

L-13

フィンガープリントキャッシュ機構の Web ブラウザプラグインによる実装

木場 雄一 †

木村 康浩 †

吉井 謙一郎 †

關 俊文 †

† (株) 東芝 研究開発センター

1 はじめに

インターネット上の Web トラフィックを軽減する手段として Web キャッシュは欠かせない技術である。

近年、ポータルサイトや EC(Electronic Commerce) サイト、Web をフロントエンドとした企業内業務プログラムなどの Web の利用形態が増えてきている。これらのサイトのコンテンツは、CGI や JAVA サブレット等の機構を使ってアプリケーションプログラムを起動し、起動されたプログラムは必要に応じてデータベースから検索したデータを使って生成されるものが多い。このようなコンテンツを、我々は動的コンテンツと呼んでいる。動的コンテンツは、従来のキャッシュ技術ではキャッシュできない。

そこで我々は、従来のキャッシュと共に、動的コンテンツもキャッシュ可能にするフィンガープリントキャッシュ技術を開発した。

2 フィンガープリントキャッシュ

フィンガープリントキャッシュ技術 [1] は、コンテンツを一意に識別するフィンガープリントと呼ぶ識別子を与え、その識別子をキーとしてコンテンツを管理する技術である。

フィンガープリントとは、コンテンツ自身から MD-5 や SHA-1 などのハッシュ関数により一意に計算されたハッシュ値を指す。指紋 (fingerprint) が人間個人を識別できるように、フィンガープリントを用いることでコンテンツそのものを識別することができる。

フィンガープリントキャッシュ技術によってコンテンツを一意に定めることができるので、動的コンテンツがキャッシュヒットできるようになる。また、異なる URL のコンテンツでも、中身が同じであればキャッシュヒットできる。さらに、類似コンテンツの場合でも、フィンガープリントキャッシュを使って差分情報のみを転送することができる。

フィンガープリントキャッシュを実現するために、フィンガープリントとコンテンツを対応付けたテーブルを持つ。このテーブルによりフィンガープリントとコンテンツのユニークな対応を構築する。しかし、このテーブルだけでは従来のキャッシュには対応できない。そこで、更に URL とフィンガープリントを対応付けるテーブルも併用する。上述した 2 段のテーブルを持った装置はクライアントとサーバ間に 2 つ直列に用意する。そのうちオリジンサーバ側の装置はフィンガープリントを計算する。クライアントからのリクエストに対するリプライデータは、サーバ側の装置、次いでクライアント側の装置を経由してクライアントへ渡る。この際、サーバ側の装置からクライアント側の装置へ、生成したフィンガープリントを送る。この装置間のデータ転送量が、フィン

ガープリントキャッシュによって削減されるため、2 つの装置をサーバ側ネットワークとクライアント側ネットワークに置くことによって、クライアントとサーバ間のデータ転送量を削減できる。

我々は実現モデルとして、2 つのプロキシキャッシュからなるキャッシュ機構の実装を完了している [1]。2 つのプロキシは、1 つをオリジンサーバ側に置き (サーバ側プロキシ)、もう一方のプロキシをクライアント側のローカルネットワークに配置する (クライアント側プロキシ)。

しかし、デュアルプロキシ構成では、必ず 2 つのプロキシを設置する必要がある。そこで我々は、フィンガープリントキャッシュ導入の際に、導入先の環境により柔軟に対応するため、クライアント側プロキシを無くし、代わりに同等の機能を Web ブラウザのプラグインとして実装する方式を開発した。

3 プラグインを用いたフィンガープリントキャッシュ機構

3.1 プラグイン構成の設計

プラグインを用いた構成を図 1 に示す。プラグインを用いたフィンガープリントキャッシュ機構は、サーバ側プロキシと Web ブラウザのプラグインモジュールの 2 つで構成される。プラグインモジュールはフィンガープリントキャッシュを管理する。

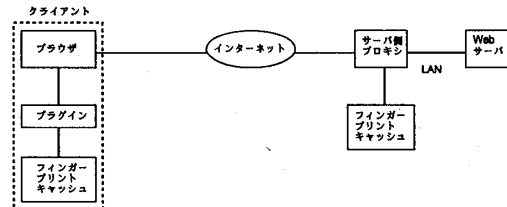


図 1 プラグインを用いた場合の構成

プラグイン構成での処理の流れを説明する。クライアントからのリクエストに対しては、プラグインモジュールは何も動作をしない。クライアントのリクエストはサーバ側プロキシを経由し、オリジンサーバへ送られる。サーバ側プロキシはリクエストを受けた際、通常のプロキシキャッシュと同じ動作をする。

サーバ側プロキシは、オリジンサーバからのリプライデータを受信すると、受信コンテンツのフィンガープリントを計算する。そして、計算されたフィンガープリントを元に、同一コンテンツがサーバ側プロキシのフィンガープリントキャッシュ存在するか調べる。サーバ側プロキシがクライアントに渡すリプライデータの形式は、3 種類に分けられる。

1. 通常転送
2. 圧縮転送
3. 差分転送

Implementation of fingerprint cache using web browser plugin.

Yuichi Koba, Yasuhiro Kimura, Kenichiro Yoshii and Toshibumi Seki.

Corporate R & D Center, Toshiba Corporation.

