。具体的には、クラウドストレージと呼ばれる分散データストアが生成する通信トラフィックを理論的、実験的に調べた。

(2) 開発してきた大規模実験プラットフォームを元に、1. マシン単体で扱えるノード数を向上させ、2. 複数台での分散シミュレーションを可能にすることで、実験可能ノード数を向上させた。

(3) 分散シミュレータを作りこむ、というこれまでのアプローチとは異なり、汎用の分散処理システムを用いてシミュレーションを行うというアプローチを始めた。

４．研究成果

(1) 実験プラットフォームに求められる性能の見積もり

分散システムのシミュレータがどの程度の通信を取り扱えればよいかを調べる目的で、現実の分散システムが生むトラフィックを調べた。本研究の主なターゲットは peer-to-peer システムであるが、それよりも通信量が多く、シミュレータへの負担が重いことが予想されるクラウドストレージを対象とした。具体的には、SNS 最大手、米 Facebook 社が自社向けに開発した後にオープンソース化した非集中型クラウドストレージ Apache Cassandra を対象とし、これがノード数に応じてどの程度の通信トラフィックを生成するのかを測定した。

測定用に用意したマシンの台数よりもはるかに多いノード数について測定する手法を考案・開発し、測定を行った。PC10 台上でノード数を 300 まで増やして測定したところ、ノード数 N に対してトラフィックは O(N^2) となること、また、ノード数 1,000 の際に 229 Mbps に達するだろうことが判った（図１）。

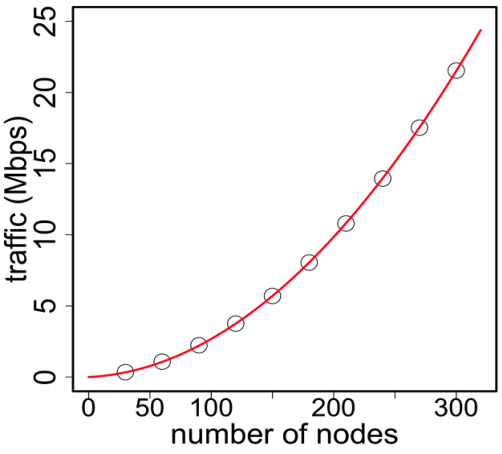

図１ ノード間トラフィック

(2) 実験可能ノード数の向上

開発してきた実験プラットフォームを元にして、その効率を向上させた。それまで実験可能だった規模は 60 万、効率向上後は 100 万ノードの実験が可能となった。2010 年度中には 125 万ノードの実験を達成した。

加えて、複数台のマシンを束ねてシミュレーションを行う機能を用意し、それによって、さらに数倍の規模の実験を可能とした。

(3) 汎用の分散処理システムを用いたシミュレーション

大規模化に向けた(2)のアプローチには限界が見えてきた。コンピュータ数台ならともかく、それ以上の台数を束ねて分散シミュレーションを行うためには、シミュレータの側に高い要件が求められてくる。具体的には、台数を増やすに従ってノード数やシミュレーション性能が向上していくというスケーラビリティ、通信トラブルなどに対応できる耐故障性などである。つまり、シミュレータ自体に、分散システムとして高度な機能、性能が求められてくる、ということである。この方向の研究・開発は依然続けていくとして、別の方向の模索も始めた。

2011 年度からは、それまでとは異なるアプローチに着手した。つまりは、汎用の分散処理システムを用いたシミュレーションである。例えば、その有用性が認知され始めた分散グラフ処理系を用いて分散システムのシミュレーションを行う。このアプローチでは、汎用ソフトウェアに起因する多くの利点を享受できる。つまり、シミュレーションだけでなく様々な目的に用いられることによるソフトウェアとその作り方の成熟、また、耐故障のための機能などである。現在、グラフ処理系などを用いたシミュレーションを試行しており、その課題が明らかになりつつある。

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）

〔学会発表〕（計３４件）

① 宮尾武裕、長尾洋也、首藤一幸、構造化オーバレイにおける経路表の順序関係に基づくネットワーク近接性の考慮手法、インターネットコンファレンス 2011（IC2011）論文集、査読有、福岡市、2011 年 10 月 27 日、pp. 41-48

