

解説

計算機ベンチマークの最新動向



2. X ウィンドウシステムのベンチマークテスト†

竹内陽一郎†

1. はじめに

ウィンドウシステムは、現在のコンピュータシステムで、最も広く使用されている代表的なヒューマンインタフェースである。現在のコンピュータの大半は、ディスプレイ上に表示される図形や絵文字を、マウスなどのデバイスで指し示したり動かしたりすることによりコンピュータとの対話を行っていく。ディスプレイへの図形の表示と入力デバイスの処理を担当するウィンドウシステムの処理性能が、システム全体の操作の快適性に大きく影響を与えるため、その適切な性能評価が重要であることはいうまでもない。さらに、ウィンドウシステムの性能評価は、今後発達していくと思われる、より高度なヒューマンインタフェースに対して、以下の観点から重要であると考えられる。第1に3次元CG、動画、アニメーション、仮想現実などの分野でも、2次元図形処理の比重はきわめて高くこの部分の評価にはそのまま適用できるという点。第2に、ネットワークを介した分散処理が今後ますます重要な役割を果たすと思われるが、その原型となるサーバ/クライアントアーキテクチャが考慮されているという点。第3に、性能評価方法の標準化が進んでおり、その標準化手法が今後の標準化にとって参考になるだろうという点においてである。

本稿では、ワークステーション分野での代表的なウィンドウシステムであり、また、アーキテクチャが整理されており、性能評価方法も確立されていることから、X ウィンドウシステムを主として取りあげ、ベンチマークテストの実例とともにその性能評価方法の解説を行う。

2. X ウィンドウシステムの概要

システム性能評価技術にとって、その基盤となるシステムアーキテクチャは、非常に重要な要素である。ここでは、まずX ウィンドウシステムのアーキテクチャについて、性能評価の観点から重要であると思われる点について取りあげる。

2.1 サーバ/クライアントアーキテクチャ

X ウィンドウシステムの開発者たちは、システムの開発にあたって、アプリケーションプログラムを入出力装置と独立に実行できるようにすることを最重要項目として取りあげた。これによって、リモートログインしているマシンで実行したアプリケーションの表示および入力を、自分が現在操作しているマシンで行うことが可能となり、これが現在のネットワークコンピューティングの基礎を築いたといっている。これを実現するため、X ウィンドウシステムは、表示や入出力を行う部分をサーバとし、アプリケーションから呼び出されるプリミティブをクライアントとする、サーバ/クライアントアーキテクチャを採用している。

実際にビットマップへの描画処理を行うグラフィック処理部分や、キーボードやマウスからの入力を処理するサブシステムは、X サーバと呼ばれ、アプリケーションとは独立に動作する。アプリケーションと、X サーバの間はソケットで結ばれ、X プロトコルと呼ばれる手順で、コマンドやデータの交換を行う。X サーバとの通信を行う最下層のアプリケーションプログラムインタフェースはXlibと呼ばれ、アプリケーションプログラムは、最終的にここを呼び出すことによって、X サーバの提供するサービスを受けることができる。

† Bench Mark Tests for X Window System by Yoichiro TAKEUCHI (Information & Communication Systems Laboratory, TOSHIBA CORPORATION).

† (株)東芝情報・通信システム技術研究所

2.2 X プロトコルと非同期処理

サーバ/クライアントアーキテクチャの欠点は、サーバ/クライアント間の通信によってオーバーヘッドが生じ、これが頻繁に発生すると性能が著しく低下することにある。X プロトコルでは、この欠点を解消する手法をいくつか採用している。

(1) 一方向プロトコル

画面表示処理では、いったん表示要求をだしてしまえば、エラーが起きない限り、サーバから返答をもらう必要はない。このため、X プロトコルでは、大半のプロトコルで、要求をだすだけで確認を行わない、一方向プロトコルが採用されている。クライアント側は、サーバに要求をだしてしまえば、次の作業に移れるため、並列処理という面でも都合がいい。エラーが発生した場合、ステータスは、エラーイベントとして非同期にクライアント側に返される。

