

外部からの測定による Web サイト性能限界 予測手法の開発

西岡大祐 川崎正人 瀬賀 聡
(株)ケイ・ジー・ティー
インターネットソリューション事業部

測定対象の Web サイトの特定ページの応答時間、ファイル総容量、送受信トラフィックデータ量、ネットワーク遅延時間および、ネットワーク帯域幅を外部から測定し、その結果を 1 層のブラックボックスシステムに抽象化したシミュレーションモデルのパラメータとして適用することで、Web サイトの応答時間とスループットを予測し、測定時の負荷に対する相対値として限界性能を算出する方法を開発した。これにより、クライアント側から見た Web サイトの性能を容易に推測することが可能となった。

Development of Predicting Web Site Performance by Outer Measurement Data

Daisuke Nishioka Masato Kawasaki Satoshi Sega
Internet Solution Division, KGT Inc.

This paper describes a method of predicting a web site performance by simulation based on the measured data from clients. We applied the measured data of response time of the web site, total amount of data transferred, network latency and bandwidth to our abstracted model of one-tier black box system to predict response time and throughput of the web site. This method does not require server side data; therefore we are able to speculate as to the outcome of performance limitation easily.

1. 結言

1.1 目的

ブロードバンドや携帯電話によるインターネット接続がごく当たり前に行われるようになり、E コマースサイトなどでは、いわゆる「8 秒ルール」などといったパフォーマンスや可用性は重要な指標である。そのため、多くの Web サイト測定サービスが行われているが、これらのサービスの問題として、稼働中のシステムにクライアント側からシステムに高負荷を掛けて測定することができないことが挙げられる。そのため、測定時の応答時間のレポートは得られるが、Web サイトの稼働状況や限界性能を得ることは困難である。

本報告では、クライアント側からの測定結果を元に、米国 HyPerformix 社の IT システム性能シミュレーター IPS OptimizerTM1) に構築した Web サイトブラックボックスモデルを用いたシミュレーションにより、アクセス限界やスループットなどの性能情報を類推することを目的としている。

1.2 測定対象

本報告では、Web サイトの特定のページにアクセスする状況を測定して、性能予測を行う。暗号化通信、クライアント側のスクリプト記述については対象外とした。

2. Web サイト性能限界予測における 技術的課題

前提条件として、本来の役割である Web サイトの業務を妨げないことが必要である。そのため、Web サイトに対して過度な負荷を掛けて限界性能を調べるといった方法は現実的ではない。Web サイトのサーバーログを調査するという方法も考えられるが、この場合は、クライアントからの負荷に対する、Web サイト内の各サーバーの稼働率やスループットを求めることはできないが、クライアント側の視点である応答時間の予測は困難である。

本報告では、このような問題に対して、クライアント側から過度な負荷を掛けることなく測定対象の Web サイトの限界性能やスループットを予測することが本報告の課題である。

3. 外部からの測定による Web サイト の性能予測手法

3.1 Web サイトのブラックボックスモデル

本報告の手法では、測定対象となる Web サイトに対して、クライアントからのアクセスで得られる情報だけを用いて Web サイトの性能予測を行う。そのため、

Webサイトの全体構成をひとつのサーバーのひとつのプロセスに抽象化したブラックボックスモデルとして定義した。システム全体をひとつのサーバーとひとつのプロセスに代表させるため、システムの詳細な振る舞いを制限することはできないが、抽象化により考慮すべきパラメータの数も少なく済むという利点がある。

3.2 応答時間の内訳の算出方法

クライアントの振舞い（ユーザワークロード）は、次のようになる。

- (1) クライアントからの、Web サイトのトップページへのリクエストは各ネットワークおよびインターネット（遅延が発生する）を通して、Web サーバーに到達する
- (2) Web サーバーでは、基礎データから計算された「Web サイトサーバープロセス応答時間」だけ CPU を使用する
- (3) Web サーバーからインターネット（遅延が発生する）を通してクライアントにリプライが返る
- (4) クライアントはリプライが返ってきてからランダムな時間待機して再度アクセスを行う

このうち、Web サイトの応答時間の内訳は次の四つの要素に集約できる。

- (1) クライアント→サーバー転送量
(リクエストサイズ) [Bytes]
- (2) サーバー→クライアント転送量
(リプライサイズ) [Bytes]
- (3) サーバープロセス応答時間 [秒]
- (4) ネットワーク遅延時間 [秒]

