マルチメディア通信と分散処理 71−24
（1995. 7. 14）

多点間接続でのTCP性能評価システムDBSの提案

村山 公保† 門林 雄基‡ 山口 英† 山本 平一†

†奈良先端科学技術大学院大学情報科学研究科
‡大阪大学基礎工学部情報工学科

TCPの性能評価によく用いられているベンチマークは2点間のホストという限定された環境下でのスループットしか測定できない。そのためTCPの一部の機能の性能しか測定できないという欠点がある。

本稿では多点間接続でのTCP性能評価システムDBS(Distributed Benchmark System)を提案する。これは，現実に運用できるさまざまな環境において，TCPの持つ機能全ての性能測定を行うことを目的としたシステムである。このシステムを使用して現在のTCPの性能・能力の評価・分析を行い，来るべき未来のネットワークに適した次世代のTCPの設計および評価に利用することを目指している。

A proposal of DBS:performance evaluation for TCP over multipoint connection

Yukio Murayama† Youki Kadobayashi‡ Suguru Yamaguchi† Heiichi Yamamoto†

†Graduate School of Information Science, Nara Institute of Science and Technology
‡Department of Information and Computer Science, Osaka University

TCP/IP benchmark systems, mainly used for TCP performance measurements, have the limitation that their application are limited to point-to-point configuration. Their major drawback is that they measure only a subset of the entire TCP functions.

In this paper, we propose a TCP performance measurement system for multipoint configuration, named DBS (Distributed Benchmark System). Its goal is the performance measurement of the entire TCP functions in various operational environments. We plan to use this system for the performance evaluation and analysis of the current version of TCP, and for the design and evaluation of the next generation of TCP, which will be suitable for future networks.

1 はじめに

インターネットの急速な発達により，TCPはネットワークの重要なプロトコルとして定着してきた。しかしTCPは従来から使われている低速なネットワーク上で発展してきたため，本来TCPが想定していなかった高速なネットワークでは問題があることが指摘されている。

最も大きな問題はウインドウサイズの制限である。この制限のため，遅延の大きなネットワークではバンド幅に関係なく遅延時間によってスループットの上限が決まってしまう。この制限はTCPのオプション機能で拡張できるように提案されているが[3]，その提案には問題があることが指摘されている[1]。

また，Douglas E. Comer らは，小さなパケットの送信を抑制する機構 (Silly Window Syndrome Avoidance, Delayed ACK) の副作用により，100MbpsのATMリンク間でも1Mbps以下のスループットしか出ない場合があることを明らかにした[2]。T. Luckenbach らはメッセージサイズに着目してデータ伝送のスループットを測定し，OSのメモリ管理機構や，コンピュータのハードウェアアーキテクチャの特性がスループットに大きな影響を与えることを明らかにした[4]。

このように，スループットの測定からいくつかのTCPの問題が明らかになっている。だが，これらに用いられているベンチマークは2点間の通信という限定された環境下での性能測定しか行われておらず，フロー制御などのTCPの一部の性能しか評価できないという欠点がある。

本稿では2点間の性能測定の問題点を指摘し，多点間での性能測定の必要性について述べる。そして，多点間での性能測定を可能にするDBS(Distributed Benchmark System)を提案する。さらに実装について述べ，Ethernet，ATM，衛星回線での測定結果の例を示す。

2 TCPの機能と評価

TCP/IPはインタネットを支える基本となるプロトコルである。IPはリンク間の制御を行いTCPはエンドホスト間の制御を行う。TCPは信頼性のあるストリームを提供するトランスポート層プロトコルである。TCPは大きくわけて次の3つの制御機構を持つ。これらはTCPのパフォーマンスに大きく関係する。

フロー制御
エンドホスト間で最大のスループットが得られるように制御する。

再送制御
パケットが消失したかどうかをできるだけ素早く正確に予測して再送する。

輻輳回避制御
ネットワークの混み具合に合わせてパケットの送信量を制御し，パケットの破棄を防ぐ。

この3つ機能が協調して動作しなければ，ネットワークを有効に利用し，かつ，高スループットを得ることはできない。

また，データ伝送の性能はTCPの機構だけで決まるものではない。エンドホストのアーキテクチャやOSのメモリ管理などの機能および処理能力，データリンクの伝送速度や特性，中間ノードの機能や性能など，ネットワークを構成する全てのものが総合的に影響を与える。さらに，同じデータリンクを流れる他のトラヒックから影響を受けたり，逆にそのトラヒックに影響を与えたりする。

