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　６月　１日現在

機関番号：１４４０１
研究種目：若手研究（Ｂ）
研究期間：２０１０～２０１１
課題番号：２２７０００２９
研究課題名（和文）大規模障害に耐えられる広域分散アーカイバルストレージの実現

研究課題名（英文）Realization of wide-area distributed archival storage tolerable to large-scale failures

研究代表者
阿部　洋丈（ABE HIROTAKE）
大阪大学・サイバーメディアセンター・助教
研究者番号：００４５６７１６

研究成果の概要（和文）：

インターネットの広範囲に影響を及ぼすような大規模な障害が発生した場合においても複数拠点に分散したデータへのアクセスをできる限り維持できるようにする広域分散アーカイバルストレージの実現に向けての基礎的検討を進めた．具体的には，長期観測データからインターネットで発生した大規模障害の痕跡を見つけ出す方法と，複数のクラウド環境を協調させる際のデータ転送スループットの推定手法の実現を進めた．

研究成果の概要（英文）：

We have conducted fundamental research that are bases for realizing wide-area distributed archival storage system that will retain accessibility as much possible under large-scale failures in the Internet. Specifically, we concentrated on the following two topics: (1) methodology for finding the vestiges of the large-scale failures actually happened in the Internet by investigating daily snapshots of Internet topology that are accumulated for approximately 8 years, (2) methodology for predicting data transfer throughput between two hosts in cloud computing environment.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2010 年度	800,000	240,000	1,040.000
2011 年度	600,000	180,000	780,000
年度			
年度			
年度			
総　計	1,400,000	420,000	1,820,000

研究分野：　総合領域
科研費の分科・細目：　情報学・ソフトウェア
キーワード：　並列処理・分散処理，分散ストレージ，広域分散システム

１．研究開始当初の背景

人類の様々な活動におけるインターネットの重要性はこの 10 年間で飛躍的に高まった．しかし，インターネットは依然として障害に対する脆弱性を孕んでいる．2009 年 4 月 10 日には，シリコンバレーの中心地であるサンノゼ地区の地下に通信会社 AT&T が敷設し

ていた光ケーブルの 500 本以上が何者かによって切断されるという事件が発生した．シリコンバレーに設置されているサーバの利用者までを含めると，このアクシデントの影響範囲はサンノゼ地区における社会インフラの停止にとどまらず，全世界に及び，影響を受けた利用者の数は計り知れない．この事件は，インターネットが依然として持つ脆弱性の存在を如実に世間に知らしめる結果に至った．今や社会的基盤の一つとなりつつあるインターネットの利用可能性(Availability)の向上は，この 10 年間続いた発展を停滞させないためには一刻を争う急務であると考えられる．

２．研究の目的

本研究の目的は，前述のような大規模なインターネット障害が発生した状況下においてもできる限り多くの利用者に対してサービスを提供し続けられるようにするための実現に向けて，そのための基盤となる各種技術の研究を行うことである．

今回の研究では特に，広域分散アーカイバルストレージの実現を念頭に置いた上で研究を進めた．アーカイバルストレージとは，次々と生み出されるデータを蓄積保存することを主な目的としたストレージシステムである．そのようなシステムに蓄積されるデータとしては，科学実験の結果として得られたデータから，行政やビジネスの結果として生成される保存文書，さらには歴史資料や美術品などの文化財を電子化したデータなど非常に多岐に渡る．

アーカイバルストレージシステムは，通常のストレージシステムに比べて，データの読み出し／書き込みのバランスが書き込み側に極端に偏っているという特徴がある．しかし，アーカイバルストレージにとって最も重要なのは，稀にしか起きない読み込み操作を適切に処理できるかどうかである．いくらデータ蓄積時にトラブルが生じていないとしても，読み込み時に必要なデータを読み出すことができなければ意味が無い．

データの読み出し可能性を向上させるためには，データの複製を複数の拠点に置いておくことが効果的である．そうすることで，たとえある拠点が火災等のために完全にアクセス不能に陥ったとしても別の拠点からデータを読み出すことが可能となる．その際，大規模停電や地震などの発生する可能性を考えると，各拠点間の距離は物理的に十分離れていることが望ましい．本研究で対象とする広域分散アーカイバルストレージシステムは，物理的に離れた複数の拠点がインターネットを介して通信を行いながらデータの蓄積・複製・保存を行うシステムを想定している．

インターネットを介して接続された拠点間で広域分散アーカイバルストレージを構成する場合は，各拠点における障害に加えて，インターネットそのものの障害に対する備えも同様に重要なポイントとなる．なぜなら，いくら各拠点が正常に動作していたとしても，それらへのアクセス網であるインターネットが通信不能な状態に陥っていては，その間はデータの読み出しが出来なくなってしまう恐れがあるためである．

