

分散 Web システムのための優先アクセス機構の実装と評価

15G485 溝淵 久哲 (最所研究室)

卒業研究にて開発した、キャッシュサーバを用いた分散型の Web システムにおいてキャッシュの品質を高く保つための優先アクセス機構を、クライアントも含めた環境にて評価し有効性を確認した。またキャッシュサーバの追加や削除に対応させるための機構を実装し機能評価を行なった。さらに Web サーバに負荷をかけるツールを開発し、今後の評価実験に向けた環境を整えた。

1 はじめに

スマートフォンなどの普及によってアクセスしたいときに Web にアクセスできる環境が整い、地震やテレビ放送など何らかのトリガによって Web サーバへのアクセス集中が起りやすくなっている。また CMS の利用が進んだことによる動的なページの増加もあり、サーバへの負荷は高まる傾向にある。本研究では、オリジナルのサーバの負荷状態に関わらずキャッシュサーバからのキャッシュ更新を安定して高速に行えるようにすることで、天気予報や株価といった数秒の遅れは問題にならない動的なコンテンツについてもキャッシュ可能とすることを目指す。

卒業研究において、キャッシュサーバからのアクセス (以下優先アクセス) を、その他のアクセス (通常アクセス) のリクエスト量に関わらず処理するための機構 (優先アクセス機構) を開発し、単体での機能評価を行なった。本論文では、クライアントを含めた環境下での本機構を用いたキャッシュサーバからのキャッシュ更新の効果を実験にて確認した。また優先アクセスの対象を動的に変更する優先アクセス管理機構の実装と機能評価、負荷量変動する負荷ツールの開発を行なった。

2 優先アクセス機構の概要

当研究室の曾川ら [2] によって設計と一部の実装が行われた優先アクセス機構を卒業研究において実装した。優先アクセス機構は、優先アクセスとそれ以外のアクセスを判定し、異なる処理部に振り分けて処理する。当研究室の加地らによって開発された NAP-Web [1] をベースにして処理部を開発した。NAP-Web は動的負荷制御機構を持っており、アクセス集中によって予め規定した時間 (応答上限時間) でリクエストを処理できないと判断した場合は、一旦リクエストを処理せずに返す。その際に次回アクセス可能となる時間と仮想的な整理券を渡し、再度アクセスしてきたリクエストは必ず処理する。処理部を優先アクセスと通常アクセスで分けることによって、それぞれのアクセスによる負荷がもう一方の処理に影響を与えないようにした。

3 キャッシュサーバを用いた機構の評価

本機構を用いることで、オリジンサーバでも、クライアントに対してキャッシュサーバが混雑時の安定して応答できることを実験で確認する。キャッシュサーバへアクセスするクライアント (間接クライアント) とオリジンサーバへアクセスするクライアント (直接クライアント) を用意し、直接クライアントからの負荷でオリジンサーバを過負荷にした状態でのキャッシュ更新の状況を再現した。本機構あり/なしの 2 種類のオリジンサーバで実験し、間接クライアントへの応答時間、直接クライアントへの応答時間を比較する。アクセス内容は掲示板の表示、キャッシュ有効期限は 5 秒、NAP-Web の応答上限時間は 4 秒とした。

優先アクセス機構を用いた際の間接クライアントの応答時間を図 1 に示す。図 1 を見ると、ほぼ 0 秒の応答時間と、0.25 秒程度のアクセスがある。後者がキャッシュの更新を必要としたリクエストであることを、オリジンサーバのログから確認している。

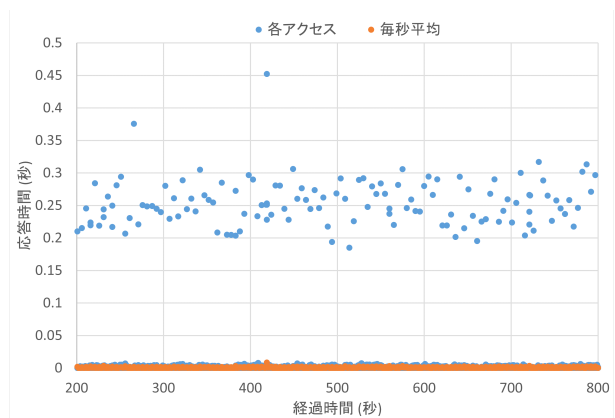


図 1: 間接アクセス 応答時間 (機構あり)

優先アクセス機構なしで実験した際の間接クライアントの応答時間を図 2 に示す。図 2 を見ると、図 1 と同様にキャッシュの更新を必要としたアクセスの応答に時間がかかっていることがわかる。しかし、優先アクセス機構ありの際は 0.25 秒程度だった応答時間が、機構なしでは 7 秒程度と長くなっている。オリジンサーバの負荷状態がキャッシュ更新に影響を及ぼしていることが確かめられた。

