社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－EVA－17 (4)
2006／6／20

仮想計算機環境における資源管理オーバヘッドの評価

網代 育大 † 田中 淳裕 †

† NEC システムプラットフォーム研究所

VMware ESX Server の仮想マシン上に構築した .NET アプリケーションを使って、仮想化に伴う CPU とディスク資源の管理オーバヘッドを評価した。その結果、仮想マシンの性能が仮想マシンの稼働数や CPU 使用率の合計値に依存することがわかった。また、ディスクに関しては、ある仮想マシンのディスクに高負荷がかかることによって他の仮想マシンのディスク性能も劣化する可能性の高いことがわかった。本論文では、ESX Server に対する評価の詳細と、運用上の留意点について述べる。

Measuring Resource Management Overhead for Server Virtualization

Yasuhiro AJIRO † Atsuhiro TANAKA †

† System Platforms Research Laboratories, NEC Corp.

We have measured virtualization overhead of CPU and disk resources with a .NET application deployed in VMs of VMware ESX Server. The result of measurement shows that CPU performance of VMs depends on both the number of VMs and the total CPU utilization. For disk performance, the result implies that heavy load in a VM may degrade another VM. This paper describes the details of the measurement, and discusses how to manage ESX Server performance.

1 はじめに

企業内の PC サーバは、これまで業務や組織の変化に応じて場当り的に導入されてきたため、企業は計算資源の有効活用されていない多数のサーバを保有しているケースが多い。特に大企業ではサーバ数が数百台規模に達しており、管理コストだけでなく、フロア代などの設置コストの増大を招いている。これを受けて近年、無秩序に増え続けたこれらのサーバを少数の高性能サーバに集約するサーバ統合の動きが盛んである。

統合対象となる 4 年以上昔に導入されたサーバでは、サポートの終了した Windows NT や Windows 2000 Server とともに、独自開発のアプリケーションや正体不明のプログラムが動作しており、新しい OS やミドルウェアを使って一から作り直すには多くの投資が必要となる。そこで、VMware のような仮想計算機環境を利用して、統合前のサーバ環境をそのまま新しいサーバの仮想マシン上に移行する統合手法が大きな注目を集めている。

PC サーバの仮想化には、多くの場合、現在のデファクトである VMware ESX Server (以下、ESX) が用いられるが、ESX の性能には未知な点が多く、統合計画時の性能設計や統合後の性能管理に関する明確な指針が存在しなかった。そこで本研究では、仮想マシン上のサンプルアプリケーションを用いて実験を行ない、

- 仮想マシン (VM) の稼働数と性能との関係や、仮想化していない物理マシンとの性能比、
- ならびに仮想マシン上のゲスト OS で計測した資源使用率に関する性能上の閾値

について調査を行なった。本稿では、実験内容とその結果を示し、ESX を用いた性能設計、運用管理のための指針について論じる。

2 実験環境

2.1 プラットフォーム

まず、評価用の物理マシンとして、サーバ統合対象のマシンとスペックの近いデスクトップマシン (PentiumIII 1.0GHz, 512MB of memory, ATA100 HDD

x1) と、Dell PowerEdge 1850 (Dual Core Xeon 2.8GHz, 2GB of memory, U320 SCSI HDD x1) を用意した。PowerEdge 1850 は、OS 起動時のパラメータを使ってメインメモリを後述の仮想マシンと同じ 512MB に縮退させた。PowerEdge 1850 は複数台用意し、ドメインサーバや負荷発生のためのクライアントとしても利用した。また、計測用マシンとして、Pentium4 2.0GHz を搭載したデスクトップマシンを用意し、Windows Server に標準添付のシステムモニタを使って、評価対象マシンの計測データをオンライン収集した。

仮想マシンの評価には Dell PowerEdge 2850 (Dual Core Xeon 2.8GHz SMP, 8GB of memory, U320 SCSI HDD x6, PERC4e/Di RAID controller) 1 台を用意し、VMware ESX Server 2.5.2 をインストールした。デュアルコア CPU を 2 基搭載し、かつ Hyper-Threading を有効にしたため、ESX からは 4 つの物理 CPU と 8 つの論理 CPU が認識される。ただし、Virtual SMP モジュールは利用せず、各仮想マシンが同時に利用できる CPU は 1 つにした。一方、HDD は、実運用環境に近づけるため RAID1 (ミラーリング) で構成し、3 つの論理ボリュームとして利用した。このうちの 1 ボリュームを ESX が占有し、残りの 2 ボリューム上に仮想マシンをそれぞれ 6 個ずつ、計 12 個作成した。各仮想マシンのメインメモリは 512MB に統一した。また ESX では仮想マシンの物理 CPU 使用量に関する上限/下限を設定できるが、今回は特に設定しなかった。

