

計測内容のシナリオ記述とタスクの集約を用いた 分散アクティブ計測基盤

板谷諭[†] 池部実[†] 藤川和利[†] 砂原秀樹[†]

[†] 奈良先端科学技術大学院大学 情報科学研究科

概要

近年、インターネットがインフラとして発展することにより、通信品質面での要求も高まっている。それに伴い、ネットワーク計測技術の重要性が増してきている。広域ネットワークの計測には、複数のプロバイダが管理するネットワークを越えての計測や、多数の計測ホスト、計測ツールの管理、運用を可能とする分散計測基盤が必要となる。本研究の目的は、スケーラビリティだけでなく、ユーザごとの異なる計測要求に対応し、多数の計測ホストと計測ツールを利用するための統一されたユーザインタフェースを提供する分散計測基盤を開発し、広域ネットワークでの計測を支援することである。提案システムでは、広域ネットワークの品質推定のための計測手法としてアクティブ計測を利用する。また、ユーザインタフェースとして計測シナリオを記述するためのシナリオ記述言語を提供し、ユーザ毎の計測要求に対応する。さらに、多数の計測ホストのある特定の条件でのグループ化と複数のユーザからの重複する計測の集約により、計測トラフィックを削減し、スケーラビリティを向上させる。実証実験の結果、提案システムによる広域での計測が可能であるか検証した

Distributed Active Measurement System that offers an scenario description language for measurement, task merging

Satoshi Itaya[†] Minoru Ikebe[†] Kazutoshi Fuzikawa[†] Hideki Sunahara[†]

[†]Nara Institute of Science and Technology Graduate School of Information Science

Abstract

The Internet has become an infrastructure, which requires a high quality of the communication. Besides, network measurement technologies are getting more important in the Internet. The measurement within wide area networks has to be done through a number of networks and has to control many hosts. This research aims to develop a distributed measurement system, which is able to support for measurements of users in a wide area network. The proposed system uses active measurement in order to estimate the quality of wide area networks. Users can specify the rules for measurement with the scenario description language of the system. The system, then, will merge the same measurement rules. In this research, it has been confirmed that the proposed system is able to measure in wide area network.

1 はじめに

近年、インターネットが急速に発展し、重要なインフラとして活用されるようになり、接続性だけでなく通信品質も問われるようになってきている。そのた

め、ネットワークの状態や品質を計測する計測技術の重要性が増してきている。

ネットワーク計測技術は様々な場面での利用が期待されている [6]。一つ目の例として、アプリケーションやサービスの最適化が挙げられる。例えば、

通信経路の帯域や遅延などの情報があれば、ネットワークゲームやFTPなどのサービス利用時に、クライアントからネットワーク距離が近いサーバを選択することが可能となる。二つ目の例としてQuality of Service(QoS)が存在する。QoSは動画のストリーミングやEコマース等の様々な品質要求を持つアプリケーションやサービスへの利用が期待されている。また、ネットワーク運用における計測の利用例として、Service Level Agreement(SLA)や、ネットワーク上のボトルネックの発見やインターネットのトラフィックの成長予測などがある。ほかにも、新しいプロトコルやアプリケーションなどの研究開発に役立てることが期待されている。

広域ネットワークの計測では、パッシブ計測の場合、観測地点の設定が困難となる。そのため、アクティブ計測が必要となる。その際、多数のノードの操作を行う。よって、多くのノードを一括して利用できるシステムが求められる。また、ノードが何らかの事情により動作しなくなったときにでも安定して計測できなければならない。本研究ではこれらの条件を満たした分散計測システムの提案と実装を行い、ユーザの広域ネットワーク計測の支援を行う。

2 関連研究

広域計測基盤とは、広域ネットワークの計測を支援するためのシステムである。通常、広域計測基盤はネットワーク上に分散する多数の計測ノードを有しており、各計測ノードを用いて計測を行う。