図2にそのモデルを示す。3種類全てには、プラグインモジュールを呼び出す命令と、クライアントが要求したコンテンツのフィンガープリントが、リプライデータのボディ部に埋め込まれる(図2中(a))。記述を解釈できるコンテンツはHTMLデータのみなので、プラグイン構成でのフィンガープリントキャッシュ対象コンテンツはHTML形式のみである。

1. の通常転送の場合、オリジナルコンテンツの中に図中(a)のデータが埋め込まれる。2. の圧縮転送の場合、HTTPボディ部にオリジナルコンテンツは付かず、図中(a)のデータのみ付加してクライアントへ送る。3. の差分転送の場合、上述共通付加データの他に、再利用するコンテンツのフィンガープリント値、オフセット、利用するデータ長といった差分情報と、差分圧縮にヒットしなかった部分のコンテンツ部分がHTTPボディ部に付加される。

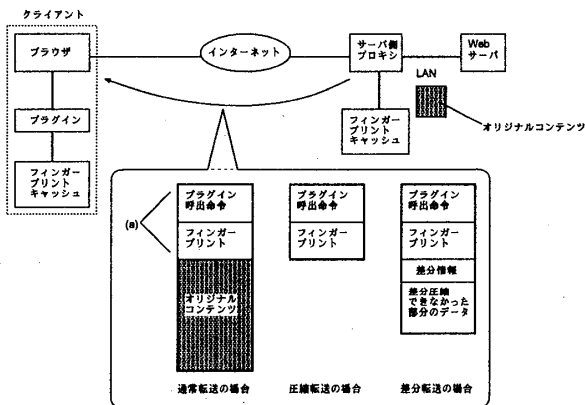


図2 プラグインに渡されるリプライデータ形式

ブラウザにデータが届くと、プラグインモジュールがブラウザへのリプライデータをフックする。プラグインモジュールはフィンガープリントや差分情報を解析し、クライアントが要求したコンテンツデータを生成する。そのコンテンツをブラウザへ渡してブラウザに表示される。プラグインモジュールにおけるフィンガープリントキャッシュテーブルは、フィンガープリントをファイル名とし、コンテンツと関連付けて保存・管理する。

3.2 プラグイン構成の実装

実装では、対象ブラウザに Internet Explorer を採用した。また、プラグインモジュールは ActiveX Control で実装した。ActiveX Control ではローカルディスクへの書き込みができるので、キャッシュファイルをローカルディスクに保存することができる。また、ActiveX Control は、Internet Explorer 上からスクリプトによって呼び出すこともできる。記述スクリプトは JScript を使い、サーバ側プロキシでリプライコンテンツに付加するようにした。更に、ActiveX Control は、HTML の OBJECT タグを使って、ブラウザ上からプラグインモジュールを自動インストールすることができる。サーバ側プロキシのサーバプログラムは、拡張性のある Web サーバアプリケーション Apache の拡張モジュールとして実装した。

4 性能評価

3.2の実装に基づき、プラグインを用いたフィンガープリントキャッシュ機構において、測定を行った。

今回の実験環境は、クライアントを配置する LAN から、インターネットへは 64 kbps の PIAFS 接続で繋がっている。サーバ側プロキシは PIAFS 接続の先に設置する。

クライアントは社内の Web サイトにアクセスし、サーバ側プロキシと、クライアント PC 間を流れるデータ転送量を計測した。サーバ側プロキシでは従来キャッシュの場合に、プロキシからクライアント PC へ転送されるデータ転送量を計算しておくことで上記の転送量と比較し、本技術によるデータ転送量削減の効果を調べた。図3に結果を示す。

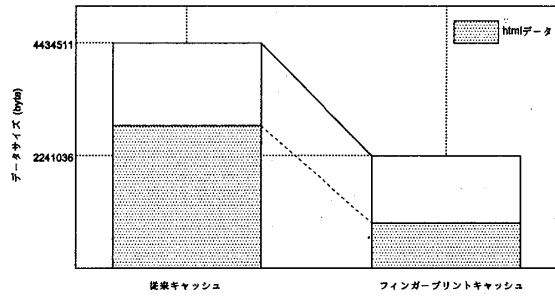


図3 データ転送削減量

その結果、フィンガープリントキャッシュの対象である HTML ファイルのネットワーク転送量が、従来キャッシュの場合に比べ約 1/4 に減った。全ネットワーク転送量で見ても、従来キャッシュの場合に比べ約 1/2 に減り、フィンガープリントキャッシュによってデータ転送量を削減できていることが確認できた。

5 おわりに

本稿では、我々が開発したフィンガープリントキャッシュ機構において、以前開発した方式では少なくとも2つの計算機が構築に必要だった点を改良するため、Webブラウザのプラグインを用いて計算機を1つ設定するだけでフィンガープリントキャッシュが可能になるような方式を提案し、実装した。

実装システムにおいて測定した結果、この方式でもフィンガープリントキャッシュの効果が得られることが判った。

今後、実際の環境での性能測定実験を行い評価をすると共に、フィンガープリント技術を使ったキャッシュ機構の更なる機能拡張をしていきたいと思っている。

参考文献

[1] 吉井謙一郎, 金井達徳, 關俊文, 吉田英樹, "フィンガープリントキャッシュと動的 Web コンテンツ配信への応用", 第4回インターネットテクノロジーワークショップ (WIT2001), September 2001.