

Proxyサーバを利用した新たなサービスの可能性について

村瀬 茂樹 北 英彦 林 照峯

三重大学工学部

近年、インターネットが急速に普及し、中でもWWW (World Wide Web) への関心が高まっている。インターネットではしばしばアクセス制御やトラフィックの軽減のために Proxy サーバが使用される。

Proxy サーバで収集したアクセスログとキャッシュの解析、Proxy サーバによる Web ページへの情報の付加、及び、ユーザの利用状況の把握によって、Proxy サーバは本来の使用目的以外にもグループにおける情報共有の支援、個々のユーザに合わせたサービスの提供を行うことができる。本研究では Proxy サーバの新たな利用法について提案し、その実現例を示す。

Personalized Service and Collaboration Support Using WWW Proxy Server

Shigeki Murase Hidehiko Kita Terumine Hayashi

Faculty of Engineering, Mie University

Recently, the Internet, especially WWW (World Wide Web), is popularized rapidly. Proxy server is often used for access control and reduction of traffic on the Internet.

Proxy server can support not only the original purpose but also the information sharing and the personalized service by the following methods: (1) analysis of the access log and cache which proxy server collected, (2) addition of the information to Web pages, and (3) grasping of the user's use situation. We suggest new service using proxy server.

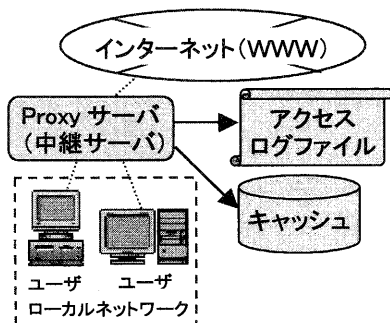
1 はじめに

1960年代後半に米国で誕生したインターネットはその後全世界に広がり、特にここ数年、WWWへの関心は爆発的に高まっており、そのユーザ数を急激に伸ばしている。

ローカルネットワークとインターネット間のアクセスを中継するために Proxy サーバを設置することがある。これは、外部からローカルネットワークのセキュリティを確保すること、及び、インターネットへのアクセス制御をすることを目的としている。WWWの場合には、上記の目的以外にキャッシュ共有によるインターネットのトラフィックの軽減を目的とする。

Proxy サーバで収集したアクセスログとキャッシュの解析、Proxy サーバによる Web ページへの情報の付加、及び、ユーザの利用状況の把握によって、Proxy サーバは本来の使用目的以外にも、グループにおける情報共有の支援、個々のユーザに合わせたサービスの提供を行うことができる。本研究では Proxy サーバの新たな利用法について提案し、その実現例を示す。

2 WWW の Proxy サーバについて



Proxy サーバは、特にインターネットの中の1つのサービスである WWW (World Wide Web) で使用されることが多い。WWW において Proxy サーバがインターネットと LAN の間に設置される理由は、以下に示す3つである。

①セキュリティの確保

インターネットから LAN 内のコンピュータに直接アクセス出来なくする。

②Web ページへのアクセス制限

Web ページの URL をチェックして、指定した Web ページを表示不可にする。

③トラフィックの軽減

Web ページのキャッシュを LAN で共有して、WWW のアクセス量を少なくする。

ユーザが Proxy サーバを中継して WWW にアクセスすると、そのときに Proxy サーバが行なった処理が「アクセスログファイル」に、Web ページのコピーが「キャッシュ」に保存される。

3 Proxy サーバを利用した過去の研究

Proxy サーバのアクセスログファイルとキャッシュを解析することで、Proxy サーバの本来の使用目的以外にも別のサービスをユーザに提供することが可能である。

そこで我々の研究室では、これまでに Proxy サーバを利用した様々な研究を行ってきた。Proxy サーバのアクセスログを解析することにより研究室内からアクセスが多かった Web ページを推薦する「やじうまくん」システム[1][2]、Proxy サーバのアクセスログとキャッシュの双方を解析することによって Web ページの読み込みにかかった時間を計測するシステム[3]がある。

このように Proxy サーバを利用すれば様々なサービスをユーザに提供することが可能である。従ってこれ以外にも Proxy サーバを利用すれば様々なサービスをユーザに提供することが出来るのではないかと考えられる。