(2) グラフィックコンテキスト

個別の描画コマンドに対して、色や、パターン、背景など、レンダリングのための細かな情報を毎回要求時に指定すると、パラメータが多くなり、通信のオーバーヘッドが増大する。X ウィンドウシステムでは、これらの情報はグラフィックコンテキスト (GC) としてあらかじめ作成し、サーバに登録しておく。描画プロトコルでは、GC を示して要求をだすため、通信の負荷が軽くなる。

(3) 連続描画インタフェース

点や線を連続して描画する処理では、点や線 1 つずつに毎回要求をだすと、クライアントサーバ間の通信量が大きくなるため効率がよくない。X プロトコルではこれを避けるため、1 つの要求で、複数の線や点の描画ができる手順を設けている。

2.3 パイプライン処理

GC や連続描画要求などの X ウィンドウシステムのアーキテクチャは、サーバ/クライアント間の通信を減らすだけでなく、サーバ側のレンダリング処理をパイプライン化できるように考慮して設計されている。現在、大部分のワークステーションは、このような X ウィンドウシステムの設計思想を生かすよう設計されたパイプライン処理構造を持つ専用グラフィックハードウェアを備えている。

3. X ウィンドウシステムの性能評価技術

X ウィンドウシステムでは、アプリケーションプログラムは必ず Xlib を使用してサービスを受けることになるため、Xlib の個々の性能をおさえておけば、アプリケーション全体の性能を予測することは比較的容易である。しかし、前章で述べたように、X ウィンドウシステムはサーバ/クライアント方式を採用しており、また、その欠点である通信オーバーヘッドを削減する種々の工夫を行っているため、性能評価を行う場合、システムのアーキテクチャをよく考慮して、適切な評価を行わないと意味のある測定結果を得られない。

以下、Xlib の測定を行う上で、注意しなければならない点についてまとめる。

(1) 通信オーバーヘッド

アプリケーションと X サーバの間で、実際に LAN などを経した通信が行われるか、同一マシン上でのプロセス間通信で済むかにより測定結果は大きく異なることがある。基本的には、サーバとアプリケーションを同一マシンとして測定し、通信のオーバーヘッドをあとで補正の方が評価しやすい。

(2) サーバ/クライアント間の同期

多くの X プロトコルは一方方向であるため、対応する Xlib の手続きの前後で時間測定を行っても、サーバが実際に描画するのに必要な時間は求められない。これを行うには、描画したビットマップをもう一度クライアントに読み出すといったテクニックが必要になる。

(3) パイプライン処理の考慮

グラフィックハードウェアによっては、パイプライン処理を行っているため、GC の変更が頻繁に行われた場合の描画性能が著しく低下する場合がある。また、連続する点や線分の描画では、一般的に要素の数が多い方が一要素あたりの処理性能はあがる傾向にある。

4. X ウィンドウシステムのベンチマーク

本章では、X ウィンドウシステムでの代表的なベンチマークテストである、Xbench/xstone、X. 11 perf/Xmark について述べる。

4.1 Xbench について

(1) Xbench の概要

Xbench は、Siemens の Claus Gittinger が作成、1989年3月に公表されたベンチマークである。

X 11 R 3用として作成されたプログラムであるが、基本的な X プロトコルしか使用していないため、X 11 R 6でも、再コンパイルすることにより使用できる。

(2) Xbench の性能測定方法

Xbench は、X ウィンドウシステム的环境においてクライアントプログラムとして動作し、一定時間の間 X サーバに対するリクエストをだし続け、X サーバが単位時間あたりに処理できる X プロトコルの処理数を計算することにより X サーバの描画性能を測定する。クライアント側で動作する他のデーモンなどによる測定値への影響を最小限にするため、各項目の測定は数回繰り返して行われ、最も良い数値を採用する。