３．研究の方法

インターネットを介して相互接続されたアーカイバルストレージの実現に向け，本研究では次の二点について研究を行った．一つは，前節で述べたようなインターネット自体の大規模障害に備えるために必要十分な複製数を合理的に決定するために，インターネットで実際に発生している障害事例を自動的に収集するシステムに関する研究である．もう一つは，インターネットを介して接続された複数拠点間でサイズの大きなデータを複製する場合に重要となる，データ転送スループットの予測技術に関する研究である．

(1) インターネット障害の痕跡発見

インターネットとは，その名前の通り，複数のネットワークの相互接続（インターコネクション）によって形成されている．インターネットを構成する各ネットワークは AS（Autonomous System；自律システム）と呼ばれ，それぞれ独自のポリシーを持って運用されている．

我々の先行研究の結果により，インターネットを AS の相互接続網のレベルで捉えた場合，ある次数（他の AS への接続リンク数）を持つ AS が通信障害を起こした場合に他の AS がその影響を受ける確率についての関数モデルが既に得られている．その関数は指数関数の形で表され，各パラメータは実際の AS レベル相互接続のトポロジから決定されている．

その関数モデルに基づいてインターネット障害に対応するための必要十分な複製数を決定するためには，次数を確率変数とした障害発生件数の確率密度関数についてもモデル化する必要があるが，それらを決定するための網羅的な事例データはこれまでに存在していない．その理由は，障害の発生事例は，各 AS 単位では把握されていたとしても，ビジネス上の機密保持もしくはセキュリティ上の懸念により，その情報が外部に公開されるケースは殆ど無いためである．

そこで我々は，2000 年から 2008 年という

長期間にわたって定期的に AS 間相互接続のトポロジを記録したデータセット（Skitter AS Links Dataset(CAIDA)）を元にして，その中から障害事例と疑わしいケースを自動的に抽出する方法に関する検討を行った．我々のアイディアは以下の通りである；まず，前述のサンノゼの例のように新聞等で報道されるくらい大規模な障害の事例を人手で収集する．次に，それらのデータが前述のデータセット内にどのように現れているかを調べる．そして，そこで現れた特徴と類似する特徴を持つケースを機械的に収集する方法の確立を目指す．

(2) データ転送スループット予測

インターネットでデータを転送する際，データはパケットと呼ばれる単位に分割された上で送信元から受信先へ配送される．また，複数のパケット通信が一つの機器に集中した場合，機器の受信バッファがあふれ，パケットの廃棄が起こる場合がある．パケットの廃棄が起きると，パケットが到達しなかったことを送信元が検知し，そのパケットを再送するという手順が必要となるため，スループットなどの通信性能は低下する．

インターネットでは，非常に多くのユーザが同時に通信網を利用しているため，パケットの混雑の状況は刻一刻と変化し，その現状を把握することや，その結果としてどのようなスループット性能が得られるかを予測するということは容易ではない．これまでにも通信性能の予測の試みはいくつか行われているが，予測のために必要とされるデータの収集に大きな手間やコストがかかるものや，データ測定は単純だがノイズの影響を大きく受けてしまうものなど，満足の行くものはまだ存在していなかった．

そこで我々は，データ測定の容易さと高い精度を兼ね備えたスループット予測システムの実現に向けた研究を行った．予測のためのデータには，プローブ転送と呼ばれる，比較的小さなサイズのデータ転送にかかる時間を用いる．プローブ転送は，OS やハードウェアによる特別な支援が無くても簡単に実行できる測定方法であり，既存研究でも用いているものがあるが，そのようなものはノイズに対する影響が大きいという問題があった．

我々のアプローチは次の 3 つである．まず，近年のクラウドコンピューティング環境の大半が仮想計算機技術を利用している点に注目し，通信混雑によってスループットが低下しているケースと，同一ホスト上の他の仮想計算機の影響でスループットが低下しているケースの切り分けを行う．次に，プローブ転送のためのデータサイズ等のパラメータを実際の環境に合わせて適切に選択するために，ノンパラメトリックな順位相関係数を用いてパラメータのチューニングを行う．最後に，前述 2 つのアプローチによって得られた学習データに対して，Support Vector Regression (SVR) を用いた非線形回帰分析を適用することで精度の高い予測曲線を得る．

４．研究成果

(1) 大規模インターネット障害の痕跡発見

2008 年の 1 月末から 2 月初めにかけて地中海で発生した海底ケーブル切断事故の事例を題材に，その結果が前述の Skitter AS Links Dataset においてどのように現れているかを調査した．AS 間の相互接続グラフはサイズが大きく，また，測定方法に起因するノイズも多く，目視での確認は難しい．そこで我々はグラフの特徴を数値として表すことのできる指標値に着目し，問題のグラフに対して様々な指標値を計算し，その値の変化から障害の発生を読み取ることができないか試みた．その結果に基づき，頂点数とクラスタリング係数の 2 つの指標値が障害痕跡の発見に有望であるとの結論を得た（図 1）．本