本研究では、上記以外の Proxy サーバの新たな利用法について考察・提案し、そのサービスを提供するのに必要な新たな Proxy サーバの提案を行う。

4 Proxy サーバの新たな利用法

4.1 Proxy サーバを用いることで可能なこと

まず初めに、「Proxy サーバを利用することで可能なサービスには何があるか？」を考えてみる。

1つ目にアクセスログとキャッシュの解析がある。これによって得られる情報を元に、グループにおける情報共有の支援や個々のユーザに合わせたサービスの提供が可能である。

2つ目に Proxy サーバが中継サーバであることを利用して、Web ページの加工をしてユーザに提供する、ということが出来る。

3つ目に同じく中継サーバであることを利用して、リアルタイムによるユーザの利用状況の把握が出来る。これを元にして、同じくグループにおける情報共有の支援や個々のユーザに合わせたサービスの提供が可能である。

4.2 Proxy サーバが提供可能な具体的なサービス

Proxy サーバを用いることで提供可能なサービスを以下に具体的に挙げる。

①Web ページの予想読み込み時間の表示

これは解析によって得られた情報と Web ページの加工を利用したものである。Web ページの閲覧時に、そのページが表示されるまでの読み込み時間を測定してデータベースに蓄え、他のユーザがそのページを開こうとしたとき、もしくは開く前にその予想読み込み時間をユーザに提供する(図1)。データベースにヒットした Web ページへのリンクがある場合に、Proxy サーバによって予想読み込み時間を付加する。

②Web ページの評価を入力

これは Web ページの加工を利用したものである。各 Web ページに評価入力用のボタン等を付加することによって、そのページの評価を

ユーザに入力してもらい、データベースに蓄える。そのデータを元に、他のユーザに Web ページを推薦する(図2)。

③Web ページの先読み

これはユーザの利用状況の把握を利用したものである。現在、ユーザが閲覧している Web ページの次のページを Proxy サーバが先読みすることで、Web の巡回を快適にする。

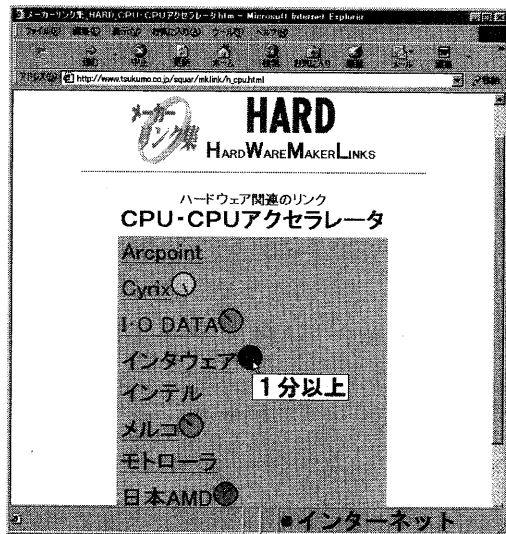


図1. Web ページの予想読み込み時間の表示

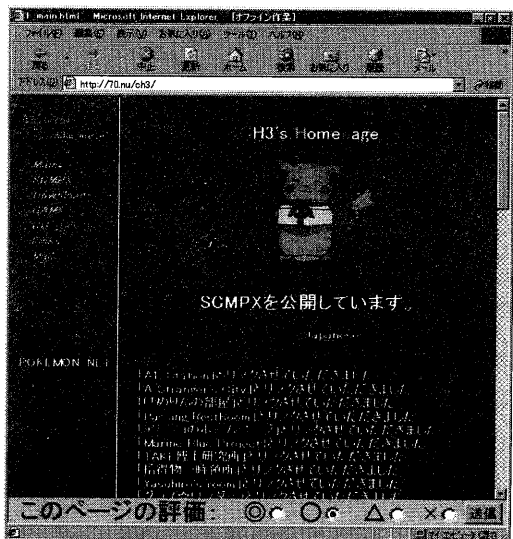


図2. Web ページの評価を入力

④お知らせ情報の挿入

Web ページの加工を利用したものであり、Web ページの一部にお知らせやニュース等を付加して、ユーザに提供する。

⑤広告ページの除去

Web ページの加工を利用して、ポップアップ Window 等で表示される、煩わしい広告ページを表示させなくする。

⑥自動翻訳

Web ページの加工を利用して、外国語で書かれた Web ページを日本語に自動翻訳して、ユーザに提供する。

4.3 既存の Proxy サーバの問題点

しかし、以上に挙げたサービスは、既存の Proxy サーバでは提供することが事実上、不可能である。

(1) 既存の Proxy サーバは、キャッシュを取ることを禁じている HTML ファイルと、CGI が出力した HTML ファイルはキャッシュを取ってはいけないことになっている。

(2) 既存の Proxy サーバのアクセスログからでは、データ解析に必要十分な情報が得られない。

(3) 既存の Proxy サーバには Web ページを加工する機能がない。

以上の理由によって、既存の Proxy サーバでは上記のようなサービスの提供が不可能なのである。これら3つが可能な Proxy サーバを製作すれば、様々なサービスをユーザに提供することができると考えられる。