② 安藤泰弘、長尾洋也、宮尾武裕、首藤一幸、FRT-2-Chord: Multi-Hop と One-Hop のシームレスな移行が可能な DHT アルゴリズム、日本ソフトウェア科学会第 28 回大会 萌芽セッション、那覇市、2011 年 9 月 27 日

③ 島村祥平、長尾洋也、宮尾武裕、首藤一幸、複数データセンター間での DHT の性能評価、日本ソフトウェア科学会第 28 回大会 萌芽セッション、那覇市、2011 年 9 月 27 日

④ 華井雅俊、中村俊介、首藤一幸、グラフデータベースを用いた PageRank 実装の試み：スケーラブルなグラフ処理に向けて、日本ソフトウェア科学会第 28 回大会 萌芽セッション、那覇市、2011 年 9 月 27 日

⑤ Hiroya Nagao、Shuji Suzuki、Kazuyuki Shudo、A 3D Visualization System for Structured Overlays、Demonstration、Proc. 11th Int'l Conf. on Peer-to-Peer Computing (IEEE P2P'11)、京都市、2011 年 9 月 1 日、査読有、pp.168-169

⑥ Hiroya Nagao、Kazuyuki Shudo、Flexible Routing Tables: Designing Routing Algorithms for Overlays Based on a Total Order on a Routing Table Set、Proc. 11th IEEE Int'l Conf. on Peer-to-Peer Computing (IEEE P2P'11)、査読有、京都市、2011 年 8 月 31 日、pp.72-81

⑦ 宮尾武裕、長尾洋也、首藤一幸、構造化オーバレイにおける経路表の順序関係に基づくネットワーク近接性の考慮手法、情報処理学会 研究報告、2011-OS-118(-15)、鹿児島市、2011 年 7 月 28 日

⑧ 長尾洋也、首藤一幸、構造化オーバレイにおけるバーチャルノード融合、情報処理学会 研究報告、2011-DPS-147(-24) / 2011-MBL-58(-24)、岡山市、2011 年 6 月 3 日

⑨ 宮尾武裕、長尾洋也、首藤一幸、構造化オーバレイにおける柔軟な経路表を活用したネットワーク近接性の考慮、情報処理学会 研究報告、2011-DPS-147(-23) / 2011-MBL-58(-23)、岡山市、2011 年 6 月 3 日

⑩ 中村俊介、首藤一幸、読み出し性能と書き込み性能を両立させるクラウドストレージ、先進的計算基盤システムシンポジウム (SACSIS2011) 論文集、査読有、東京都千代田区、2011 年 5 月 26 日、pp.171-180

⑪ 長尾洋也、首藤一幸、柔軟な経路表: 経路表空間上の順序関係を利用したオーバレイネットワークルーティング方式、先進的計算基盤システムシンポジウム (SACSIS2011) 論文集、査読有、東京都千代田区、2011 年 5 月 25 日、pp.117-125

⑫ 宮尾武裕、長尾洋也、首藤一幸、構造化オーバレイにおける柔軟な経路表を活用したネットワーク近接性の考慮、ポスターセッション、先進的計算基盤システムシンポジウム (SACSIS2011) 論文集、査読有、東京都千代田区、2011 年 5 月 25 日、pp.240-241

⑬ 長尾洋也、FRT に基づくルーティングアルゴリズムデザイン、第二十四回 P2P SIP 勉強会、東京都目黒区、2011 年 4 月 23 日

⑭ 中村俊介、首藤一幸、読み出し性能と書き込み性能を両立させるクラウドストレージ、情報処理学会 研究報告、2011-OS-117(-24)、那覇市、2011 年 4 月 14 日

⑮ 奥寺昇平、中村俊介、長尾洋也、首藤一幸、非集中型クラウドストレージのスケーラビリティ評価、情報処理学会 研究報告、2011-OS-117(-22)、那覇市、2011 年 4 月 14 日

⑯ 奥寺昇平、長尾洋也、中村俊介、首藤一幸、非集中型クラウドストレージのスケーラビリティ評価、情報処理学会 講演論文集、東京都目黒区、2011 年 3 月 4 日