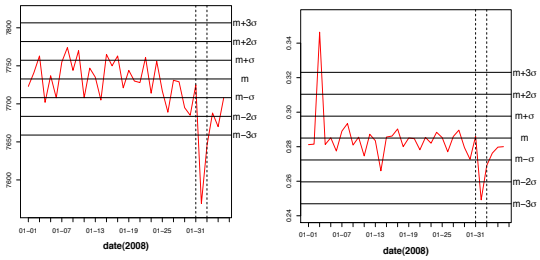

図 1.　頂点数（右）とクラスタリング係数（左）の変化

その後，この知見に基づいて約 8 年間分のデータセット全体に対して網羅的に障害痕跡と思われるケースを抽出するソフトウェアの実装と実験に現在も取り組んでおり，今後も継続して行う予定である．

(2) データ転送スループット予測

本研究では，PlanetLab と呼ばれる広域分散テストベッド環境上でプローブ転送とデータ転送を繰り返し行って収集したデータセットに対して前述の 3 つのアプローチに基づくスループット予測手法の適用実験を行った．また，比較のために，累積密度関数(CDF)に基づく予測手法の再現も行い，我々の提案手法と比較を行った．その一例を図 2 に示す．様々なケースで比較検討を行った結果，全体的な予測精度やノイズに対する頑健

性において，既存手法よりも提案手法の方が優れているということが確認できた．

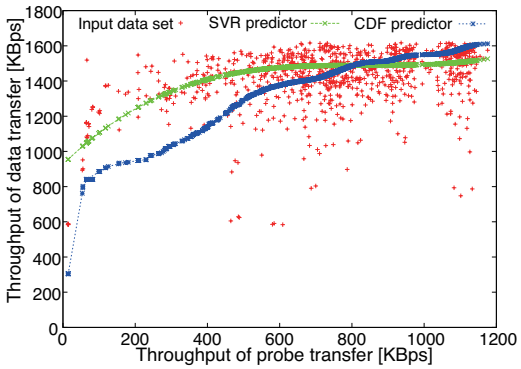

図２． 提案手法(SVR)と既存手法(CDF)の比較の例

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〕（計 2 件）

① Chunghan Lee, Hirotake Abe, Toshio Hirotsu and Kyoji Umemura. Analytical Modeling of Network Throughput Prediction on the Internet. IEICE TRANSACTIONS on Information and Systems. 査読有. (accepted for publication)

② Chunghan Lee, Hirotake Abe, Toshio Hirotsu and Kyoji Umemura. Traffic Anomaly Analysis and Characteristics on a Virtualized Network Testbed. IEICE TRANSACTIONS on Information and Systems, 査読有, Vol. E94-D, No. 12, pp. 2353--2361, Dec 2011.

〔学会発表〕（計 7 件）

① Chunghan Lee, Hirotake Abe, Toshio Hirotsu and Kyoji Umemura. A Statistical Approach for Selecting Throughput Prediction Parameters on the Internet. The 6th International Conference on Ubiquitous Information Technologies & Applications (CUTE 2011), pp. 37--40, December 15, 2011 ,Seoul, Korea.

② Chunghan Lee, Hirotake Abe, Toshio Hirotsu and Kyoji Umemura. Predicting Network Throughput for Grid Applications on Network Virtualization Areas. NDM2011: The 1st Workshop on Network-Aware Data Management, November 14, 2011, Seattle, USA. Awarded the Best Paper Honorable Mention.

③ Chunghan Lee, Hirotake Abe, Toshio Hirotsu and Kyoji Umemura. Estimating Traffic Anomalies for Throughput Prediction on Network Virtualization. The Second Workshop on High Speed Network and Computing Environments (HSNCE 2011), July 19, 2011, Munich, Germany.

④ Hirotake Abe, Hirotaka Moriya and Kyoji Umemura. On Finding Vestiges of Internet Backbone Failures for Optimizing Wide Area Data Replication. WOSD: The 1st International Workshop on Open Systems Dependability：Adaptation to a Changing World, pp. 230--233, June 27, 2011, Hong Kong, China.

⑤ Chung-Han Lee, Hirotake Abe, Toshio Hirotsu and Kyoji Umemura. Internet Traffic Characteristics of Virtual Machine on Amazon EC2. 情報処理学会 第 117 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, 2011 年 4 月 14 日, 沖縄県那覇市.

⑥ Chung-Han Lee, Hirotake Abe, Toshio Hirotsu and Kyoji Umemura. A PCA Analysis for Traffic Anomaly Estimation. Internet Conference 2010 (IC2010), October 26, 2010, Tokyo, Japan.

⑦ Chunghan Lee, Hirotake Abe, Toshio Hirotsu, Kyoji Umemura. Analysis of Anomalies on a Virtualized Network Testbed. The 2010 IEEE International Conference on Computer and Information Technology (CIT 2010), June 29, 2010, Bradford, UK.

６．研究組織
(1)研究代表者

阿部　洋丈（ABE HIROTAKE）
大阪大学・サイバーメディアセンター・助教
研究者番号：００４５６７１６