5 Proxy サーバを用いた新サービス提供システム

5.1 システム構成

図3は本システムの基本的なシステム構成図である。ローカルネットワークとインターネット間に専用の Proxy サーバを設置して、サーバから送られるデータに対して様々なサービスを付加してクライアントに配信する。

また、解析に必要な詳細なログファイルと、キャッシュとは別に解析用として、本来はキャッシュとしては使用してはならないことになっている HTML ファイルも含めて、全ての HTML ファイルを保存する。そして、これらを解析することによってもまた、ユーザに様々なサービスを提供する。

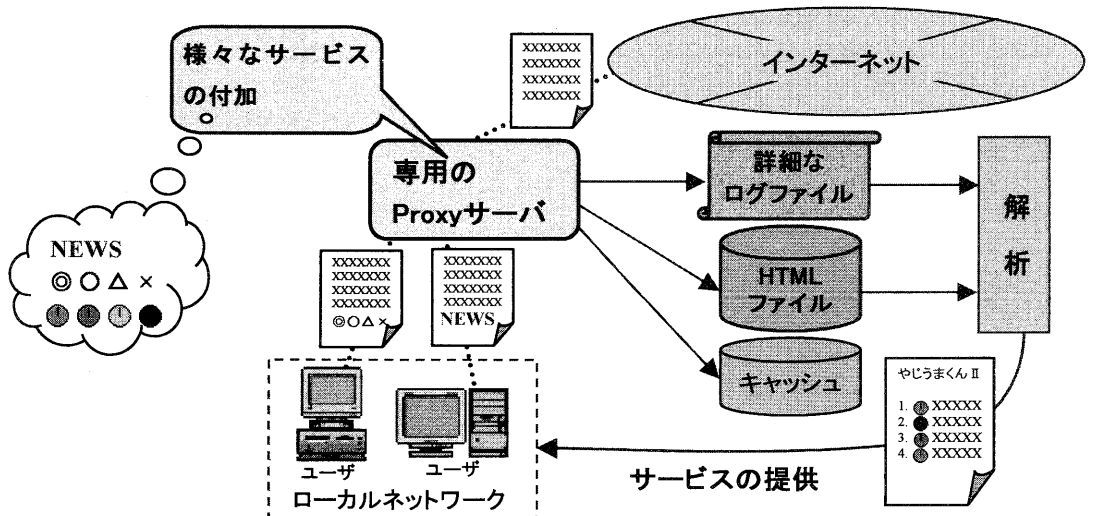


図3. Proxy サーバを用いた新サービス提供システムのシステム構成図

5.2 試作システム(読み込み時間の表示)

これまでに挙げてきたサービスを如何に実証するか?と言うことで、全 HTML ファイルのコピーを保存し、また Web ページの予想読み込み時間を付加してクライアントへ配信する機能を持つ Proxy サーバを製作した。

5.2.1 システム構成

(読み込み時間の求め方)

図4が本システムの「読み込み時間の求め方」に関するシステム構成図である。プログラム簡略化のために、既存の Proxy サーバと併用する形式をとった。既存の Proxy サーバである DeleGate のアクセスログファイルと今回製作した Proxy サーバが残す HTML ファイルのコピーの双方を解析することによって、過去に Web ページを閲覧したときにかかった読み込み時間を測定している。

5.2.2 Web ページの読み込み時間の求め方

本システムでは、プロキシサーバのログファイルと HTML ファイルの双方を解析することで、Web ページの読み込み時間を求めている。これは、Web ページは主に HTML ファイルと複数の画像ファイル等から構成されているが、

ログファイルからでは Web ページがどのファイルから構成されているかを知ることが出来ないため、ログファイルのみの解析では読み込み時間を求めることができないのである。

そのため本システムでは、以下のような方法で Web ページの読み込み時間を求めている。

- (1) ログファイルの中から HTML ファイルの読み込み開始時刻を抜き出す。
- (2) 保存されているその HTML ファイルのコピーを解析して、その Web ページの構成ファイルを調べる。
- (3) 再びログファイルに戻って、その構成ファイルの中で一番最後に読み込みが完了したファイルを探す。
- (4) その読み込みが完了した時刻と HTML ファイルの読み込みを開始した時刻の差をとって、Web ページの読み込み時間を求める。