このようなさまざまな構成要素からなるネットワークにおいて，アプリケーションはTCPに，スループットが常に最大になるようにトラヒックをコントロールすることを要求される。この要求をふまえると，TCPの性能の目標は「TCPのサービスを利用するアプリケーションにとってスループットが大きくなること」と考えられる。このことから，TCPの性能を評価する上でアプリケーションレベルでの性能測定は有効な手段だと考える。

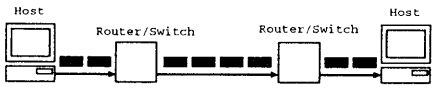

図1: 既存のベンチマークが想定している環境

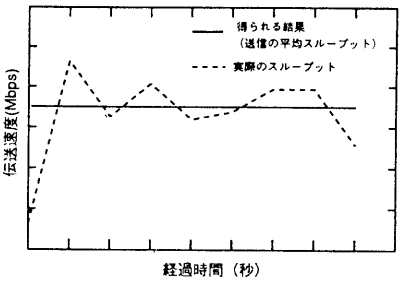

図2: 既存のベンチマークでの測定結果

3 既存のベンチマークの問題点

代表的なベンチマークの特徴とその問題点について述べる。

3.1 代表的なベンチマーク

TCP/IPネットワークの性能測定を目的とした代表的なベンチマークにはBallistics Research Laboratoryの Mike Muuss によって作成されたttcp，Hewlett-Packard Companyによって作成されたNetperfがある。なお前出の[4]と[2]ではスループットの測定にttcpを利用している。これらのベンチマークは輻輳やパケット消失のない2点間のホストでのスループットを求めることを目的に作られている(図1)。1回の測定結果で得られ結果は，

$$\text{スループット} = \frac{\text{伝送した総データ量}}{\text{データ伝送にかかった総時間}}$$

で求められたスループットである。そのため，実際のスループットが図2の点線で示されるように変化していたとしても，実線で示された平均値しか測定することができない。

3.2 問題点

これらのシステムは，トラヒックの無いネットワークの2点間でデータ伝送のスループットを測定するという限られた測定しかできない。またスループットの時間変動も測定できない。そのため現実のネットワークのような複数のホスト間でデータ伝送を行っ

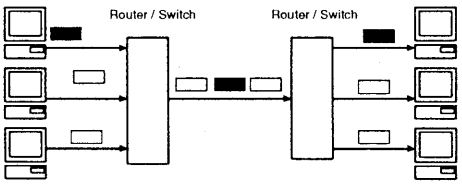

図 3: 提案システムの測定環境の例

た時のスループットや，同じデータリンクを流れる他のトラフィックから受ける影響や与える影響を測定できない。また，1つしかトラヒックの無いネットワークでは，TCP のフロー制御，再送制御，輻輳回避制御の内のフロー制御以外の機構はほとんど動作しない。そのため3つの機構が協調して動作しているかについては調べることはできず，TCP の問題点の全貌を明らかにすることはできない。

以上のことから，より現実に即した TCP の性能測定システムの開発が求められている。

4 DBS(Distributed Benchmark System) の提案

本節で DBS(Distributed Benchmark System) を提案する。システムの目的と概要を述べる。

4.1 目的

DBS は実際のさまざまな TCP/IP ネットワークの性能を測定することにより，TCP のプロトコルとその実装の総合的な性能を評価するシステムである。特に，TCP のフロー制御，再送制御，輻輳回避制御の3つの制御機構の性能と，それぞれがどの程度協調して動作しているか明らかにする。任意のネットワークトポロジーに対応し，現実のネットワークの利用形態のように多点間のホストでデータ伝送を行い，そのホスト間のデータ伝送が相互に与える影響を測定する。輻輳を発生させてトラヒックの変動の測定も行う。これら，TCP の総合的な性能を評価することが目的である。

