

NAP-Web のためのリソース解析機能を持つプロキシの開発

08T273 山田 大樹 (最所研究室)

次回アクセスを保証する NAP-Web の動的ページへの対応、
時間予測の精度向上のためにプロキシツールの開発を行う。

1. はじめに

我々の研究室では、アクセス集中時にサービスできないクライアントに対して整理券を配布することによって次回アクセス時間を通知し、その時間に行われるアクセスを受け付けることを保証する Web システム NAP-Web (Next Access Promised Web) の開発を行っている[1]。NAP-Web では複数のコンテンツから構成されるページにアクセスする場合、連続アクセスが生じるがこれらに十分に対応していなかった。

この問題を解決するために Web コンテンツの情報をもとに待ち時間の予測を向上する「ページ情報取得システム」の開発が検討された[2]。しかし動的ページに対応しておらず、時間予測をサーバ側で行うので NAP-Web への負荷が大きいなどの問題があった。

そこで本研究では、これらの問題に対処するために、NAP-Web を支援するリソース解析機能を持ったプロキシツールの開発を行うことにした。

2. NAP-Web の概要

NAP-Web は図 1 に示すように、サーバにアクセスが集中した際にアクセス制御を行い、サーバダウンを防ぐシステムである。アクセス制御により、拒否されたクライアントに対しては、次回アクセス可能な時間を予想し、その時間が書かれた整理券を発行する。その整理券に書かれた時間にクライアントが再アクセスすると、優先的に応答を返すというシステムである。

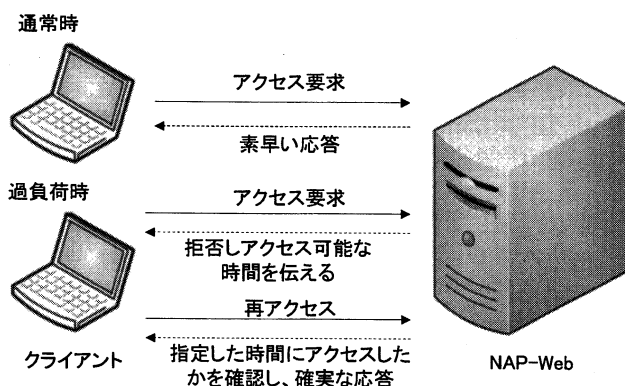


図 1. NAP-Web の概要

3. NAP-Web の問題点

NAP-Web では時間予測精度の悪さ、複数のコンテンツから構成されているページへの不十分な対応などの問題がある。この問題を解決するために新田氏の研究では連続アクセスに対応するために、Web コンテンツごとのリソース構成および、リソースごとの処理時間の統計情報を収集し、それらの情報を用いて、将来アクセスされるリソースを予約し、処理時間を予測する手法を提案していた。しかしこの手法は、動的ページに対応しておらず、サーバの負荷が大きいという問題がある。

動的ページに対応するために動的ページのリアルタイムでの解析を行い、サーバの負荷を軽減するためにクライアント側で動作し、解析した結果をサーバに送信する機能を導入したプロキシツールの開発を行う。

4. プロキシツールの概要

上記で述べたプロキシツールを用いた処理の概要を図 2 に示す。

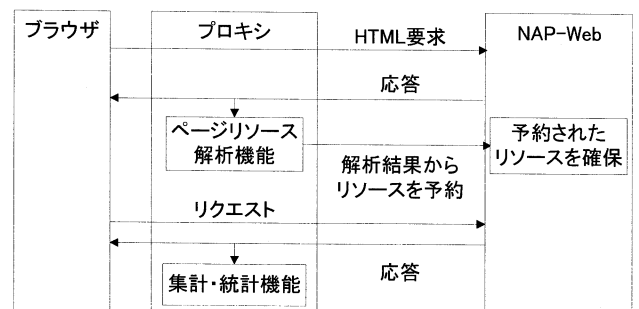


図 2. プロキシツールの概要

ページリソース解析機能ではブラウザからサーバに HTML (ページリソース) の要求があったとき、返ってきた HTML リソースをプロキシが受け取りブラウザに渡すとともに、そのときそのページを構成するリソース構成を調べる。リソースを抽出し、抽出したリソースを NAP-Web に予約する。予約する際には POST メソッドを用いて、NAP-Web の専用の CGI ページに送信する。リソースをリクエストされたときに素早い応答をすることが可能となる。集計・統計機能では、サーバからリソースを受け取った際のファイルの種類、サイズ、応答時間などを調

べ、解析を行う。解析結果から集計・統計を行い、時間予測に利用する。

5. 動的ページの調査

提案する手法の有用性を確認するために動的ページがどのような構成を調査する。動的 Web ページにアクセスし、アクセスしたログから各ページの解析を行う。調査方法は、キャッシュ機能を切ったブラウザで、Yahoo などのアクセス数が多い 10 種類の動的 Web ページに 10 回ずつアクセスを行いリソース名、ファイルサイズの収集を行う。収集したログの解析結果を図 3、4、5 に示す。