(3) Xbench の主なテスト項目

Xbench は、40 項目の測定を行う。以下に主なテスト項目をあげる。

●線分

縦、横、斜め方向について、破線、実線、幅、長さなどを変化させ描画時間を測定する。

●矩形領域

1 辺の長さを変化させながら、それぞれ、枠だけの描画、タイルによる塗りつぶし、ステップルパターン (stipple: 点描) による塗りつぶし、および、矩形領域の反転速度について測定を行う。

●円弧および扇形

5~360 度までの円弧および、塗りつぶした扇形について測定を行う。

●多角形

5 角形の塗りつぶし速度を測定する。

●領域のコピー

1 辺の長さを変化させながら、それぞれの正方形領域のコピー時間を測定する。

●ビットマップコピー

●スクロール

●文字列

固定サイズフォントの描画速度を測定する。

●複合テスト

ウィンドウの作成、ウィンドウのマップ、GC の作成、領域の消去、文字列描画、スクロール、ウィンドウの破壊などを組み合わせた総合性能測定を行う。

(4) xstone 値による評価

Xbench の測定結果は、ウェイトづけがなされ、いくつかの性能指標にまとめられる。

●lineStones

線描画の性能指標

●fillStones

塗りつぶしの性能指標

●blitStones

ブロックイメージ転送の性能指標

●arcStones

円弧描画の性能指標

●textStones

文字列描画性能指標

●complexStones

複合テスト性能指標

●xstones

総合性能指標

総合性能の xstone 値は、リファレンス値として、サンマイクロシステム社のワークステーション、Sun 3/50 の X 11 R 3 モノクロ X サーバの測定結果を用い、その値を 10,000 としてノーマライズされる。

xstone 値を求めるときのウェイトづけは、初期の X ウィンドウシステムの使用状況を反映して、テキスト処理とスクロール処理のウェイトが他に比べて非常に大きくなっている。このため、グラフィック描画処理のウェイトの大きい現在の使用状況では、xstone 値はやや陳腐化した感がある。

4.2 X. 11 perf について

(1) X. 11 perf の概要

X. 11 perf は、DEC の Joel McCormack らによって作成された X サーバのベンチマークプログラムである。

X. 11 perf は、v 1.2 が X 11 R 4 に収録されたのが最初で、X 11 R 6 では最新版の v 1.5 が収録されている。

測定項目のレパートリーは、Xbench と同様であるが、測定項目が多く (現状 450 項目) かつ詳細で、ほぼ X ウィンドウシステムの全機能をカ

バーしているといつてよい。

(2) X. 11 perf の性能測定方法

X. 11 perf では、さまざまな速度の X サーバ/クライアントの環境に対応するため、各テスト項目の測定にはいる前に、サーバ/クライアント間のマルチプロトコルの往復時間を測定して、これを補正值として使用する。このため、サーバ/クライアント間の通信系のオーバーヘッドには独立な測定結果が得られるようになっている。

各テスト項目は最初に仮実行して実行時間の見積もりを行い、X サーバの性能にかかわらず測定時間がほぼ同じ値になるように測定回数を決める。(デフォルトは5秒)したがって、性能が速すぎることによって、測定誤差が大きくなり測定結果の信頼性が損なわれるようなことはない。

測定はデフォルトで5回行いその平均値が採用される。(Xbench が最良値を採用するのは対照的)

(3) X. 11 perf の主な測定項目

以下に、X. 11 perf の主な測定項目をあげる。

- 点 (Xbench には、点は含まれていない)
- 線分
- 矩形領域
- 円、円弧、扇形
- 多角形
- テキスト
- 領域コピー
- スクロール
- ウィンドウ作成、破壊、マップ、アンマップ
- グラフィックコンテキストの変更

X. 11 perf では、描画がパイプライン処理される場合なども想定しており、各測定項目はそれらを考慮した条件で何度か測定が繰り返される。

(4) Xmark 値による評価

X. 11 perf の測定項目は多岐にわたり、測定項目の一覧表だけでは、総合的な描画処理能力を比較することは難しい。そこで、X. 11 perf v1.3 から、Xmark という総合指標が用いられるようになった。

Xmark は xstone と同様、X. 11 perf の各測定項目の加重平均を特定の基準値にノーマライズした値である。基準値はサンマイクロシステム社のワークステーション SPARCstation 1 での測定値 2119 であり、これを Xmark の 1.0 として値

