

1N-03 待ち行列モデルを用いたWWWサイトレスポンスタイム変動予測システム

坂田 祐司、小山 貴夫、箱守 聡、松田 栄之
(株) NTT データ 開発本部 技術開発部
e-mail: {ysakata,kym,hako,matu}@rd.nttdata.co.jp

1. はじめに

現在、WWW を用いたビジネスが急激な勢いで立ち上がっている。商用 WWW サイトにおいては顧客の確保や維持のために高い信頼性で快適にサービスを提供することが不可欠である。特に”8 秒ルール”^[1]といわれるように、レスポンスタイムが顧客の WWW サイトへの評価に与える影響は大きい。一定のレスポンスタイムを維持するためには、レスポンスタイムの悪化を予測し、予防的な対処を行うことが有効である。

本稿では、WWW サイトを構成する WWW サーバや DB サーバなどのリソースを、待ち行列網によってモデル化し、WWW サイトのレスポンスタイムの予測を行う手法を提案する。また、この手法を実装した予測システムについても述べる。

2. レスポンスタイムの予測

筆者らは、これまでに WWW サイトのレスポンスタイムをその時系列データの回帰分析によって予測する手法を提案している [2]。この手法では WWW サイトのワークロードとレスポンスタイム値の変動に高い相関があることを仮定していた。しかしながら、負荷が集中し主記憶メモリ上だけの処理が不可能になり、不揮発ディスク上のキューを使用する必要が生じた場合などに、回帰分析のみでシステムの特性をモデル化することが困難な場合があることを実験の結果によって得た。

本稿では、このようなリソース状態の変化を考慮し、精度よく予測できるシステムを構築するため、対象 WWW サイトへのワークロードと対象 WWW サイトのリソースの状態値からレスポンスタイム値を算出可能な性能モデルを構築する。その上で対象 WWW サイトから定期的にワークロードと状態値を取得し、さらにワークロードの予測を行う。そして、それらの値を性能モデルに入力することによりレスポンスタイム値の予測を行う。この処理の流れを図 1 に示す。また、個々の処理内容に関し、特徴的な処理の概要を以下の各節で述べる。

2.1. 待ち行列網による性能モデルの構築

本手法では待ち行列網を用いてリソースの挙動を

An estimation system of WWW server performance using queuing networks

Yuji Sakata, Takao Koyama, Satoshi Hakomori, and Shigeyuki Matsuda

NTT Data Corporation

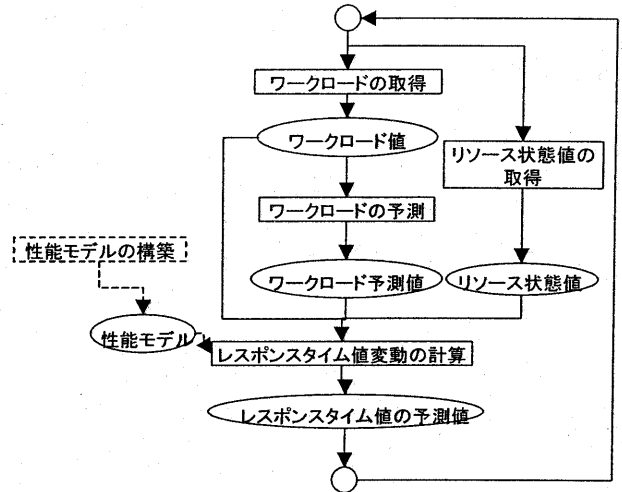


図 1: 本提案手法の処理の流れ

モデル化して表現する。すなわち WWW サイトを構成する個々のリソースを待ち行列で表現し、それらの相互作用を考慮に入れて全体のモデル化を行う [3]。このモデルでは、リソースの初期状態に相当するモデルの初期状態値と、その時点以降のワークロードに相当する待ち行列網に到着する呼量によって、レスポンスタイム値などの性能値を導出することができる。

また、待ち行列網によってモデル化された性能モデルの解決には解析的な手法 [4] とシミュレーションを用いた手法があるが、様々な WWW サイトに対応可能であることから時間追跡法によるシミュレーション [5] を用いて行う。

2.2. 回帰分析を用いたワークロードの予測

本手法では、2.1 で示した性能モデルを用いてレスポンスタイムを予測するため、ある時点でのリソースの状態値とそれ以降のワークロードを予測する方法を導入する。ワークロードの予測値は定期的に取得されるワークロードの時系列データから回帰分析 [6] によって計算する。

3. 予測システムの実現

2 章で示した手法による WWW サイトのレスポンスタイム予測システムの詳細と実装について、以下に述べる。

3.1. 予測対象 WWW サイトの概要

提案方式の妥当性を評価するため図 2 で示す TPC-W 仕様 [7] で規定されている WWW サイトを用いる。TPC-W は物品の販売を行う典型的な商用サイ

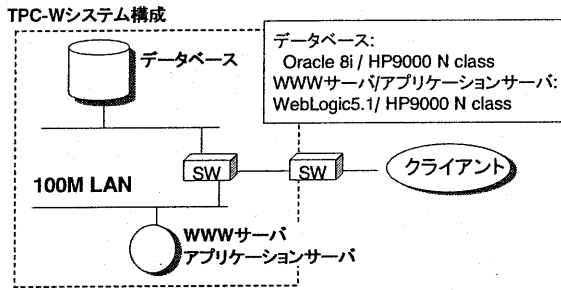


図 2: 予測対象 WWW サイトのシステム構成

トをモデル化したサイトであり、14 種類の URL で区別されるリクエストを処理する。また図 2 で示されているクライアントは TPC-W の仕様で定められた動作を行うものとし、クライアント数を変化させることにより、ワークロードの変動を模擬するものとする。また、今回ネットワークにおける遅延時間は無視できるほど小さいため、レスポンスタイムの実測はクライアントにおいて行う。