また、求めた読み込み時間はフリーのデータベースである PostgreSQL[4]に保存している。

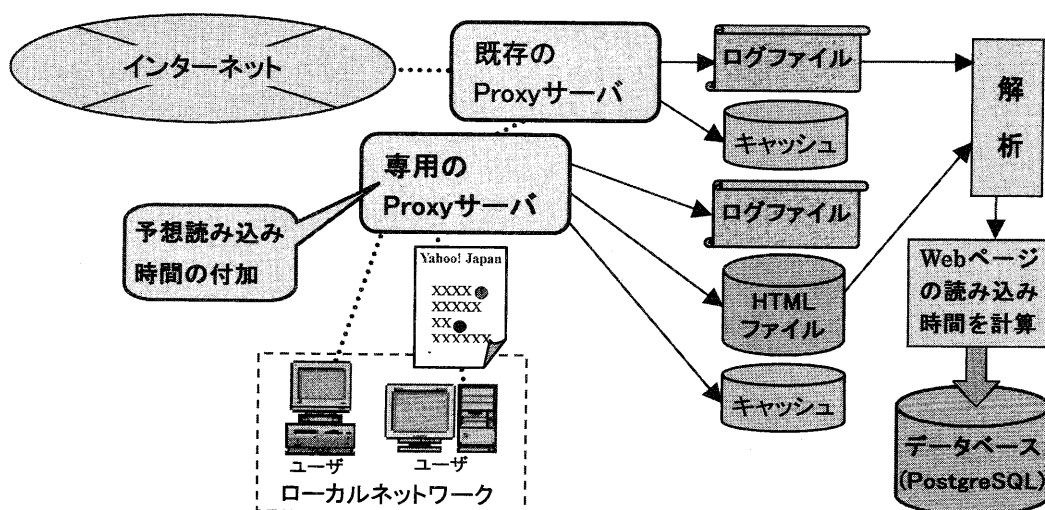


図4. 試作システムのシステム構成図 (1)読み込み時間の求め方

5.2.3 読み込み時間の付加

Web ページの予想読み込み時間を提供するサービスのシステム構成図を図5に示す。

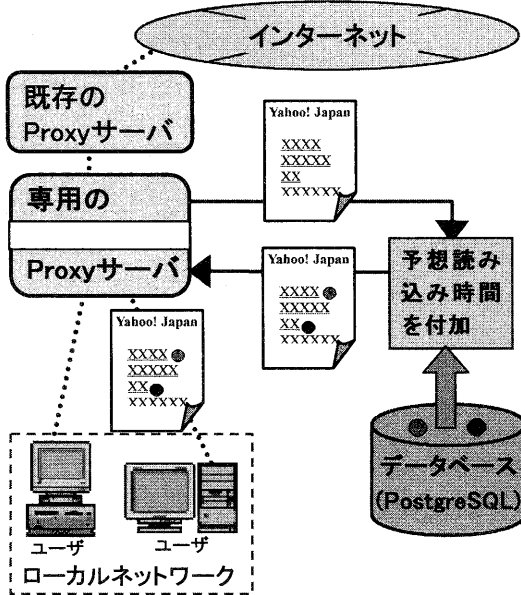


図5. 試作システムのシステム構成図
(2) 読み込み時間の付加

クライアントへと渡されるファイルのうち、HTMLのみを行単位で Perl のプログラムに渡す。そしてその行にアンカータグが記述されていたら、PostgreSQL にアクセスして予想読み込み時間を付加する。そして Proxy サーバにその行を返し、クライアントへと配信している。

また読み込み時間は、そのまま数字で表示すると見づらくなるので、本システムでは読み込み時間の早さを4段階の時計のアイコンで表示している(図6)。

5.2.4 試作システムの評価

予想読み込み時間を膨大なデータベースから検索して付加するためかなりの時間が必要で、実際に利用するには改善の必要がある。



図6. 本システムを用いて閲覧した Web ページ

6 終わりに

本研究では Proxy サーバを利用することで提供可能な新たなサービスについて具体的に提案し、その実現例を示した。

Proxy サーバを利用した新たなサービスの可能性を示すことができたと考えられる。

参考文献

- [1] 杉井俊彦, 北彦彦, 林照峯: “アクセス回数を利用した WWW の人気ホームページ道案内システム”, 情報処理学会グループウェア研究会, 研究報告 97-GW-21, p.235-240, 1997
- [2] 杉井俊彦, 北彦彦, 林照峯: “WWW のお勧めホームページについての情報共有システム”, 情報処理学会グループウェア研究会, 研究報告 98-GW-26, p.103-108, 1998
- [3] 村瀬茂樹, 北彦彦, 林照峯: “読み込み時間を用いたWWWページのフィルタリング”, 情報処理学会グループウェア研究会, 研究報告 98-GW-29, p.37-42, 1998
- [4] PostgreSQL Inc. <http://www.pgsql.com/>