これらの物理マシンやデスクトップマシンは、すべて単一のギガビットイーサネットスイッチに接続した。アプリケーションプラットフォームとしては、Windows Server 2003 (Enterprise Edition SP1) と IIS 6.0、SQL Server 2000 (Enterprise Edition SP4) を選択した。その上に Microsoft 社がリファレンスアプリケーションとして公開している Pet Shop 3.0 [5] を配備し、計測と分析を行なった。

2.2 Pet Shop の特徴

サンプルの Pet Shop は、C# .NET を使ってオンライン上のペット購入サイトを実現したものである。デフォルトで購入できるペットは全部で 28 種類が用意されており、鳥、犬、猫、爬虫類の分類や、それぞれにつき 3 種類程度の品種、雌雄等の区別がある。会員情報やペットの詳細、購入履歴、在庫情報等のデータは DB (SQL Server) 側で管理され、IIS/ASP.NET 上の Pet Shop アプリケーションが DB のデータを適宜参照/更新しながら、ユーザと HTTP によるインタラクションを行なう。

サイトを訪れたユーザは、ペット情報の閲覧や、会員としてログインした後にペットの購入を行なうことができる。ペットの種類が少なく、データがほぼメモリ上に展開されるために、DB の更新時以外はディスクアクセスがほとんど発生しない。しかし、ペットの種類を大きく増やすと、Web サイトや DB テーブルのデザインを大幅に設計し直さなければならなくなるため、今回はデフォルトのままで実験を行なった。

2.3 負荷発生クライアント

配備した Pet Shop サイトに安定した負荷をかけるため、ユーザの HTTP 通信をエミュレートする負荷発生クライアントプログラムを開発した。開発環境は Linux 2.4.31, gcc 3.3.2 であり、以下の機能を提供する。

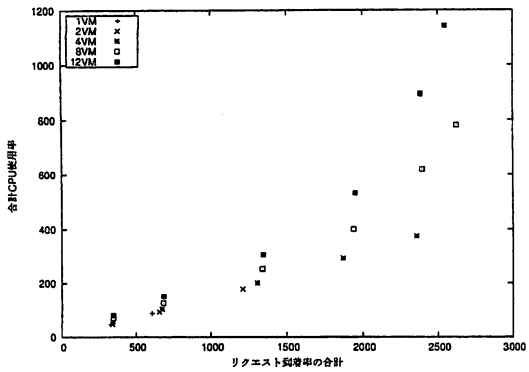
- マルチスレッドによるサイトアクセス。各スレッドは独立にペット情報の閲覧や、購入を行なう。
- ポワソン到着するリクエストの発行。各スレッドは、リクエストのレスポンスを受け取ってから次のリクエストを発行するまでの間に、指数乱数で決めた秒数だけ思考時間 (think time) をはさむ。
- 各リクエストの応答時間の計測と、平均応答時間、スループットの算出。

ペットを閲覧する場合、各スレッドは、トップページへの来訪、犬や猫等の分類の選択、品種の選択、雌雄等の個体の選択の 4 トランザクションを繰返し行なう。購入の場合では、閲覧時と同様のトランザクションのほかに、ログインやショッピングカートへの追加、住所の確認等を含む 14 トランザクションを繰り返す。今回の実験では、HTML ファイルのやりとりのみを行ない、画像の取得は省略した。

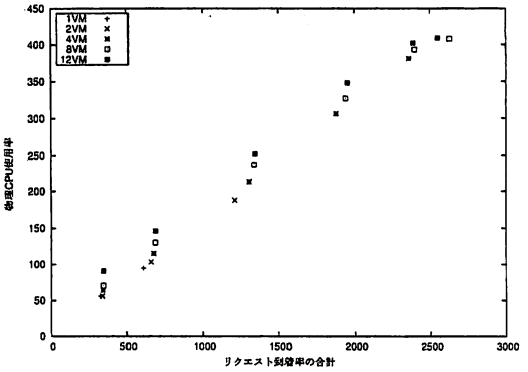
3 実験と評価

3.1 手順

トランザクション中にペットの閲覧比率が高い場合、Pet Shop アプリケーションのボトルネックは、IIS が動作するマシンの CPU になる。逆にペットの購入比率が高い場合は、SQL Server マシンのディ



