

# クラウドに適した Web システムの負荷監視機能の改善と評価

10T263 堀内 晨彦 (最所研究室)

クラウド環境において負荷量に応じて動的にキャッシュサーバ数を増減させることで、応答性を確保しつつ運用コストを低減する Web システムについて、負荷監視機能の改善と評価について述べる。

## 1 はじめに

近年、クラウド環境の発展に伴い、キャッシュサーバをクラウド上に仮想マシンとして構築することが容易になり、これらの仮想キャッシュサーバを用いて Web サービスの能力を向上させることが出来るようになった。しかし、キャッシュサーバ数が負荷量に対して少なければ期待した応答性が得られなかったり、負荷量に対してキャッシュサーバ数が多ければリソースの余剰により無駄なコストが発生したりする。そこで当研究室では、クラウド環境において負荷量に応じて動的にキャッシュサーバ数を増減させることで、応答性を確保しつつ運用コストを低減する Web システムを開発している。ここで、キャッシュサーバ数を負荷量に対して最適化するためには、Web サーバにかかる負荷を正確に把握する必要がある。本研究では、Web サーバの CPU 使用率と処理数を、Web サーバデーモンと外部プログラムで測定する方法について提案し、その実装と評価を行う。

## 2 システムの概要

本システムでは図 1 に示すように、ソフトウェア L7 スイッチに機能を追加した拡張 L7 スイッチと、キャッシュ元のコンテンツを持つオリジンサーバ、取得したキャッシュを提供するキャッシュサーバ群を用いることで、第 1 節で述べた Web システムを実現する。監視及びキャッシュサーバの増減は拡張プログラムを用いて行い、リクエストの制御はソフトウェア L7 スイッチの機能を用いて行う。ソフトウェア L7 スイッチに追加する機能を以下に示す。

- A. サーバの負荷量を監視する。
- B. 負荷量に応じて、仮想キャッシュサーバの起動・停止を行う。
- C. 仮想キャッシュサーバ数の増減に合わせて、サーバ数に応じたリクエストの振分先を設定する。

## 3 負荷監視機能の改善

先行研究 [1] では、第 2 節の機能 A と C を実装したプロトタイプが作成されており、HTTP の応答時間を

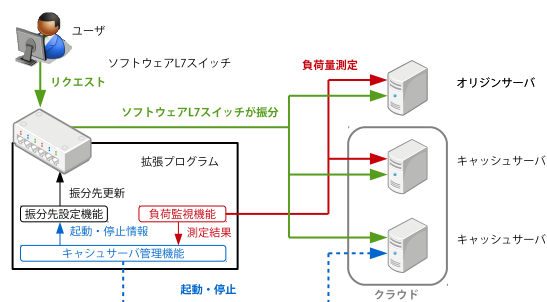


図 1: システムの構成

負荷量として用いていた。しかし、評価実験では、過負荷状態であるにもかかわらず負荷量の測定結果が小さかったために振分を停止する場合があった。

そこで本研究では負荷監視機能の精度を向上させるために、以下のパラメータを負荷量として用いることが可能であるか検討する。Web サーバとしては、代表的な Apache を用いる。

- a. Web サーバの CPU 使用率
- b. Apache の最大同時処理数に対する現在の処理数の割合 (稼働率)

これらの負荷量を測定するために、以下の方法を用いる。

- i. Apache の動作状況ページを取得・解析し負荷量を測定
- ii. Web サーバ上で動作する外部プログラムを用いて負荷量を測定

これらのパラメータ・方法の組み合わせうち、どれがシステムの負荷監視機能に最適かを評価するための実験を行った。仮想 Web サーバに対して大量のリクエストを行いながら、上述の方法で各パラメータの測定を行った。その結果、CPU 使用率はサンプルページの CGI の処理のため、リクエストが少なくとも 100%に達してしまい、過負荷状態の判別が出来なかった。稼働率はリクエスト数に追従して増減したが、方法 i では過負荷時の測定に時間がかかり、方法 ii では過負荷状態を脱した後の測定値の変化が緩やかだった。

## 4 システムの再設計

プロトタイプソースコードが複雑で機能の追加・修正を行うことが困難であったため、システムをオブジェクト指向に基いて再設計することにした。システムを構成する要素を以下のように現実世界に即したクラスとして定義することで、各クラスから操作するソフトウェアとの親和性を高める。