Sophia[3]は、刻一刻と状態が変化するネットワーク状態を把握するための広域計測基盤である。Sophiaは、ユーザが行いたい計測内容を記述するための言語を定義している。Sophiaにおいて、各ノードは過去の計測データを保存しており、いつでも参照できる。計測データは、各ノードで分散管理されている。

蟹江[5]らの研究は、一台の統括ホストと複数の計測ホストから構成され、複数の計測ホストが協調してアクティブ計測を行うことを目的としている。計測時に複数の計測ホスト間の同期問題とユーザが個々の計測ホストを選択し、アクティブ計測のためのジョブ割り当て作業の負担を削減することに重点をおいたシステムである。計測内容を記述するための専用言語によってシナリオを記述する。また、計

測ホストのグループ化を行っており、グループ単位での計測を行う。

HP Labsのyalagandulaらが開発したA Scalable Sensing Service[4]は、広域ネットワークでのリアルタイムかつ正確な計測データを収集するための分散計測システムである。アプリケーションのパフォーマンスの改善やリンク切れ、DoS攻撃などの原因解明への利用を目的としている。

これらの関連研究においては、計測ノードをネットワークアドレスやASレベルでグループ化している。また、過去の計測データを蓄積しておきユーザの要求に応じて過去のデータを返信するなどの工夫が行われている。しかしながら、いずれのシステムも複数のユーザからの要求が重複した場合の処理の集約機能は実現されていない。

3 広域計測基盤システム要件

広域計測基盤システムに求められる要件は以下の4点である。

- ユーザごとに異なる計測目的への対応
- システムのスケールビリティ
- ユーザが多数のノードを用いて計測結果を得るための操作手段
- 計測データの再利用

1つめは、ユーザごとに異なる計測目的へ対応するというのは、システム側に各種計測ツールをモジュールとして組み込む事でユーザが自由に計測できるという事である。この機能によってユーザ自身が各計測ツールを組み合わせるより計測が容易になる。

2つめは、分散アプリケーションテストベッドであるPlanetLab[1]で動作させることを考慮すると、600~1000台のノードがストレスなく動作することが求められる。

3つめは、ユーザが多数のノードを用いて計測結果を得るための操作手段である。従来の研究では計測用のAPIを提供することにどまっているものが多い。しかし、ノードを用いて計測結果を得るためには一連の動作を定義でき、ユーザが操作するためのユーザインタフェースが必要となる。

4つめは、計測データの再利用である。即時性を求めない計測や過去の計測データを集めて過去の傾

向としてみる場合があるので、過去に計測したデータを保存し、いつでも利用できるにすることが求められる。

4 広域計測基盤システムの提案

本研究では、シナリオ記述言語を用いて計測内容の記述を可能とし、計測ノードを様々な単位でグループ化する機能、複数のユーザの計測内容の集約機能を提供するシステムを提案する。

4.1 システム概要

本システムは、広域ネットワーク上で多地点に存在する計測ノードを用い、アクティブ計測によりネットワーク品質を測定する。広域計測システムの概要を図1に示す。本システムの構成要素は以下に示す。

- ゲートウェイサービス

ユーザインタフェースをユーザに提供する。システムへの窓口であり、ユーザは全てゲートウェイサービスを利用して操作を行う。

- 計測ホスト

計測ホストとは、広域ネットワーク上に設置された計測ノードを意味する。各計測ホストは、計測機能、タスクスケジューリング機能、計測データ保存機能をもつ。

本システムでは、全ての計測ホストはグループ化され、ユーザはグループ単位で計測ホストを操作する。本システムでは、ユーザが行いたい1つの計測内容をタスクと呼ぶ。各グループには1台以上のホストが所属し、その内の1台が代表ホストとなる。ユーザは、ウェブベースのインタフェースを提供するゲートウェイサービスを介して、各グループの代表ホストにタスクを投入する。本システムの操作は、専用のシナリオ記述言語によって行う。専用のシナリオ記述言語を用意することにより、ユーザの異なる要求に柔軟に対応することが可能となる。タスクを投入されたホストは、シナリオ記述言語をインタプリタで解釈し命令を計測に必要なグループに送信する。命令の送受信は、各グループの代表ホストが行う。命令を受信した代表ホストは、タスクスケジューリング表に新たなタスクを追加し、同一