(a) 各仮想マシンの CPU 使用率の合計値



(b) ESX 上の CPU 使用率

図 1: 仮想マシンと ESX の CPU 使用率

スクがボトルネックとなる。そこで、CPU オーバヘッドの評価は、ペットの閲覧のみを行なうリクエストが集中したときの IIS マシンに対して行ない、ディスクオーバヘッドの評価は、購入のみのリクエストが集中したときの SQL Server マシンに対して行なった。

物理マシンの評価では、2 台の物理マシンにそれぞれ IIS と SQL Server をインストールし、その上に Pet Shop を構築した。そして、別マシン上の負荷発生クライアントから IIS 上の Pet Shop に対してリクエストを発行し、クライアントのスレッド数 (負荷) を徐々に上げながら性能の計測を行なった。サーバ側の計測項目は、CPU 使用率、ディスク使用率、メモリ使用量、ネットワーク転送量、IIS や SQL Server のトランザクション数である。メモリ使用量とネットワーク転送量については、実験中、これらのハードウェアがボトルネックになっていないことの確認に用いた。サンプリング間隔は 30 秒に設定し、計測用マシンのシステムモニタを使って計測データをオンラインで収集した。計測の際は負荷を 5 分以上かけ続け、中間の 3 分間におけるサンプリングデータ 6 個分の平均値を算出した。

3.2 CPU オーバヘッドの評価

仮想マシン数に対する CPU のオーバヘッドを見積もるため、ESX 上に 1-12 個の仮想マシンを用意した上で IIS を起動した。各 IIS に対応する SQL Server を ESX マシン (PowerEdge 2850) とは別の PowerEdge 1850 上に起動し、独立の Pet Shop サイトを複数配備した。その上で、各サイト (IIS) に対して負荷発生クライアントから均等にリクエストを発行し、IIS へのリクエスト到着率 (毎秒の平均トランザクション数) の合計値と、CPU 使用率の合計値との関係を調べた。図 1 (a) は、この結果をプロットしたグラフである。また、同時に計測した ESX 上の CPU 使用率を同図 (b) に示す。各仮想マシン上の Windows Server は、自身に割り当てられた CPU 資源量を 100% として使用率を算出するため、(a) の合計 CPU 使用率は、12VM であれば 1200% 付近まで上昇する。一方、(b) における ESX 上の CPU 使用率は 4 つの物理 CPU に関する合計値が計測されるため、400% までの値をとる。

負荷発生クライアントにおける各スレッドの平均思考時間は 100 msec に設定した。また、スレッド数は、Pet Shop サイトにエラーが発生するか、評価対象マシンの CPU またはディスク使用率が 95% を超え、スループットが頭打ちになるまで 100, 200, 400, … と増やし続けた。このとき、スレッドを各仮想マシンに均等に割りふることで、仮想マシンの負荷をほぼ均等にした。スレッド数を均等に分割できない場合は、一部の仮想マシンに対するスレッド数を 1 増やして調整した。たとえば、12 個の仮想マシンに 200 スレッドから負荷をかける場合は、4 つの仮想マシンには 16 スレッド、残りの 8 つの仮想マシンには 17 スレッドによるリクエストを発行した。

ペットの閲覧しか行なわない場合、SQL Server マシンにはほとんど負荷がかからないため、各 IIS からの SQL リクエストは、SQL Server の 1 インスタンスがすべて処理することにした。このような構

成であっても、実験を通じて SQL Server マシンの CPU とディスクの使用率は常に 0–1% であった。これは、.NET Framework のキャッシュ機構により、DB の検索結果が IIS/ASP.NET 側で保持されるためと考えられる。また、Pet Shop にエラーが発生していない状況では、クライアント側の平均応答時間は 177–284 msec の範囲内であった。

図 1 の結果は、以下の事実を示している。

- 仮想マシン数が多いほど、使用率が高く、オーバヘッドが大きい。
- ESX 上の CPU 使用率 (b) は、リクエスト到着率の増加に応じて線形に増加する。
- 仮想マシン数が物理 CPU 数以下の場合、合計 CPU 使用率 (a) も線形に増加する。物理 CPU 数より多くの仮想マシンを稼働した場合、合計 CPU 使用率は 400% 付近まで線形に増加し、仮想マシン数による差も小さいが、400% を超えた時点から急激に上昇する。