応答時間の内訳のうち、サーバープロセス応答時間以外は、外部からの測定によって得られる。なおサーバープロセス応答時間は次の式により算出可能である。

サーバープロセス応答時間: x

リクエストサイズ: $y1$

リプライサイズ: $y2$

ネットワーク帯域幅: z とすると

$$x = (y1+y2)/z \dots \text{(数式 1)}$$

外部からの測定する項目は、次に定める方法で行った。表 1 に項目と測定方法を示す。

測定には、ping, wget といったネットワーク関連のコマンドと、Web サイトへのブラウザによるアクセス時

の packets キャプチャソフトである Wireshark^[2] (旧 Ethereal) を用いて HTTP プロトコルを監視する。

表 1 外部からの測定する項目

NO.	項目	測定方法
1	平均応答時間	特定ページへのアクセスを Wireshark (Ethereal) を用いてパケットキャプチャを行い、そのログから求める
2	ネットワーク遅延時間	ping コマンドの実行で得られる遅延時間を 2 で割った時間を用いる
3	ネットワーク帯域幅	wget コマンドの実行で得られる帯域幅を用いる
4	クライアント→サーバー転送量	項番 1 より求める
5	サーバー→クライアント転送量	項番 1 より求める
6	サーバープロセス時間	数式 1 より求める

3.3 ブラックボックスモデルへの適用

このように求めた測定項目は、ブラックボックスモデルのパラメータとして利用する。米国 HyPerformix 社の性能シミュレーター IPS Optimizer を用いて作成したブラックボックスモデルのトポロジー構成を図 1 に示す。

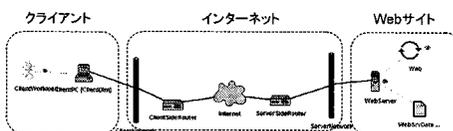


図 1 ブラックボックスモデルのトポロジー

ブラックボックスモデルでは、1 台以上のクライアント、測定で求めた帯域幅と遅延時間を持つインターネットおよび、1 台のサーバーで代表させた Web サイトからなる。

このように Web サイト全体をひとつのサーバーとサーバープロセスに抽象化しているため、サーバープロセスが、数式 1 から求めたサーバープロセス応答時間だけ、サーバーの CPU を消費し、このサーバーの

CPU 稼働率を Web サイト全体の稼働率とみなす。また、Web サイトを実測したときに同時にアクセスしている他のクライアント数を求めることができないため、測定時の実際のクライアント数（数は不明）をブラックボックスモデルの1クライアントとみなし、これをアクセスレートと呼ぶ。シミュレーション時の負荷量はこのアクセスレートを単位とする。

4. Web サイトの予測結果

本報告による Web サイトの基礎データの測定と、シミュレーションによる限界性能の予測を行った。測定対象は実際にインターネット上で情報公開を行っているサイトを使用した。

4.1 測定および、シミュレーション条件

基礎データの測定に用いたハードウェア、ソフトウェアの構成を表 2 に示す。また、測定結果を表 3 に示す。また、シミュレーション実施条件を表 4 に示す。

表 2 測定・予測環境

作業	項目	種類
測定	機種	Apple PowerBook G4 1.62 GHz
	OS	Mac OS X 10.4.3 (Tiger)
	ブラウザ	Safari 2.0
	測定ソフトウェア	Wireshark 0.99.3, GNU wget 1.53, ping
	ネットワーク	TEPCO ひかり 100Mbps(MAX)
予測	機種	DELL Optiplex GX280 (3GHz)
	OS	Microsoft Windows XP SP2
	シミュレーション制御	Microsoft Excel 2003, Windows Scripting Host 5.6
	シミュレーター	HyPerformix IPS Optimizer 3.4

表 3 Web サイト測定結果

測定日	2006年8月13日 AM 8:37
平均応答時間	0.057 秒
Ping 応答時間	2.5 ミリ秒

ネットワーク帯域幅	8128671.392 バイト/秒
クライアント→サーバー 転送量	949 バイト
サーバー→クライアント 転送量	26050 バイト
サーバープロセス 応答時間	0.053678547 秒

表 4 シミュレーション実施条件

項目	値
アクセスレート	10~250 (10きざみ)
負荷分布	指数分布に従った確率分布
負荷投入間隔 (Think Time)	平均時間 10 秒
シミュレーション 実施時間	待機 (Warm-up)時間 300 秒 シミュレーション時間 900 秒

4.2 シミュレーション予測結果

図 2 にクライアントの応答時間の予測グラフを示す。シミュレーションでは、クライアントの応答時間として、平均値、最小値、最大および、標準偏差を得ることができる。ここで応答時間平均値に着目した場合、測定時のユーザアクセスレートの 140 倍を超えると応答時間の増大傾向が顕著になる。

