様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２８ 年　 ６ 月　 ６ 日現在

機関番号：１２１０２
研究種目：若手研究(B)
研究期間：2012～2015
課題番号：２４７０００３１

研究課題名（和文）スループット予測に基づいて複製配置を行う広域分散アーカイバルストレージ

研究課題名（英文）Wide-area distributed archival storage with replica allocation mechanism based on throughput prediction

研究代表者
阿部　洋丈（Abe, Hirotake）

筑波大学・システム情報系・准教授

研究者番号：００４５６７１６

交付決定額（研究期間全体）：（直接経費）　　3,300,000円

研究成果の概要（和文）：ネットワークを介してデータ転送を行う場合のスループット予測技術を応用したシステムの実現、および、予測技術そのものの拡張を目指した研究を実施した。具体的には、予測技術に基づいた広域の相互バックアップシステムやグリッドコンピューティングメタスケジューラのプロトタイピング、TCP 輻輳制御における恒常性の分析、および、マルチパスTCP転送への予測技術の応用などに取り組んだ。

研究成果の概要（英文）：We conducted several researches based on our proposing throughput prediction of data transfer on computer network and researches for improving itself. These includes prototyping of distributed mutual backing-up system for disaster tolerance, meta-scheduler for world-wide grid computing, analysis on the homeostasis in TCP congestion control, and adapting the prediction technique to multipath TCP transfer.

研究分野：システムソフトウェア

キーワード：インターネット　スループット予測　TCP　輻輳制御

１．研究開始当初の背景

本研究を開始した当時は、2009 年 4 月 10 日にアメリカ合衆国カリフォルニア州サンノゼで発生した事件をはじめとして、インターネットを支える通信ケーブルが切断されるという事件や事故がたびたび発生していた。今や欠かすことのできない社会的インフラの１つに数えられるようになって久しいインターネットだが、他の社会インフラと異なり、このような障害による影響は現地周辺に留まらずに世界中に波及し得るという特徴がある。そのような事件・事故によってたびたび障害が発生していたのでは、社会の安定的な発展や、新たなイノベーションの創出を阻害する結果にも繋がりかねない。

ネットワーク内の複数箇所に配置された複数のコンポーネントで構成される分散システムにおいて、システムの耐障害性を高めるための手法にレプリケーションがある。レプリケーションでは、同じ内容を持つ複数のコンポーネント、つまり*レプリカ*を用意しておくことによって、コンポーネント自体の故障や、通信回線の故障による通信途絶に備える。

当研究の研究代表者らは、以前、インターネットトポロジの持つスケールフリー性に着目し、あるサイトにレプリカを作成するとどの程度の効用が得られるかを見積もる手法、および、その手法を用いて適切なレプリカの配置を決定するための研究を実施していた（科研費　若手研究(B)、課題番号 22700029）。それにより、どのサイトにも等しいコストでレプリカが配置できると仮定した場合に、どのサイトにレプリカが配置されているべきかを決定するための指針を示すことが可能となった。

しかし、それによる判断が実践的な意味で適切かというと必ずしもそうとは言えなかった。なぜなら、前述の「どのサイトにも等しいコストでレプリカが配置できる」という仮定が成り立たない状況がしばしば存在する為である。

ここでは特に時間的な配置コストについて考える。新たなレプリカをあるサイトに作成する場合、オリジナルのコンポーネントや他のレプリカの持つデータをすべてそのサイトに転送する必要がある。そのとき、転送が完了するまでの時間を左右するのはネットワークの通信性能である。ネットワークの通信性能を示すためのいくつかの指標のうち、大容量のデータを送信する場合に大きく関係しているのは*スループット*である。スループットは、一秒間で送信可能なビット数を表す。

スループットは、そのネットワークを構成している通信機器や通信回線の種類だけでなく、そのネットワークの混雑状況によっても変動するため、送信者と受信者の組み合わせによって異なる値を取りうる。また、たとえ同じ送信者と受信者の組み合わせであっても、時々刻々と変化する。そのため、いまデータを転送した場合にどの程度のスループットになるかを予測することは容易ではない。

２．研究の目的

我々は、効率的なレプリカ配置の実現をはじめ、様々な広域分散システムの性能向上のためのスループット予測手法の拡張および応用に関する研究を実施した。特に、研究代表者らがこれまでに研究を実施してきた、機械学習に基づいたスループット予測手法を核として、それに基づいたシステム設計や、予測手法の拡張などに注力した。