⑰ 建部大輔、中村俊介、首藤一幸、クラウドストレージにおける HDD と SSD の特性比較、情報処理学会 第 73 回全国大会 講演論文集、東京都目黒区、2011 年 3 月 4 日

⑱ 宮尾武裕、長尾洋也、首藤一幸、構造化オーバレイにおける柔軟な経路表を活用したネットワーク近接性の考慮、情報処理学会 第 73 回全国大会 講演論文集、東京都目黒区、2011 年 3 月 3 日

⑲ 中村俊介、首藤一幸、読み出し性能と書き込み性能を選択可能なクラウドストレージ、第 3 回データ工学と情報マネジメントに関するフォーラム (deim2011)、伊豆市、2011 年 2 月 27 日

⑳ 中村俊介、首藤一幸、読み出し性能と書き込み性能を選択可能なクラウドストレージ、情報処理学会 研究報告、2011-OS-116(-10)、福岡市、2011 年 1 月 25 日

21 Kazuyuki Shudo、How to Construct A Distributed Data Store、12th Japanese-American Frontiers of Science (JAFoS) Symposium、木更津市、2010 年 12 月 4 日

22 中村俊介、首藤一幸、読み出し性能と書き込み性能を両立させるクラウドストレージ、Work-in-Progress 発表、第 22 回コンピュータシステム・シンポジウム (ComSys 2010) 併設ワークショップ、大阪市、2010 年 12 月 1 日

23 中村俊介、首藤一幸、読み出し性能と書き込み性能を両立させるクラウドストレージ、ポスター・デモセッション、第 22 回コンピュータシステム・シンポジウム

（ComSys 2010)、大阪市、2010 年 11 月 29 日

24 長尾洋也、首藤一幸、グループ間通信を抑制するオーバレイネットワークの構成手法、ポスター・デモセッション、第 22 回コンピュータシステム・シンポジウム（ComSys 2010)、大阪市、2010 年 11 月 29 日

25 長尾洋也、プライベート／パブリックを考慮した分散ハッシュテーブルの構築、Internet Week 2010、東京都千代田区、2010 年 11 月 24 日

26 中村俊介、MyCassandra コトハジメ、NOSQL afternoon in Japan、東京都品川区、2010 年 11 月 1 日

27 長尾洋也、首藤一幸、新しいアプリケーション層ルーティング方式、IIJ Techtalk、東京都千代田区、2010 年 10 月 14 日

28 長尾洋也、柔軟な経路表に基づく Overlay Network の設計と応用、第二十一回 P2P SIP 勉強会、東京都目黒区、2010 年 9 月 19 日

29 長尾洋也、首藤一幸、オーバレイネットワークにおけるグループ間通信抑制手法、情報処理学会 研究報告、2010-OS-115(-11)、金沢市、2010 年 8 月 3 日

30 長尾洋也、首藤一幸、柔軟な経路表によるオーバレイネットワークのルーティング方式、DICOMO2010 シンポジウム、下呂市、2010 年 7 月 9 日

31 首藤一幸、大規模ネットワークテストベッドへの期待、招待講演、DICOMO2010 シンポジウム、下呂市、2010 年 7 月 7 日

32 Team HIBIKI（長尾洋也、鈴木脩司)、hibiki、Interop Tokyo 2010 クラウドコンピューティングコンペティション、千葉市、2010 年 6 月 9 日

33 長尾洋也、柔軟な経路表による Overlay Network の設計、第 4 回 広域センサネットワークとオーバレイネットワークに関するワークショップ、東京都目黒区、2010 年 4 月 28 日

34 長尾洋也、首藤一幸、柔軟な経路表によるオーバレイネットワークの設計、情報処理学会 研究報告、2010-OS-114(-11)、伊東市、2010 年 4 月 22 日

６．研究組織

(1)研究代表者

首藤　一幸（SHUDO KAZUYUKI）

東京工業大学・情報理工学（系）研究科・准教授

研究者番号：90308271

(2)研究分担者

(3)連携研究者