4.2 DBS の提案

DBS は TCP で構築された分散環境におけるデータの送受信の性能を測定するシステムである (図3)。

多点間のホストでデータ伝送を行い，各ホストがデータの送受信時刻を記録し，それを集計して処理

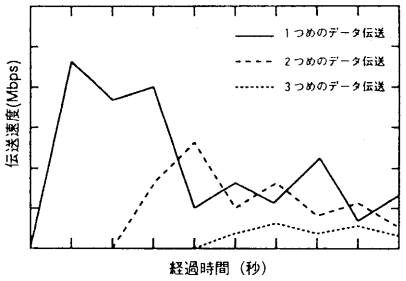

図 4: 提案システムでの測定結果

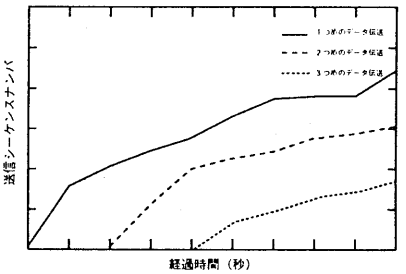

図 5: データ送信のパターン

することによりスループットとデータの送受信時刻の推移を計算する。その結果，図4，図5のようなグラフを描くことができる。これは図2で説明したような従来のベンチマーク得られる結果に比べ，はるかに情報量が多く，より多くの事象を解析できると考えられる。

5 DBS の実装

DBS は以下の環境のもとで動作するように実装した。

- 測定する全てのホスト間の時刻同期はとれている。
- バークレイソケットインターフェースをもつ UNIX システム上で動作する。

次節の測定では DBS が稼働するシステム間の時刻の同期に xntp を使用した。

本システムは，次のような特徴を備えている。

- アプリケーションレベルでのスループットを測定できる。(既存の性能測定ソフトウェアと同じ)
- 多点間のホストで複数のデータ伝送が可能。
- 各ホスト間のデータ伝送のスループットと遅延時間の時刻変動を測定できる。

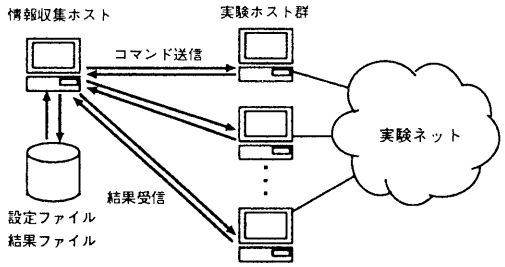

図 6: 提案システムのモデル

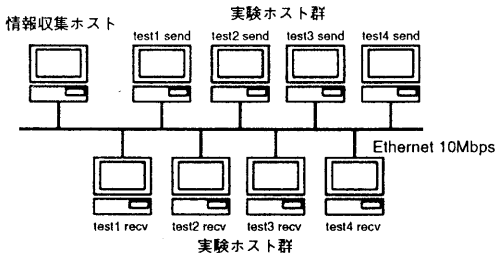

図 7: Ethernet での測定

実装したシステムの構成を図 6に示す。

DBS は次の 2 つのソフトウェアからなっている。

dbsd　実験ホスト群で動作させるデーモン
dbsc　情報収集ホストで実行するコマンド

DBS では次のような手順で性能測定を行う。

1. 情報収集ホストでテストの設定ファイルを作成する。
2. 全ての実験ホストで dbsd デーモンを実行する。
3. 情報収集ホストで dbsc コマンドを実行する。
4. 情報収集ホストから各実験ホストへのコマンドが送信される。
5. 指定された時刻にデータ伝送が始まる。
6. 指定された時刻にデータ伝送が終り，結果を情報収集ホストへ伝送する。
7. 情報収集ホストで収集された結果を解析する。

測定によって情報収集ホストに収集される結果は，各 TCP セグメントのシーケンスナンバ，データ長，送受信時刻である。この結果を処理することにより，データの送受信時刻やスループット・遅延時間の変動がわかる。

表 1: 測定環境とパラメータ

CPU	DECstation5000/25 (DEC)
OS	ULTRIX V4.3
TCP バッファサイズ	16384byte
セグメントサイズ	1024byte
データ伝送時刻	test1: 0.00 秒
	test2: 1.00 秒
	test3: 2.00 秒
	test4: 2.00 秒
送信回数	test1: 2000 回
	test2: 1000 回
	test3: 1000 回
	test4: 1000 回

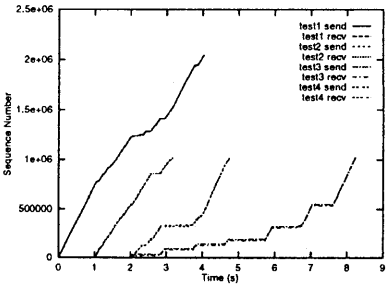

図 8: 各セグメントの送受信時刻

6　測定結果の例 (測定環境)