この 3 つめの事実を確認するため、仮想マシン数と負荷に対する CPU 処理性能を調べた。ここで、CPU 処理性能 μ は、リクエスト到着率 λ、CPU 使用率 ρ に対して、$\mu = \lambda/\rho$ で計算される値であり、単位時間あたりに処理可能な最大リクエスト数を表す。リクエスト到着率と CPU 使用率が線形の関係にある場合は、どの計測データで計算しても、CPU 処理性能は等しい値となる。結果を表 1 に示す。表には、図 1 (a) の元になっている λ や ρ の合計値を記しているが、仮想マシンあたりの値は、これらを仮想マシン数で割ったものである。また、2 つの物理マシン、PentiumIII (Pen3) マシンと PowerEdge 1850 (Xeon) の処理性能も同時に算出し、Pen3 の CPU 処理性能を 1 としたときの性能比を表の 1 番右側に併記した。結果は省略するが、物理マシンにおける λ と ρ の関係はほぼ線形であり、算出には ρ が 50% 付近にあるデータを用いた。

VM 数が 1 や 2 のときにスレッド数が 200 や 400 までのデータしか計測できていないのは、ESX 上の CPU 使用率は低くても、仮想マシンの CPU 使用率が 100% 近くまで上昇し、Pet Shop にエラーが発生するためである。つまり、仮想マシン上のサーバを運用する上では、ESX 上の使用率だけでなく、仮想マシンごとの使用率も管理しなくてはならない。また、VM 数が同数の場合、合計 CPU 使用率が 400% 以下の範囲では、CPU 処理性能や相対性能に大きな差異は見られない。この範囲における相対性能の平均値を求めた結果を以下に示す。表中の相対性能値は、以下の値から最大でも 6% しか外れていない。

表 1: 物理マシンと仮想マシンの CPU 性能

VM 数	スレッド数	Req. 到着率	CPU 使用率	CPU 処理性能	相対性能
(Pen3)	50	174	49.6	350	1
(Xeon)	400	1305	52.6	2483	7.09
1	100	329	46.4	708	2.02
	200	607	88.2	688	1.97
2	100	339	47.6	712	2.03
	200	657	94.2	698	1.99
	400	1211	179	678	1.94
4	100	343	57.5	597	1.71
	200	675	104	651	1.86
	400	1306	201	651	1.86
	600	1875	293	641	1.83
8	100	346	70.0	494	1.41
	200	685	127	537	1.53
	400	1339	253	530	1.51
	600	1939	401	483	1.38
	800	2395	620	386	1.10
12	100	346	80.8	428	1.22
	200	687	151	454	1.30
	400	1345	306	439	1.26
	600	1950	532	366	1.05
	800	2384	895	266	0.761

VM 数	1	2	4	6	12
相対性能	1.99	1.99	1.81	1.49	1.26

このようにして得られた VM 数 n と相対性能 p に対して回帰分析を行なったところ、$p = 2.09 - 0.0710 \cdot n$ となり、図 2 に示すように、うまくフィットすることがわかった。便宜上、VM 数 1 のときの CPU 性能を 1 と考えると、VM 数が 1 増えるごとに VM 性能が $0.0710/(2.09 - 0.0710) \approx 0.035$ ずつ劣化することになる。また、ESX の稼働する PowerEdge 2850 の物理サーバとしての CPU 処理性能は、SMP マシンであることから、PowerEdge 1850 の 2 倍程度の $7.09 \times 2 \approx 14.2$ と見積もられる。今回の実験では、VM 数が 1–4 のときの VM 性能が、物理サーバの処理性能を論理 CPU 数で割った $14.2/8 = 1.775$ より大きな値となった。サーバを物理マシンから仮想マシンへ移行/統合する際には、このような相対性能値が重要となる。