図 3 は動的ページに 10 回アクセスした際の、リソースごとの取得回数を示す。取得するリソースの大部分が、毎回アクセスされているものであるとわかる。それ以外ではアクセス回数の少ないリソースの取得回数が多い。

図 4 は数の代わりに、リソースサイズの合計をプロットしたものである。図 3、4 を見てみると動的ページの構成の大部分は、毎回アクセスされるリソースであり、リソースの取得回数の割合より、サイズの割合が大きいことから、出現回数の少ないリソースのサイズは大きい傾向があると考えられる。いずれにしても、共通するリソースは良くリクエストされるリソースであり、これらをより正確に統計・集計を行い精度の高い時間予測を行うことができると考えられる。

図 5 は同一のページにアクセスしたときの前のアクセス (ページ A) と、後のアクセス (ページ B) のリソースサイズの構成を比較した平均を示している。ページ A とページ B の差分の比率はほぼ一致している。このことから各アクセスの処理量はほとんど変化がないと考えられ、動的ページでも正確な時間予測を場合ができると考えられる。

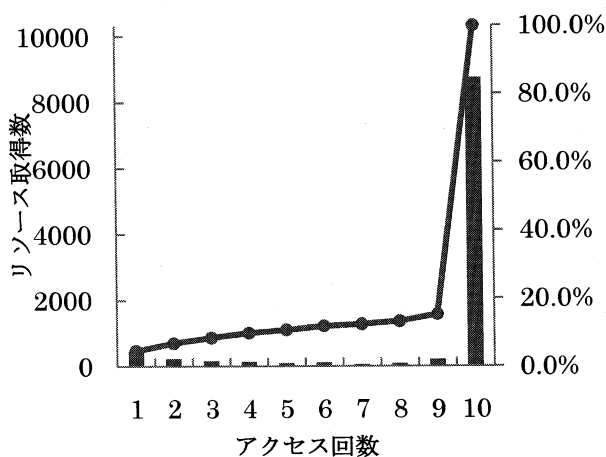


図 3.リソース取得回数

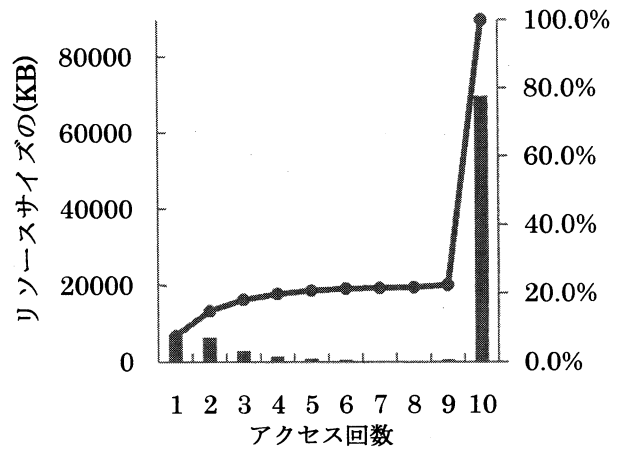


図 4.リソースサイズ

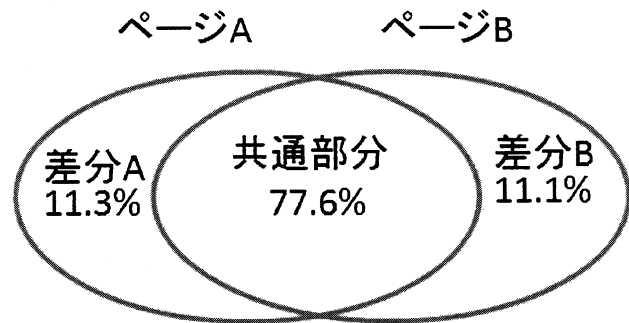


図 5.ページサイズ構成比率

6. まとめ、今後の課題

今回の調査では動的ウェブページにおけるコンテンツの構成について確認することができた。動的 Web ページの大部分が毎回アクセスされるリソースであった。ページサイズの構成を比較すると、サイズ差分の平均値は、ほぼ一定であった。このことからページの構成がほぼ一定のページでは、共通部分と差分の集計・統計を行うことにより、動的ページでも時間予測ができると確信した。

今後はさらに多くの動的ページを調査し、今回の調査から得られた情報より、より良い時間予測、統計・集計の手法の検討、プロキシツールの開発、実装を行っていく。

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] 新田淳二, “アクセス情報を用いた Web サービスにおけるサービス時間の予測”, 香川大学工学研究科, 修士論文, 2011