DBS を使用して，Ethernet，ATM，衛星通信上での TCP の性能を測定した例を示す。

6.1　Ethernet

Ethernet で測定した例を示す。図 7のような環境で測定を行った。使用したコンピュータおよび OS を表 1に示す。結果を図 8，図 9に示す。この測定では，test3 が異常なトラヒックを表している。特に 5 秒後には test3 以外にはデータ伝送は存在していないため全ての帯域を使い切る権利がある。しかし 8 秒後まで間欠的なデータ伝送をしている。これは，何らかの原因で輻輳制御またはフロー制御がうまく働かなかったと推測できる。

6.2　ATM LAN

ATM LAN を使用して測定した例を示す。図 10 のような環境で測定を行った。使用したコンピュータおよび OS を表 6.2に示す。結果を図 11，図 12に示す。test2 の伝送開始後 test1 は 1.5 秒の間データ伝送が行われていない。パケットが消失して再送タイマが働いたとも考えられるし，同一マシンで送信プログラムを動作させたため OS のプロセス管理に

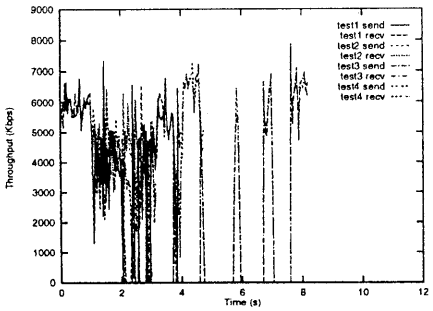

図 9: スループットの変化 (0.2 秒間隔)

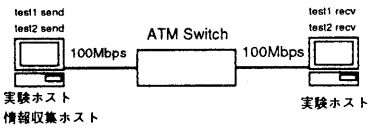

図 10: ATM LAN での測定

表 2: 測定環境とパラメータ

CPU	SPARCstation10(SUN)
OS	SunOS 4.1.3
ATM NIC	ASX-100 (FORE)
ATM 交換機	ATOMIS 5 (NEC)
TCP バッファサイズ	52428byte
セグメントサイズ	4096byte
データ伝送時刻	test1: 0.00 秒
	test2: 1.00 秒
送信回数	test1: 5000 回
	test2: 2000 回

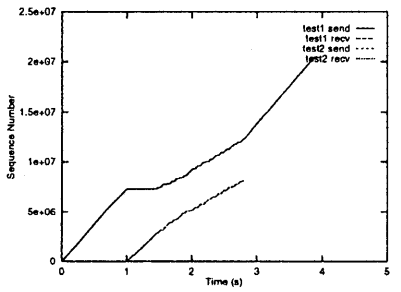

図 11: 各セグメントの送受信時刻

関係があるとも考えられる。

今後は複数のホストを ATM LAN で結んでセル落ちが発生するような環境を作り性能測定を行う予定である。

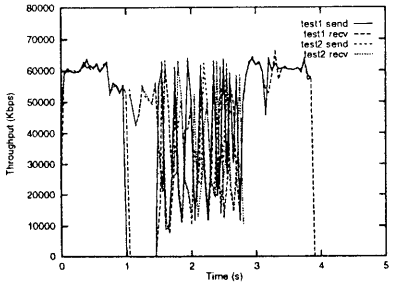

図 12: スループットの変化 (0.05 秒間隔)

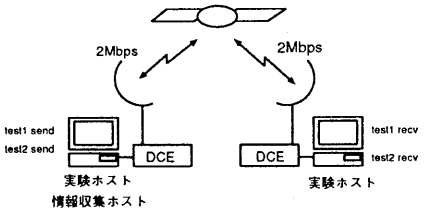

図 13: 衛星回線での測定

表 3: 測定環境とパラメータ

CPU	SPARCstation10(SUN)
OS	SunOS 4.1.3
DCE	D1230A IDU (NEC)
TCP バッファサイズ	52428byte
セグメントサイズ	1024byte
データ伝送時刻	test1: 0.00 秒
	test2: 2.00 秒
送信総バイト数	test1: 300 回
	test2: 300 回