が求められる。

xstone では、テキスト処理に極端に偏ったウェイトづけがなされていたが、Xmark はこれと比べると、比較的均等なウェイトづけになっている。これは、現在のワークステーションでは、グラフィックユーザインタフェースの発達にともなう、テキスト処理以外のグラフィック処理の比重が相対的に高まってきていることを反映しているといえる。

4.3 X. 11 perf による性能予測

X. 11 perf の測定項目は、ほぼ X のプロトコルを網羅しているため、アプリケーションごとに固有の重みづけを求めて、Xmark 値の計算と同様にそのアプリケーション固有の性能指標を求めることができる。

X ウィンドウシステムでは、使用した X プロトコルの統計をとる機能が備わっているため、これを用いて、アプリケーションごとに正確な重みづけが求められる。さらに、X. 11 perf の測定値は、通信オーバーヘッドが補正されているため、実際の通信オーバーヘッドを測定し、X. 11 perf の測定値を加重平均した値に加算すれば、マシンやシステムの構成が変わったときのアプリケーション性能を実際に測定することなく、かなり正確に実行時間を見積もることができる。

グラフィックユーザインタフェースを含むアプリケーションは性能分析が難しく、また、他のマシンへの移植には時間がかかることが多いため、この X. 11 perf ベンチマークを用いた性能見積もりは非常に有効である。

5. 3次元グラフィックのベンチマーク

2次元の基本的な描画性能は、X. 11 perf など詳細に測定することができるが、3次元では、さらに、座標変換や隠面処理、レイトレーシングやシェーディングなどの特有の処理が必要となるため、これらを含めたベンチマークテストが必要となる。この分野では、標準化団体の米 GPC (Graphic Performance Characterization) 委員会がベンチマークの標準化を行っている。

5.1 GPC PLB

PLB (Picture Level Benchmark) は、GPC 委員会による代表的なベンチマークテストである。GPC PLB は、X. 11 perf などのように、描

画プリミティブの性能を測定して加重平均するアプローチとは異なり、描画すべき絵を標準化し、その描画時間の性能を比較するミニアプリケーション型のベンチマークテストである。

テスト項目として、10枚の絵が用意されており、各測定値ごとに、PHIGSなどの標準ライブラリを用いて測定した通常の値と、各ベンダごとに最適化した値とを公表できる。ベンダごとの最適化値の測定では、ベンダ固有のライブラリの使用が認められ、さらに、一定のルール内でのチューニングが許されている。各測定結果は、2次元、3次元のワイヤフレーム、3次元のソリッドモデリングの3つの分類に従って加重平均がとられ、それぞれ、PLB 2d (2次元)、PLBwire (ワイヤモデリング)、PLBsurf (ソリッドモデリング) という指標にまとめられる。

5.2 GPC PLBの測定項目

GPC PLBの10個の測定ファイルは、CG分野やCAD分野の代表的なアプリケーションに対応するように定められている。各ファイルの概要を以下に示す。

(1) 2次元

- pc __board
電子回路のCAD図面
- floor __plan
オフィスのフロア見取り図
- site __plan
建物と道路の配置プラン

(2) 3次元ワイヤフレーム

- sys __chasis
コンピュータのシャーシ設計図
- race __car
レーシングカーアニメーション
- seafloor
海底の鳥瞰図

(3) 3次元ソリッドモデリング

- cyl __head
自動車エンジンのシリンダヘッド
- head
人間の頭部
- shuttle
スペースシャトルのアニメーション
- studio
スタジオ内見学の3次元CG