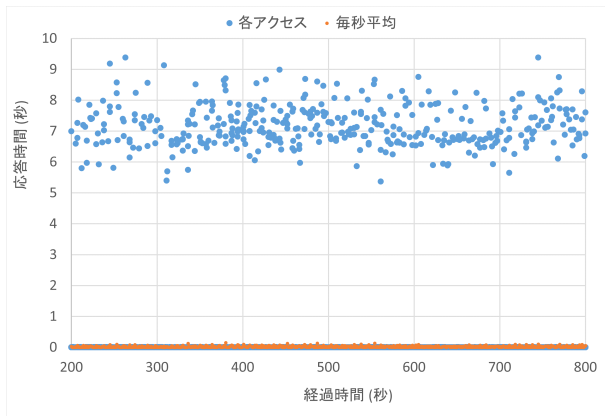


図 2: 間接アクセス 応答時間 (機構なし)

以上のことから、本機構をキャッシュ更新に用いることで、オリジンサーバが過負荷状態にあっても、キャッシュの更新が短時間に行われていることが確認できた。このことから優先アクセス機構を用いることで、オリジンサーバの負荷状態に関係なく、動的コンテンツのキャッシュを頻繁に更新することができることを示している。

4 優先アクセス管理機構

これまでの実装では優先クライアントが静的に指定していた。このため、動的に IP アドレスが割り振られたキャッシュサーバを利用する場合や、コンテンツ管理者などの特定のユーザがオリジンサーバを Web インタフェースで操作する際に優先アクセスを提供可能とするため、優先アクセス対象を動的に追加・削除するための機構を実装した。

クライアントの識別には IP アドレスと Cookie を用いる。IP アドレスはキャッシュサーバを、Cookie はそれ以外からのアクセスの識別に用いる。キャッシュサーバの識別に IP を用いるのは、キャッシュサーバに手を加えずに利用できる機構を目指したためである。キャッシュサーバの追加や削除の指示を出すキャッシュサーバ管理機構は堀内ら [3] によって開発されており、そこから利用しやすい形が望ましいと考えた。そのため API を用意してリクエストを送ることで、追加や削除が行われるようにする。Cookie についても優先キーを発行・削除する API を実装する。

優先 IP アドレスの登録と削除を行う IP-API と、優先 Cookie の発行と削除を行う Cookie-API を実装した。それぞれの動作を確認し、正常に登録と削除が行われること、API を用いて登録したクライアントからのアクセスが優先アクセスとして扱われていることを確認した。

5 負荷テストツールの開発

これまで実験に用いていた負荷テストツールは、一定量の負荷をかけることしかできなかった。しかしこのツールでは、実環境を想定した急激なアクセスの

増加に対する評価や、実際のログを基にしたアクセス量の変化に対する評価が行えない。既存の負荷テストツールも調査したが、NAP-Web の再アクセスに対応しつつ負荷量を変動させられるツールは存在しなかった。そのため、設定された毎秒の負荷量に従い Web サーバに負荷をかけるツールを開発した。

実装したツールを用い、実験を行なった。50 から 1000 まで変動するシナリオで実験し、キャッシュサーバに対して負荷テストを行なった。実験で得られた実測値 (サーバのアクセスログから確認した値) とシナリオの値を比較したところ、シナリオで毎秒 200 以上のアクセスを行っているところでもスループットが 200 で頭打ちになっていた。CPU がボトルネックになっている可能性を考え、GNU Parallel を用いてツールを並列実行した。4 コアのクライアントを用いて単調増加するシナリオで、頭打ちになる箇所を調べたところ、コア数分を並列実行するところまではスループットは上昇したが、想定した程は上昇せず 600 程度で頭打ちとなった。この原因の調査は今後の課題とする。

また複数台でツールを動かした結果、1500 を超えるスループットが得られた。副次的なことであるが、同程度の性能のオリジンサーバのスループットが 130 程度であったことから、キャッシュサーバを用いることで少ない台数でサービスが可能になることが確かめられた。

6 おわりに

優先アクセス機構をキャッシュサーバのキャッシュ更新に利用することで、常に安定したキャッシュ更新が行えることを確認した。また、優先アクセス管理機構を実装し、固定 IP を持たないサーバや、コンテンツ管理者からのリクエストを優先アクセスとして扱えるようにした。さらに、負荷量テストツールを開発し、課題はあるものの指定したシナリオに沿った負荷をかけることが出来ていることを確認した。

今後の課題として、実際のアクセスログや様々なコンテンツを用いた評価、CPU 使用量や電力量などを評価軸とした評価、想定する分散 Web 環境での評価がある。

参考文献

- [1] 加地智彦, 最所圭三, "NAP-Web における次回アクセスの効率化", 電子情報通信学会論文誌, Vol.J92-B, No.10, pp.1600 1610, 2009
- [2] 曾川直也, "NAP-Web における優先アクセスの設計と評価", 香川大学, 学士論文, 2014
- [3] 堀内辰彦, 最所圭三, "負荷予測に基づく分散 Web システムにおけるキャッシュサーバ管理機能の改良", 2016 年電子情報通信学会総合大会講演論文集, B-6-18, 通信講演論文集 2, p.18, 2016.3