3.2. 予測対象 WWW サイトのモデル化

WWW サイトの待ち行列網としてのモデル化は、WWW サイトの構成要素の複雑さと、期待する予測精度に依存する[3]。今回の予測対象 WWW サイトでは、図 2 で示すように一台のサーバ機に対して一つ

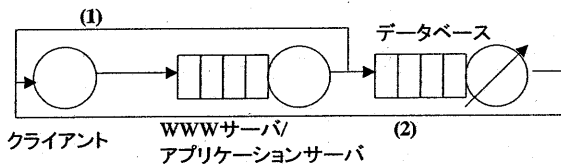


図 3: 予測対象 WWW サイトのモデル化

のアプリケーションのみが稼動している。よって、アプリケーションが複数のサーバ機を共有資源として利用することはない。よって、サーバ機を一つの待ち行列としてモデル化を行うこととした。

クライアント数が有限である場合のモデルを図 3 に示す。なお、記述方式は[3]において用いられている方式に従っている。また、各待ち行列の条件を表 1 に示す。この WWW サイトに対するリクエストは、その種類によってデータベースにアクセスしないリクエスト(図 3 の(1))とアクセスを行うリクエスト(図 3 の(2))が存在し、また個々のリクエストが WWW サーバやデータベースなどのリソースにかかる負荷は異なる。これらの条件は各リクエストの行う動作ロジックと事前のリクエストごとの負荷試験により確定した。

表 1: モデル化に用いた条件

	最大待ち行列数	同時処理サービス数	共有リソース: 待ち時間
WWWサーバ/アプリケーションサーバ	最大TCPコネクション数	サーバの最大同時スレッド数	
データベースサーバ		WWWサーバ/アプリケーションサーバのデータベースに対する最大コネクション数	
クライアント			TPC-Wの仕様に従った待ち時間

3.3. 予測システムの実装

図 4 に、予測システムの構成を示す。

本予測システムは予測対象 WWW サイトのリソースに導入される状態値取得エージェントと予測サーバから構成される。予測サーバは図 1 で示した処理を実現し、その予測結果を定期的に結果表示モジュールで表示し、また SNMP を用いて本システム外の運用管理ツールなどに通知する。

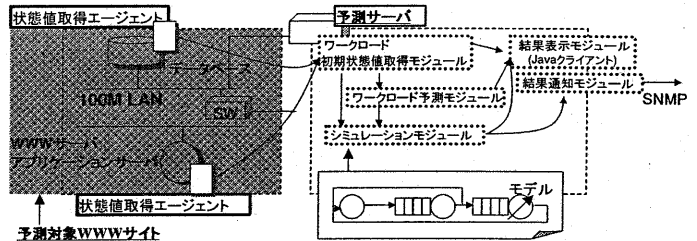


図 4: 予測システム構成

4. 構築した性能モデルの検証

今回の予測対象 WWW サイトにおいて、クライアントの数を変化させ、各リクエストのレスポンスタイムを計測した。更にこの計測結果と同条件のワークロードが WWW サイトにかけられた場合のレスポンスタイムを提案する性能モデルを用いて算出した。それぞれの結果を図 5 に示す。各リクエストにおける実測値と計算値の相関係数は 0.91 から 0.99 の間の値となった。よって、今回検証したシステムにおいてはワークロードの予測を行えば、レスポンスタイム値の予測が可能であるという点において 2 章で示した手法が妥当であることが示唆される。

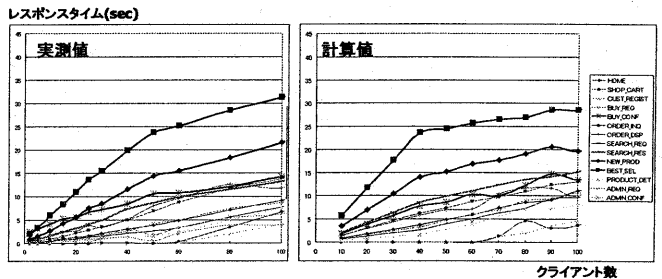


図 5: レスポンスタイム値の実測値とシミュレーションによる計算値

5. おわりに

本稿では待ち行列網を用いることによる WWW サイトの内部構成を考慮に入れたレスポンスタイム値の予測手法を提案し、この手法を用いたレスポンスタイム予測システムの実装を示した。また平易な構成の WWW サイトにおいて待ち行列網によるモデル化の妥当性を検証し、その有効性を示した。

参考文献

- [1] Zona Research, Inc Reports <http://www.zonaresearch.com/info/press/99-jun30.htm>
- [2] 坂田, 小山, 吉谷, 松田, "回帰分析を用いた WWW サーバの短期的性能予測システムの検討と実装", 情報処理学会第 61 回全国大会
- [3] D.A.Menasce, V.A.F. Almeida, "Capacity planning for web performance", Prentice Hall
- [4] D.A.Menasce, V.A.F. Almeida, L.W.Dowdy, "Capacity Planning and Performance Modeling", NJ: Prentice Hall, 1994
- [5] 秋丸 春夫, 川島 幸之助, "情報通信トラフィック", 第 7 章, (株) オーム社
- [6] 久米 均, 飯塚 悦功, "回帰分析", 岩波書店
- [7] TPC-W specification <http://www.tpc.org/wspeg.html>