

分散 Web システムにおけるクライアントに対するキャッシュサーバの効果 を評価するためのシミュレータの開発

14G466 杉山 友啓 (最所研究室)

本稿では、クライアントのアクセスに対するキャッシュ更新の評価を行うためのシミュレータの開発について述べる。

1. はじめに

我々の研究室では Web サーバとキャッシュサーバ群からなる分散 Web システムを開発している[1]。キャッシュサーバからの更新が滞らないために Web サーバでキャッシュサーバからの更新アクセスのためにリソースを確保し、通常のアクセスよりも優先的に処理を行う機構の開発も行っている[2]。本研究では、キャッシュサーバの更新のためのリソースが確保されていることを前提としたキャッシュ更新のシミュレータを開発する。

卒業研究ではキャッシュサーバのキャッシュを逐次的に更新するシミュレータの開発を行った。しかし、開発したシミュレータでは逐次的なキャッシュ更新のみの評価になり、現実のキャッシュ更新に即した複数のキャッシュの同時更新について評価はできなかった。そのため、キャッシュを逐次的な更新ではなく複数のキャッシュを同時に更新できるようにシミュレータの改良を行った。さらに、クライアントも含むシミュレーションを行うことができるようにも拡張した。本稿では、クライアントを含むシミュレーションについて述べる。

2. 分散 Web システム

図 1 に分散 Web システムの構成を示す。クライアントからアクセスがあった場合、まず振り分け機構に送られ Web サーバもしくはキャッシュサーバにアクセスを振り分ける。Web サーバのアクセスが多くなり処理できなくなってきた場合はキャッシュサーバに振り分けられ、負荷を軽減させる。

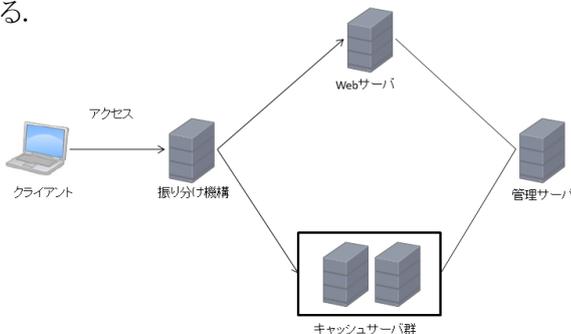


図 1:分散 Web システム

これにより、クライアントは高速にアクセスすることができ

る。本研究ではキャッシュサーバ上のキャッシュ管理を対象としたシミュレーションの開発を目指している。

3. キャッシュデータの品質

キャッシュサーバが保持しているキャッシュは時間が経過していくにつれて提供できる情報の品質が低下する。キャッシュするコンテンツはさまざまな種類があり、キャッシュの品質低下のパターンは数多くある。たとえば広告は宣伝の内容によっては期限があるのもがあり、期限が過ぎてしまうと広告の意味がなくなってしまう。このようなコンテンツは有効期間中、キャッシュの品質は低下しないが、有効期間が過ぎてしまうとキャッシュの品質が 0 になる。それに対してブログや掲示板などはアクセスするたびに更新されている可能性があるが、最新情報でなくても十分な情報が得られると考えられる。そのためキャッシュの品質は急激には落ちず、徐々に低下していくと考えることができる。

本研究では、この 2 つのパターンでそれぞれのキャッシュの品質が変化するコンテンツについて扱うことにする。キャッシュの有効期限を超えるとキャッシュの品質が 0 になるキャッシュをタイプ A とし、時間の経過とともにキャッシュの品質が低下していくキャッシュをタイプ B とする。

4. シミュレーション

タイプ A とタイプ B 両方ともキャッシュの品質を 100%から開始する。クライアントのアクセスがない場合、キャッシュ更新はタイプ A を優先して行い、空いている時間を利用してタイプ B の更新を行う。タイプ B を更新するタイミングとして以下の 3 つを考えた。

- タイプ A と次のタイプ A の間にできる更新の空き時間で更新できるタイプ B があれば更新を行い、ない場合はそのままにする。
- 後ろにあるキャッシュ更新が期限に間に合わない場合が出てくるが、タイプ B を更新する。
- 次のタイプ A の更新を前倒しにし、まとまった空き時間を後ろに作り、タイプ B の更新を行う。

今回は実装が容易かつタイプ B の品質低下が a.より少ない b.の方法を扱う。

クライアントのアクセスが発生した場合、キャッシュが存在し、その有効期限が切れていない場合はキャッシュを配信し、そうでなければ Web サーバにアクセスしそのリソースを獲得し、キャッシュして配信しなければならない。キャッシュを獲得する必要がある場合、空いている時間があれば必要なキャッシュの獲得を開始できる。しかし、空き時間がない場合は、できる限りクライアントへの応答時間を早くするために、更新中のキャッシュを中断してクライアントが必要なキャッシュの更新を開始するようにした。中断する対象は、現在更新中のものからキャッシュの更新時間が一番長いものとする。クライアントのアクセスを含めない場合と含めた場合について以下のパラメタでシミュレーションを行った。結果を図2、図3に示す。