３．研究の方法

(1) スループット予測を用いて効率的なデータ転送を行う分散システムの実現

スループット予測が高い精度で実現できると仮定すると、それに基づいて分散システムの動作を最適化することができる。たとえば、あるデータの転送先サイトの候補が二箇所あった場合、どちらを選択すれば先に転送が完了するかということは、それぞれのサイトとの間の通信スループットを予測することで予想可能である。また、あるサイトにデータを転送することが決定した後でも、転送元から転送先までに複数の経路がある場合は、どの経路をつかうべきかをスループット予測によって決定することも可能となる。

本研究では、広域分散環境における相互バックアップシステム、および、グリッドコンピューティングシステムにおけるメタスケジューラという 2 つのシステムを題材として、スループット予測を組み込んだシステム設計を実践し、その性能評価を実施した。

(2) スループット予測手法の深化と拡張

インターネット上の通信でしばしば用いられている TCP（Transmission Control Protocol）は、非常に多くの利用者が同時にインターネットを破綻させずに維持することに貢献している一方で、通信スループットの挙動の理解や予測を非常に困難なものにしている。また、ネットワーク環境の高度化・複雑化により、複数の経路を同時に使用したマルチパス TCP が普及の兆しを見せてきており、輻輳制御の更なる複雑化、さらにはスループット予測の更なる困難化が予想される。

本研究では、TCP スループットの挙動に関して新たな一解釈を示すこと、および、マルチパス TCP におけるスループット予測手法を実現することを目指した。

４．研究成果

(1) スループット予測を用いて効率的なデータ転送を行う分散システムの実現

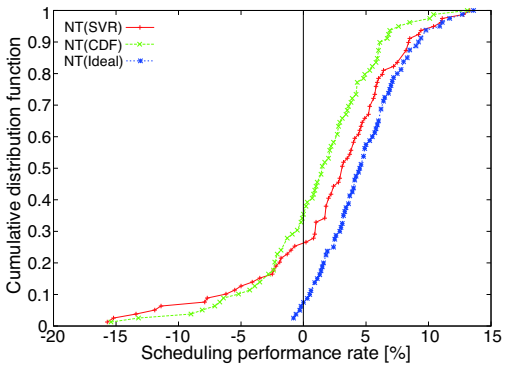
(a) スループット予測に基づいた広域分散バックアップシステム

企業や自治体における情報システムの災害対策を想定し、広域ネットワークで接続されたサイト同士が相互にシステムのバックアップを行うシステムの試作と評価を行った。試作したシステムは、サイト間のネットワークで OpenFlow プロトコルによる経路制御が可能であると仮定し、その上で機械学習に基づくスループット予測を行うことで、バックアップ完了までの時間を短縮させることができる。さらに我々は、OpenFlow の利点を更に活用するために、始点と終点までの経路を複数作成し、その上でマルチパス TCP 転送によるデータ送信を行うシステムのプロトタイピングを実施した。

(b) スループット予測に基づいたグリッドコンピューティング用メタスケジューラ

グリッドコンピューティングでは、あるサイト上で生じた膨大な量の観測データを世界各地の計算サイトへ分割して割り当てることで処理している。このとき、処理時間を短縮するためには、純粋に計算にかかる処理時間を短縮するだけでなく、データをサイト間で転送するためにかかる時間も短縮することが望ましい。

我々は、機械学習に基づいたスループット予測技術を、グリッドコンピューティングにおけるメタスケジューラに適用し、スループット予測の精度向上がどの程度グリッドコンピューティングの性能向上にどの程度寄与するかを評価した。この評価は、世界的な広域分散テストベッドである PlanetLab を利用して収集したスループットの実測値を用いたシミュレーションを通して行われた。既存の素朴なスループット予測手法と、機械学習に基づくスループット予測手法を入れ替えてシミュレーションを実施した結果、より精度の高い機械学習ベースの手法を用いた場合の方が約 11%の処理時間の短縮が達成可能であることが示された（下図）。



(2) スループット予測手法の深化と拡張

(a) TCP 輻輳制御に見られる恒常性

スループット予測をさらに高精度にするためには、TCP に対する理解を更に深める必要がある。スループット予測の分野では、TCP の挙動を数式で表現することで行う Formula-based アプローチが知られている。しかし、現状では、機械学習ベースの方式のように、History-based のアプローチの方が高い精度を達成可能である。History-based は TCP の挙動に関する知識を明示的には用いずに予測を行っているが、TCP が示すカオス的な振る舞いまでを捉えた結果であると考えることは難しい。そのため、スループットの予測には依然として改善の余地があると予想される。