グループ内の計測ホストにタスクを割り当てる。計測結果は、計測を行ったグループの代表ホストに保存される。ユーザは必要に応じて、過去の計測履歴を検索、閲覧することが可能である。また、タスクスケジューリング表は、予定を格納するだけでなく、複数のユーザからタスクが投入された場合に重複するタスクを集約する。

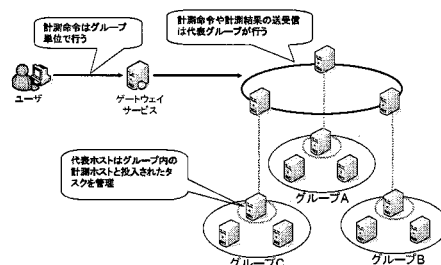


図1: システム概要図

4.2 計測ホストのグループ化

本システムでは、全ての計測ホストはグループ化され、ユーザはグループ単位で計測ホストを操作する。グループとは何らかの共通点をもつ計測ホストの集合である。グループ化の具体的な仕組みを以下に示す。

- グループ化の概要

計測ホストにグループ属性を表すための識別子であるタグを付与する。全ての計測ホストはタグによりグループ化される。タグによるグループ化を行うことで、計測ホストをネットワーク単位だけでなく、サービス単位などで分類することが可能である。タグによるグループ化は、計測トラフィックを削減するだけでなく、ユーザに計測ホストを選出する基準にもなる。

- 計測ホストのグループへの参加

計測ホストを既存グループへ登録する場合を説明する。ユーザは、ゲートウェイサービスで計測ホストと付与するタグを入力する。ゲートウェイサービスは、タグが示すグループの代表ホストに新しく参加する計測ホストの情報を通知する。代表ホストは、管理リストに登録するホストを追加し、その情報を返信する。参加す

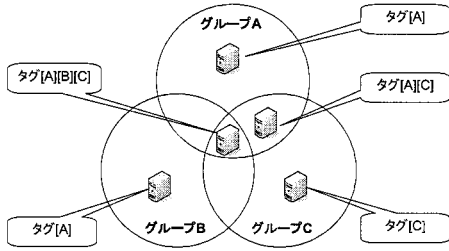


図 2: タグによるグループ化の例

る計測ホストは、返信を受け取り参加が承認されたことを確認し、グループに参加する(図 3)。

● 新規グループの登録

新規グループを作成する場合を説明する。ゲートウェイサービスは、新規登録したタグと計測ホストの情報を全てのグループの代表ホストに通知する。新規グループ作成時に最初に関連付けされた計測ホストは自動的に代表ホストとなり、他のグループの代表ホストのリストをゲートウェイサービスから受け取り、広域計測システムネットワークに参加する(図 4)。

● 代表ホストの交代

グループ作成時に最初に関連付けされた計測ホストが代表ホストになるが、代表ホストがなんらかの原因で機能しなくなった場合には代替ホストを選出し代表ホストの交代を行う。グループ内に最も長く所属している計測ホストが代替ホストとなる。