5.3 GPC委員会での今後の標準化動向

GPCのPLBはPHIGSなどの標準ライブラリを対象としてきた。一方、3次元描画ライブラリとしては、シリコングラフィクス社が開発したOpenGLが近年、急速に普及しており、3次元グラフィック処理プリミティブの事実上の標準になりつつある。このため、OpenGLに基づいた標準ベンチマークテストの制定が強く望まれている。このような背景から、GPC委員会でもOpenGLに基づくベンチマークテストの標準化に現在着手している。

6. マルチメディア関連のベンチマーク動向

本章では、最近特に注目されている、マルチメディアに関連したベンチマークの動向に触れる。

(1) PCとMS-Windows

マルチメディア用の機器として、最近パーソナルコンピュータの画面表示性能評価の重要性が高まっている。通常、パーソナルコンピュータでは、ウィンドウシステムとOSが一体化しており、画面表示性能はOS全体の性能評価の一部として評価されることが多い。パーソナルコンピュータの代表的なOSである、マイクロソフト社のMS-Windowsを対象とするベンチマークテストを例として取りあげると、代表的なベンチマークとして以下のものがあげられる。

- WinTach
- WinSock
- WinStone
- SYSmark

このうち、WinTachはWindowsのグラフィックス性能全般を対象としており、X. 11 perf/Xmarkと似た性格のベンチマークといえる。WinStoneおよびSYSmarkは、ウィンドウシステムの性能評価というよりは、PC全体にわたる総合的なベンチマークで、PC上で使用頻度の高いアプリケーションプログラムを実際に実行して、その時間からPCの性能を比較する。どちらかといえばSPECベンチマークに性格に近い。ただし、PC上のアプリケーションがテスト測定対象であるため、実際にはウィンドウ処理性能の比重がかなり高いベンチマークといえる。

(2) 動画処理性能

動画処理では、MPEGが国際標準として確立

されている。動画の場合、静止画とは異なり、どのくらい速く描画できるかという尺度ではなく、どの規模の映像まで正常に再生できるかという尺度で評価する必要がある。この動画の規模について、MPEG 2では、プロファイルとレベルという概念で標準化を行っている。たとえば、MPEG 2映像として一般的なメインプロファイル、メインレベル (MP@ML) では、画面サイズ 720×576、毎秒 30 コマ、最大転送レート 15 Mbps などのように映像の満たすべき条件が具体的に定められている。このため、MP@ML の規格を満たす性能があるかどうかといったコンプライアンステストがベンチマークテストにかわって重要になると予想される。

また、性能とならんで画質が非常に重要になることから、これを評価するための標準映像の制定も重要な課題である。

(3) インターネット

WWW サーバ/クライアントシステムのベンチマークをシリコングラフィックス社が公表するなど、この分野でもベンチマークの標準化が模索され始めている。また、インターネット上でアニメーションや、仮想現実を実現するための技術として、サンマイクロシステム社の Java やシリコングラフィックス社の VRML などのスクリプト言語が注目されてきており、今後これらの処理系に関連した性能評価の重要性が増すことが予想される。

7. おわりに

本稿では、現在のコンピュータシステムの代表的なヒューマンインタフェースである X ウィンドウシステムを取りあげ、その性能評価方法と、代表的なベンチマークテストについて述べてきた。

今日、コンピュータとネットワーク技術の進歩によって、ヒューマンインタフェース技術はますます高度化、複雑化する傾向にある。これにともなって、複雑化するシステムを解きほぐし、適切な性能評価を行う必要性もますます増大すると思われる。このため、この分野でのシステム性能評価の標準化は今後いっそう重要な課題となるであろう。

参 考 文 献

- 1) Adrian, N.: Xlib Programming Manual, O'Reilly & Associates, Inc. (1990).
- 2) Oliver, J.: INTRODUCTION TO THE X WINDOW SYSTEM, Prentice-Hall, Inc. (1989).

(平成 7 年 11 月 28 日受付)



竹内陽一郎 (正会員)

1960 年生。1985 年京都大学大学院理学研究科物理第二専攻修士課程修了。1986 年(株)東芝入社。現在、情報・通信システム技術研究所にて、マルチメディア関連の研究開発に従事。