6.3 衛星回線

衛星回線を使用して測定した例を示す。図 13のような環境で測定を行った。使用したコンピュータおよび OS を表 3に示す。衛星回線には約 0.5 秒の遅延時間がある。このため，ウィンドウサイズが最大の 64kbyte になったとしても最大 1Mbps 以下のスループットしか得られない[1]。この結果ではウインドウサイズいっぱいのデータを送信した後，ACK を待っている時間の方が長いのが分かる。

今後は 2 つのホストをルータとして使用しそのルータで輻輳が生じるような状況を作って測定することを考えている。

[1]SunOS 4.1.3 ではウィンドウサイズの上限は 52428byte なので最大 0.8Mbps。

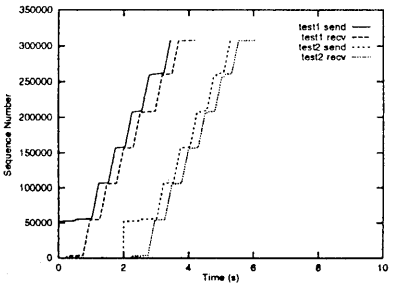

図 14: 各セグメントの送受信時刻

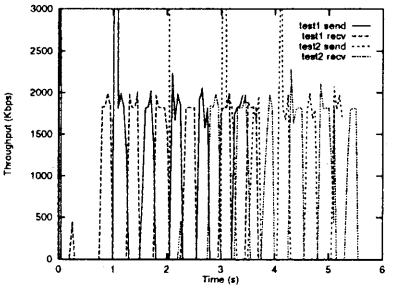

図 15: スループットの変化 (0.05 秒間隔)

7 今後の課題

本稿では TCP 性能評価システム DBS の提案それによる具体的な測定例を示した。しかし，課題も残されている。

1 番目の課題は，時刻の精度に関することである。現在の実装では時刻取得の精度は 10ms 程度しかない。そのため，10ms より細かいパケット送受信時刻を調査することはできない。高速なネットワークや，微小時間での分析には時刻の精度が問題になる可能性がある。この精度で問題があるか，また，改善できるかについて検討しなければならない。

2 番目の課題は，プログラムのコーディングに関する問題である。プログラムのコードは少なからず測定結果に影響を与えていると考えられる。測定結果の妥当性を示すためにはその程度を調査する必要がある。影響が無視できないならば，影響を小さくする方法や，データ処理によって影響を除去することを検討する必要がある。

3 番目の課題は，DBS が測定できることの網羅度である。DBS が TCP の全機能を本当に測定できるのか，あるいは，測定できないことは何なのかを明らかにしなければならない。現時点で分かっているのは，TCP のコネクションの確立や開放のオーバーヘッドを測定できないことである。DBS がカバーしていない項目を検討し，それを測定できるように DBS を拡張する必要があるだろう。

4 番目の課題は，アプリケーションレベルでの測定だけでは異常の原因を突き止めるのに限界があることである。例えばパケットの再送が起きたかどうかを知ることはできない。これを補うためには次のような方法が考えられる。

- ネットワークアナライザなどを利用して伝送中のパケットを調べる。
- OS 内部の TCP の処理部のパラメータを記録する。

これらの方法を組み合わせることにより，TCP のウィンドウサイズやパケット再送と，スループットの関係についてより詳しく分析できるようになる。その結果，TCP の問題点をより深く調査できるようになると考えている。

8 まとめ

2 点間接続モデルの性能評価の限界を示し，多点間接続での TCP 性能評価システム DBS(Distributed Benchmark System) を提案した。また，DBS を用いて Ethernet，ATM，衛星回線による測定を行い，今までのベンチマークでは測定できなかった TCP の挙動を測定した例を示した。

今後は，DBS を利用してさまざまなネットワークにおける TCP の性能評価を行い，より現実に即した TCP の性能を明らかにする。そして現在の TCP の拡張だけで性能を向上できるか検討する。その結果から導き出された結論もとにして次世代の TCP を考案し，実装を行い，DBS で評価を行う予定である。

参考文献

[1] A. McKenzie. “A Problem with the TCP Big Window Option” RFC 1110, August 1989

[2] Douglas E. Comer and John C. Lin, “TCP Buffering And Performance Over An ATM Network” Purdue Technical Report CSD-TR 94-26, 1994

[3] R. Fox. “TCP Big Window and Nak Options”, RFC 1106, June 1989.

[4] T. Luckenbach, R. Ruppelt, F. Schulz, “Performance Experiments within Local ATM Networks”, GMD-FOKUS(Berlin, D)