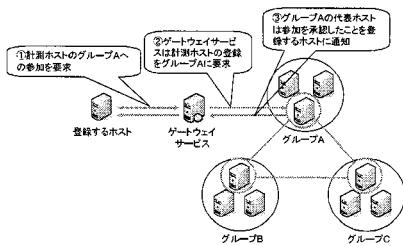


図 3: 計測ホストのグループへの参加

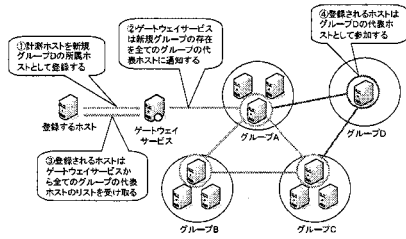


図 4: 新規グループの登録

4.3 シナリオ記述言語とインタフェース

本システムでは、ユーザの異なる要求やグループ単位での計測ホストの操作をシナリオ記述言語で表現する。ユーザはシナリオを記述しシステムへ投入する。インタプリタによって、シナリオは解釈されシステムのタスク管理表へ登録される。ノードやデータや計測ツールの扱いをシナリオ記述言語により一本化することで、ユーザビリティの向上を図る。一度、記述したシナリオは登録され、後から使うことも可能である。記述例として、グループ A,B 間の可用帯域 30Mbps 以上の場合、ジッターを続けて計測するシナリオを図 5 に示す。

```
//引数が 0 の場合は標準の設定で計測が行われる
//term=0 の場合は即座に計測
//opt は計測内容ごとに設定できるオプション
SendCmd(GrpA, GrpB, ABW, 0, 0, 0)
if(ABW > 30)
    data = reply()
    SendCmd(data. src, data. dst, Jitter, 0, 0)
end
```

図 5: 記述例

4.4 タスクスケジューリング機能

本システムは、タスクスケジューリング機能を有しており、現在の状況を調べたい場合や、長期間に渡り計測を行いたい場合などにも対応している。また、タスクスケジューリング機能により、重複しているタスクを集約することも可能である。

● グループ単位でのスケジュール管理

タスクのスケジュール管理をグループ単位で行っている。ユーザがタスクを投入すると、計測時に受信側となるグループに計測命令が送信される。受信側グループの代表ホストは、投入されたタスクをタスクスケジューリング表に追加し、同一グループ内から計測時に受信側ホストとなる計測ホストを1台選出し、送信側グループに計測命令と共に通知する。送信側グループも、受信したタスクをタスクスケジューリング表へ追加し、計測ホストを1台選出する(図6)。

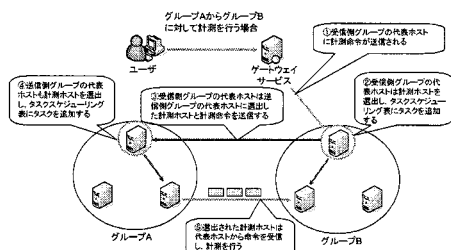


図6: スケジュール管理

● タスクスケジューリング表

タスクスケジューリング表は受信タスクスケジューリング表と送信タスクスケジューリング表の2つから構成される。受信タスクスケジューリング表は、グループが計測時に受信側となる場合のタスクを管理する。送信タスクスケジューリング表は、グループが計測時に送信側となる場合のタスクを管理する。ユーザからタスクが投入されると、計測時に受信側のグループの代表ホストが計測命令を受信する。タスクを受信した代表ホストは、送信タスクスケジューリング表を参照する。これは、競合する計測内容の有無を確認するためである。次に、受信タスクスケジューリング表を参照し、重複する計測内容の有無を確認する。同じ計測内容がスケジュールにあればタスクを追加せず、なければ追加する。また、計測内容が一部重複している場合は、その部分だけタスクを集約する。送信側のグループの場合、まず、受信タスクスケジューリング表を参照する。そして、競合する計測内容の有無を確認する。次に、送信タスクスケジューリング表を参照し、重複するタスクの有無を確認し、なければ、タスクを追加し、

計測時の送信側ホストを1台選出する。計測ホストの選出は、タスクスケジューリング表を参照し、計測タスクを割当てていないホストを優先して利用する。

4.5 データの分散保存

多数の計測ノードを用いて広域ネットワークの計測を行う場合、膨大な計測データを蓄積しなければならない。本システムでは、計測データをグループ単位で管理することにより、計測データを分散管理する。計測データは、計測時の送信側ホストが取得する。