- **VirtualMachine:** 仮想マシンクラス  
仮想 Web サーバの負荷量の測定を行う。
- **Hypervisor:** ハイパーバイザクラス  
管理する仮想マシンの起動・停止などを行う。
- **LoadBalancer:** ロードバランサクラス  
キャッシュサーバの起動・停止に合わせて振分先の更新などを行う。
- **Cluster:** クラスタクラス  
各クラスのメソッドを呼び出してシステムを動作させる。

## 5 評価

第3節で述べた負荷監視機能により、拡張プログラムが負荷量の増減に伴ってキャッシュサーバへの振分を変更させられるか評価を行う。実験にはロードバランサ・オリジンサーバ・キャッシュサーバ2台を用い、それらはクラウド環境をシミュレートするため仮想マシン上に構築した。リクエストは2台のクライアントから行う。

実験の手順を以下に示す。ここで、振分が行われているかを評価するため、キャッシュサーバは常に起動している。

1. それぞれの Web サーバにおいて、最大同時処理数を 200 に設定する。
2. 2つのクライアントを用いて同時リクエスト数を 20 秒毎に 50 ずつ増加させ、550 になるまでリクエストを増加させ続ける。
3. その後、同時リクエスト数を 20 秒毎に 50 ずつ減少させ、0 になるまでリクエストを減少させ続ける。

負荷量に稼働率を、測定に Web サーバデーモンを用いた場合の結果を図2に示す。ここで、サービスを行っている Web サーバの稼働率の平均値が 0.8 を上回ると過負荷状態、0.2 を下回ると過負荷を脱したとみなす。

オリジンサーバの負荷量が 0.8 を上回ると 1 台目のキャッシュサーバが起動し(①)、オリジンサーバと 1 台目のキャッシュサーバの負荷量の平均値が 0.8 を上回ると 2 台目のキャッシュサーバが起動している(②)。

キャッシュサーバの起動後は負荷量が低下し、リクエストが振り分けられ負荷が分散されていることが分かる。同時リクエスト数が減少すると負荷量も減少しているが、負荷量が 0.2 以下でキャッシュサーバを停止させるため、1 台目のキャッシュサーバと 2 台目のキャッシュサーバが立て続けに停止していることが分かる(③)。キャッシュサーバの停止後はリクエストの振分先が減少したことにより一時的に負荷量が上昇している。

以上のことから、負荷量に応じてキャッシュサーバへの振分・振分停止が出来ていることが分かる。また、過負荷状態におけるキャッシュサーバへの振分停止などは発生していない。平均稼働率を用いることで、リクエスト数に応じて Web サーバの負荷量を把握出来ていることが確認できた。

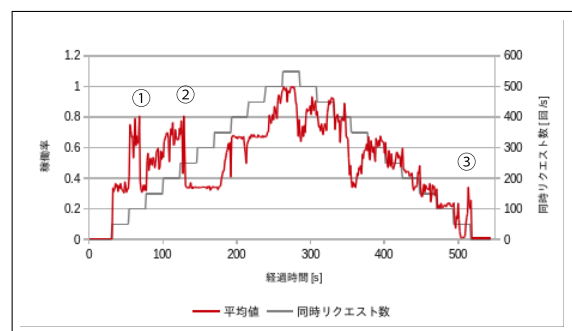


図 2: 振分実験結果 (平均値)

## 6 まとめ

クラウドに適した Web システムの概要について述べ、負荷監視の精度に問題があることを指摘した。その問題解決のため、負荷量として適切なパラメタとその取得方法について検討した。システムを再設計し、負荷監視機能の改良とその評価を行った。

今後は、以下のような内容について研究を行う予定である。

- キャッシュサーバの起動・停止を考慮した実験
- キャッシュサーバ数に応じた振分閾値の自動決定
- コンテンツ (URL など) によるリクエストの振分
- 仮想マシンの構成の違いによる評価

## 参考文献

- [1] 小笹光来, "クラウドに適したキャッシュサーバを用いる Web システムにおける管理機構の開発及びその評価", 香川大学, 修士論文, 2012