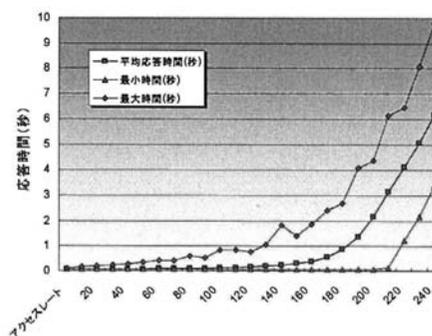


図 2 クライアント応答時間のグラフ

図 3 に Web サイトの稼働率の予測グラフを示す。測定時のユーザアクセスレートの 140 倍で、稼働率が 80%に達し、200 倍を超えるとほぼ 100%となる。これ以降 Web サイトそのものが飽和すると予測される。

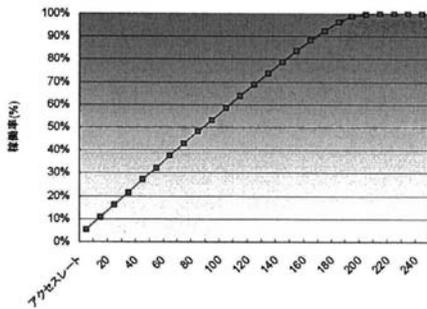


図 3 Web サイトの稼働率

図 4 に Web サイトのスループットの予測グラフを示す。スループットは測定時のユーザアクセスレートの約 190 倍で最大値に達し、その後低下することが予測される。

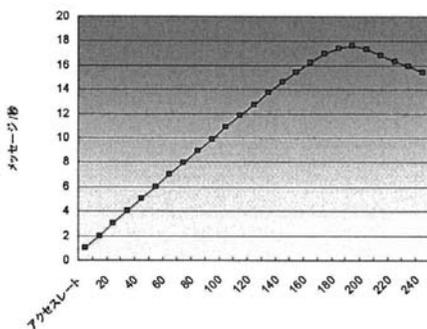


図 4 Web サイトのスループット

これらの結果から、平均応答時間の増大傾向が顕著となる測定時のユーザアクセスレートの 140 倍では Web サイトの稼働率が 80% となり、200 倍で 100% となる。同様に 140 倍を超えるとスループットが低下することが予測される。ユーザアクセスレート 140 倍以上ではスループットが低下し、クライアントの応答時間もニアに増加する傾向となり、Web サイトとしての性能限界を迎えていると考えられる。

5. 結言

5.1 結論

本報告では、測定対象の Web サイトに対して、応答時間、ファイル総容量、送受信トラフィックデータ量、ネットワーク遅延時間および、ネットワーク帯域幅を外部から測定し、その測定結果を、Web サイトのブラックボックスモデルのパラメータに適用することで、

実際の稼働中に Web サイトに対して過度な負荷を与えることなく、Web サイトの限界性能をシミュレーションにより把握することが可能となる。

本報告の手法により得られる Web サイトの性能情報は次の通りある。

- ・ アクセスレートの変化に対する応答時間の平均時間、最小時間、最大時間および、標準偏差
- ・ アクセスレートの変化に対する Web サイトの平均稼働率
- ・ アクセスレートの変化に対するスループット

本報告による手法では、複雑な Web システムを 1 層の単純なサーバーとサーバープロセスに置き換え、測定対象の Web ページも Web サイトの特定のページだけを対象としているため、基本的には、複雑なトランザクションは再現せずに、Web システムの性能を概算で求める点と、測定時に同じ Web サイトに同時にアクセスしている人数を知ることができないため、測定時の負荷に対するアクセスレート（相対比較）で結果が得られることに留意する必要がある。

5.2 今後の課題

本報告では簡便性を優先して、ブラックボックスシステムに対しての単純な 1 トランザクションに置き換えてシミュレーションを行っているが、通常 Web サイトでは複数のサービスを提供しており、それに応じたトランザクションの数がある。より現状に近い状況のシミュレーションを行うためには、本報告の手法を複数のトランザクションを扱えるように拡張することが考えられる。

また、例えば測定時の Web サーバーログを参照することで、アクセスレートを実際のアクセス数に換算することや、トランザクションの比率などを得ることが可能であるため、Web システムが通常取得しているログファイルなどを併用して、より現状に近いシミュレーションによる性能予測を行うことも考えられる。

参考文献

URL は 2007 年 5 月時点で有効である。

- 1) <http://www.hyperforni.com/>
- 2) <http://www.wireshark.org/>

連絡先

〒160-0022 新宿区新宿 2-8-8 とみん新宿ビル
 (株) ケイ・ジー・ティー Ci 技術開発グループ
 Email: hypx-info@kgt.co.jp