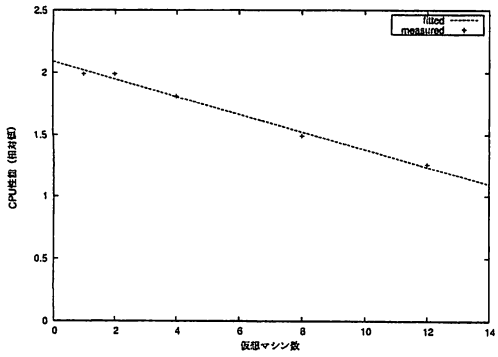


図 2: 仮想マシンの稼働数に対する CPU 性能

一方、合計 CPU 使用率が 400% を超える状況では、8 VM のときに 1.10、12 VM では 0.761 まで相対性能が悪化している。これは、合計 CPU 使用率が物理 CPU 数の 4 (= 400%) を超えると、CPU 使用率ならびに使用率に依存する応答性能が急激に劣化し始めることを意味する。稼働 VM 数を物理 CPU 数以下にすれば、このような状況を回避できるが、前述の劣化率を考慮し、かつ合計 CPU 使用率が物理 CPU 数を超えないように管理することで、CPU 性能の劣化を最小限に抑えることができる。

3.3 ディスクオーバヘッドの評価

ディスクオーバヘッドの評価には、仮想マシン上に SQL Server と Pet Shop 用のデータベースを用意し、各 SQL Server とインタラクションを行なう IIS を PowerEdge 1850 上に設定した。IIS サーバは 1 つだけ稼動し、仮想マシン数分の Pet Shop サイトを独立に配備した。その上で、前節と同様、各 IIS に均等に負荷をかけ、SQL Server へのリクエスト到着率の合計値と、ディスク使用率の合計値との関係を調べた。ディスクに関しては、VM 数 6 のときの計測も行なった。結果を図 3 (a) に示す。また、同図 (b) は、(a) における合計 CPU 使用率を各 VM 数で割った平均ディスク使用率のグラフである。グラフ (b) では、PentiumIII マシンの ATA ディスクや PowerEdge 1850 の SCSI ディスクに関する結果もあわせてプロットしている。しかし、ESX 上のディスク使用率は、計測手段が提供されていないために計測できなかった。実験中、IIS マシンの CPU とディスク使用率はぞれぞれ 73%, 3% 以下であり、クライアント側の応答時間は 280 msec 以下であった。

グラフにおいて、1-6 VM のデータは同一の論理ボリュームに配置された VM に関する計測値であり、8, 12 VM のデータは 2 つの論理ボリューム中の VM をそれぞれ 4, 6 個ずつ稼働した場合の値である。同程度の負荷に対する 4, 8 VM 時の合計ディスク使用率はほぼ等しく、8 VM 時は倍のリクエストを処理できていることから、ディスク (論理ボリューム) に対する負荷分散に成功していることがわかる。これは 6, 12 VM のデータにおいても同様である。また、内蔵の ATA ディスク (Pen3) に対する使用率はリクエスト到着率 80 付近で急上昇し、負荷に応じた使用率を計測できなかった。

さらに、図 3 からは以下の事実が読みとれる。

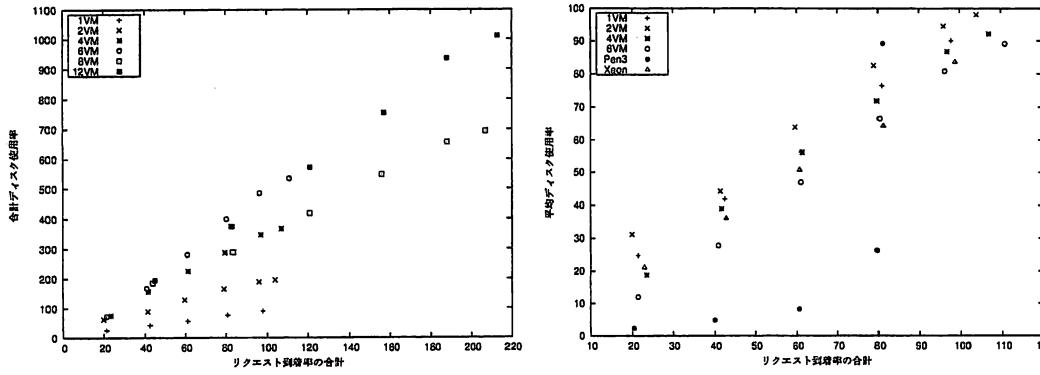
1. 1 つの論理ボリュームに配置された VM の稼働数を N 倍にすると、合計ディスク使用率はほぼ N 倍になる。