我々は、スループット予測の新たな方式の模索のために、TCP のカオス的な挙動の更なる理解を深めるための実験を実施した。実験は、ネットワークシミュレータ上に構築された 30 ノードからなるネットワークにおいて、各ノード間に間欠的なフローを発生させた。フローの発生するインターバルの長さを調節することでネットワークへの負荷を調整し、それによる輻輳ウィンドウサイズ(cwnd)の挙動の変化を観察した。負荷が高くなるにつれ、cwnd の挙動はカオス的なものとなる。

主成分分析を通じた実験結果の観察を通じ、我々は、cwnd の挙動の安定性は内部のフィードバック制御が生み出すアトラクタの自己組織化の結果であるとの結論に至った。負荷が上昇するにつれて安定性が崩れる現象は、内部の遷移状態数の爆発的増加に起因しており、これは Ashby's Law of Requisite Variety として知られる法則と合致している。

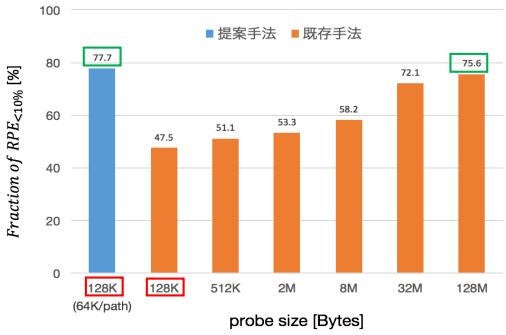
(b) 機械学習に基づく TCP スループット予測手法のマルチパス TCP への拡張

(1) (a) で述べたように、我々はマルチパス TCP を利用したシステムのプロトタイピングに取り組んでいる。その過程で、既存の機械学習ベースのスループット予測手法では、シングルパスの場合と同様の精度を達成するのが難しいことを発見した。

そこで我々は、その原因の調査を行い、マルチパス TCP 向けに予測手法の拡張を行った。マルチパス TCP では、複数のシングルパス TCP フローを束ねる形で単一の広帯域な TCP フローを実現する。このとき、元となるシングルパス TCP フローはすべて同時的ではなく逐次的に確立される。そのため、短時間のプローブ転送ではすべての経路を網羅的にテストすることができないという問題が生じる。これを解決するため、我々は、複数のシングルパス TCP のプローブ結果から、その上で実現

されるマルチパス TCP フローのスループット予測を実現する手法を実現した。

広域 OpenFlow 環境を模して作成したローカルのテストベッドにおける実験を通じて、我々の提案手法は、既存の手法をそのまま適用した場合と比較して大幅な精度向上を達成可能であることを示した。同一のプローブ転送サイズを用いる場合は、既存手法による予測精度は 47.5%であるのに対し、提案手法は 77.7%の精度を達成した。また、既存手法と同程度の精度を提案手法で達成するために必要なプローブサイズは、既存手法の場合の約 1000 分の 1 で済むことが示された（下図）。



５．主な発表論文等

〔雑誌論文〕（計　１件）

① Mizuki Oka, Hirotake Abe and Takashi Ikegami. Dynamic homeostasis in packet switching networks. Adaptive Behavior, vol. 23, no. 1, pp. 50-63, February 2015, DOI: 10.1177/1059712314556369.（査読有）

〔学会発表〕（計　３件）

① 浜崎 拓也，阿部 洋丈，加藤 和彦．Linux MPTCP kernel のネットワークスループット予測に関する研究．2015 年並列／分散／協調処理に関する『別府』サマー・ワークショップ（SWoPP2015），2015 年 8 月 4 日、ビーコンプラザ 別府国際コンベンションセンター（大分県別府市）

② 前田達憲，阿部洋丈，加藤 和彦．スループット予測による経路選択を用いた OpenFlow 環境での MPTCP 転送，情報処理学会 第 127 回 システムソフトウェアとオペレーティング・システム研究会，2013 年 12 月 3 日，芝浦工業大学 豊洲キャンパス（東京都江東区）

③ Chunghan Lee, Hirotake Abe, Toshio Hirotsu and Kyoji Umemura. Performance Implications of Task Scheduling by Predicting Network Throughput on the Internet. The 11th IEEE International Symposium on Parallel and Distributed Processing with Applications, Melbourne, Australia, 16-18 July, 2013.

６．研究組織

(1)研究代表者

阿部 洋丈（Abe, Hirotake）
筑波大学・システム情報系・准教授
研究者番号：00456716