4.6 計測トラフィック削減効果の実験

設計に基づき本提案のプロトタイプ実装を行った。本システムは、Fedora Core 5上で、Ruby1.8.5を用いて実装した。構成をシンプルにするため、グループ管理機能、タスク管理機能はゲートウェイサービスが保有する。実験は、PlanetLab[1]のホスト約300台で行った。また、各ホストを提供組織毎(サイト単位)でグループ化した。グループ数は約150であり、各グループには平均2台のホストが参加する。実験では、提案システムをPlanetLab上のホストにインストールし、RTTを計測した。そのデータを基に、以下の条件でトラフィック量を比較した。2人のユーザが計測シナリオを投入して計測する。その際に、グループ化・集約機能無し、集約機能のみ無し、集約50%、集約100%の4つのケースを比較する。集約50%とは2人のユーザのシナリオの50%が重複していることを指す。

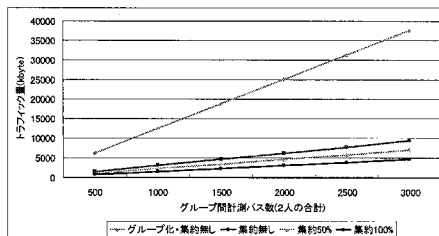


図7: トラフィック量の比較

表 1: 送受信タスクスケジューリング表

送受信タスクスケジューリング表	保有する情報
送信タスクスケジューリング表	送信側グループ 受信側グループ 選出された送信側計測ホスト 選出された受信側計測ホスト 計測内容 (RTT, jitter, ...) 計測期間
受信タスクスケジューリング表	送信側グループ 受信側グループ 選出された受信側計測ホスト 計測内容 (RTT, jitter, ...) 計測期間

4.7 提案システムのユーザビリティに対する効果

提案システムでは、ユーザはグループ名をホストの選出基準として用いる。また、シナリオ記述を行う場合、ユーザはグループ内のホストの選出や他のユーザとのタスクの重複などを意識する必要はない。

4.8 考察

実験より、グループ化と重複するタスクの集約により計測トラフィックが削減されていることが確認できる。また、グループ化と計測シナリオ記述言語によりユーザビリティの向上を図った。

4.9 おわりに

本研究では、広域ネットワーク上で多地点に設置されたノードが、ユーザの記述したシナリオによって計測を行うためのシステムの提案を行った。今後は、プロトタイプシステムの実装をすすめて、提案システムの評価を行っていく予定である。また、シナリオ記述言語はプロトタイプシステムのための必要最低限な命令セットしか定義していない。ユーザの異なる要求に対応するためには、シナリオ記述言語の拡張及び、インタプリタの実装を行う予定である。

参考文献

- [1] PLANETLAB. <http://www.planet-lab.org>.
- [2] R. Mahajan, N. Spring, D. Wetherall, and T. Anderson. User-level internet path diagnosis. *Proceedings of the nineteenth ACM symposium on Operating systems principles*, pp. 106–119, 2003.
- [3] M. Wawrzoniak, L. Peterson, and T. Roscoe. Sophia: an Information Plane for networked systems. *ACM SIGCOMM Computer Communication Review*, Vol. 34, No. 1, pp. 15–20, 2004.
- [4] P. Yalagandula, P. Sharma, S. Banerjee, S. Basu, and S.J. Lee. S 3: a scalable sensing service for monitoring large networked systems. *Proceedings of the 2006 SIGCOMM workshop on Internet network management*, pp. 71–76, 2006.
- [5] 蟹江弘士, 片山喜章, 高橋直久. 協調型アクティブモニタリングシステムの実装と評価. 信学技報, 第 105 卷 of *IN2005-155*, pp. 73–78, 高知, 2月 2006. 2006年2月16日(木) – 2月17日(金) 高知工科大学 (IN).
- [6] 鶴正人, 尾家祐二. インターネットの特性測定技術とその研究開発動向. 情報処理学会会誌, Vol. 42, No. 2, pp. 192–197, 2001.