○キャッシュの同時更新数:4

・タイプ A

キャッシュ数:100 個

更新時間:40~60 秒

更新間隔:2700~3300 秒

・タイプ B

キャッシュ数:100 個

更新時間:40~60 秒

品質の減少量:1 秒につき 0.015%~0.020%

更新開始品質:50%以下

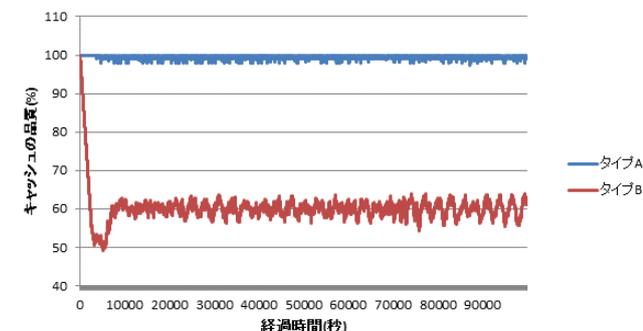


図2:クライアントのアクセスがない場合

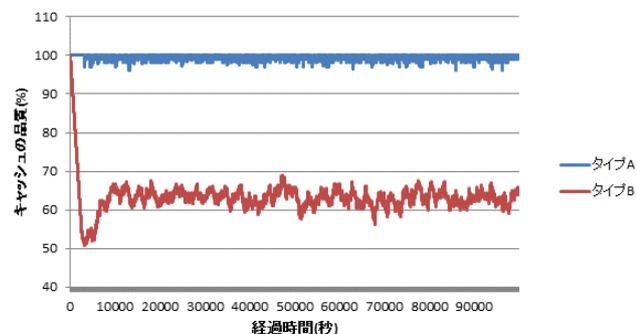


図3:クライアントのアクセスがある場合

タイプ A のキャッシュの品質はクライアントのアクセスの有無に関わらず影響が少ない。クライアントのアクセスがない場合のタイプ B のキャッシュの品質の揺れは規則性になっている。アクセスがある場合は揺れ幅も大きく、規則性も乱れているが、アクセスがない場合と比べると、キャッシュの平均品質は上がっている。これは、クライアントがタイプ B のキャッシュをアクセスし、タイプ B のキャッシュ更新が優先して行われたためだと考えられる。

図4は、クライアントの応答時間を示す。今回、クライアントが必要としているキャッシュは4つに設定している。クライアントの応答時間については、必要なリソースのキャッシュがすべて更新する必要がない場合は、すぐにクライアントに配信できるため催促の応答時間である1秒になっている。更新が発生している場合は、更新時間が40~60秒の間になっており、そのことがグラフに反映されている。以下の結果から、今回の実験ではクライアントからのアクセスに対し、たかだか1つのキャッシュミスが発生していたと思われる。

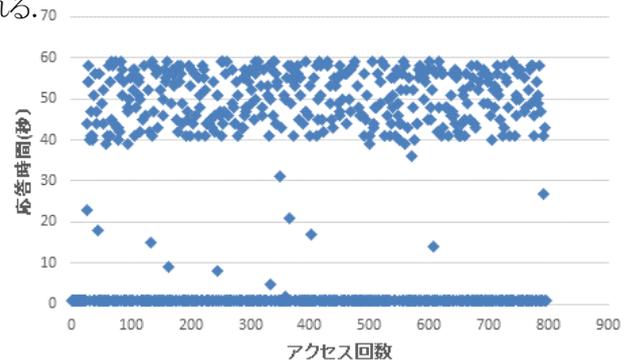


図4:クライアントの応答時間

5. おわりに

キャッシュ更新を逐次更新から同時更新が可能になるようにシミュレータの改良を行い、さらにクライアントのアクセスを入れたシミュレータの開発を行った。逐次更新から同時更新になったためより現実的なキャッシュ更新のシミュレーションが可能になった。

参考文献

- [1] 加地智彦, 最所圭三: “過負荷時のユーザの不満を抑えるために次回アクセスを保証する Web システム”, 情報処理学会論文誌, Vol.50, No.2, pp.872-881, 2009
- [2] 溝渕久哲, 最所圭三: “優先アクセス機構を持つ Web サーバの開発およびその評価”, 第14回情報科学技術フォーラム講演論文集, L-021, Vol.4, pp.231-234, 2015