2. グラフ (b) を見ると、構成の似ている PowerEdge 1850 (Xeon) の平均ディスク使用率に対して、1, 2 VM のディスク使用率は増加しており、オーバヘッドの存在を確認できる。しかし 4, 6 VM では、ディスク使用率が 1, 2 VM のときよりも逆に減少する。

CPU 資源は、負荷が軽い (特に合計ディスク使用率が物理 CPU 数以下の) ときでも一定量が割り当てられるのに対して、ディスク資源は競争的に割り当てられる。つまり、N 個の VM がディスク資源を要求すると、各 VM には $1/N$ のディスク資源が割り当てられ、各 VM 上のゲスト OS は割当て分を 100%として使用率を算出するため、上記 1 の結果になると考えられる。また 2 に関しては、4, 6 VM 時にディスク性能が向上しているとは考えにくく、実際の負荷が見えづらくなっている可能性が高い。各 VM からのディスクアクセス要求は、ESX の VMM (Virtual Machine Monitor) が処理するが、VMM によるディスクアクセスがゲスト OS から正確に観測できていない可能性が考えられる。

4 関連研究

我々は、Web アプリケーションの観点から、仮想マシンのリクエスト処理性能や資源使用率について分析した。ESX Server の基本性能は VMware の white paper 等 [1, 4] で示されているが、仮想マシン上の、特に Windows Server におけるスループッ



(a) 合計ディスク使用率　　(b) 平均ディスク使用率

図 3: 各仮想マシンのディスク使用率の合計値と平均値

ト性能や資源使用率との関係については報告がなされていない。また、オープンソースの仮想計算機環境である Xen [2] では、ESX の VMM にあたる機能を Domain0 が担うが、文献 [3] は、Domain0 の I/O 処理に伴う CPU 使用量のうち、各 VM の利用分を算出する手法を提案している。

5 まとめと今後の課題

サンプルアプリケーションを用いた実験により、仮想マシン数が 1 のときの性能を基準として、稼働数が 1 増えるごとに CPU 性能が約 3.5% ずつ劣化することを確認した。ただしこれは、仮想マシンの合計 CPU 使用率が物理 CPU 数以下の場合であり、合計 CPU 使用率が物理 CPU 数を超えると、劣化の幅はさらに大きくなる。しかし、仮想マシンの稼働数に基づく性能の劣化率と、合計 CPU 使用率を適切に管理することで、物理 CPU 数より多くの仮想マシンを稼働しても (少なくとも CPU 性能に) 大きな問題がないことがわかった。

ディスクに関しては、VM 数が 4 以上になると使用率を正確に観測できないため、オーバヘッドの分析を行なうことができなかった。現行の ESX は、ディスク転送量を出力する esxtop 等のツールを提供しているが、ディスク使用率の計測手段は提供していない。Pet Shop のような Web アプリケーションは、転送量ではなく、アクセス回数に起因してディスク使用率が増加する。今回の実験では、VM 上の SQL Server に高負荷をかけた場合でも、ESX におけるディスク転送量はわずか 0.2 MB/sec であり、ディスクの負荷を正確に反映していなかった。ディスクオーバヘッドの評価や性能の管理には、ESX のレベルでディスク使用率を計測するための手段が必要であるが、これに関しては今後の課題である。

参考文献

[1] I. Ahmad, J. M. Anderson, A. M. Holler, R. Kambo, and V. Makhija. An Analysis of Disk Performance in VMware ESX Server Virtual Machines. In *Proc. IEEE Sixth Annual Workshop on Workload Characterization*, pages 65–76. IEEE, 2003.

[2] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield. Xen and the Art of Virtualization. In *Proc. 19th ACM Symp. on Operating Systems Principles (SOSP'03)*, pages 164–177. ACM, 2003.

[3] L. Cherkasova and R. Gardner. Measuring CPU Overhead for I/O Processing in the Xen Virtual Machine Monitor. In *Proc. 2005 USENIX Annual Technical Conference*, pages 10–15, 2005.

[4] VMware Inc. ESX Server Performance and Resource Management for CPU-Intensive Workloads. VMware White Paper, 2005. Available from `http://www.vmware.com/vmtn/resources/esx_resources.html`.

[5] G. Leake and J. Duff. Microsoft .NET Pet Shop 3.x: Design Patterns and Architecture of the .NET Pet Shop, 2003. Available from `http://msdn.microsoft.com/library/en-us/dnbda/html/petshop3x.asp`.