フィンガープリントキャッシュ機構のシングルサインオンシステムへの対応

木場 雄一† 木村 康浩† 中山 大士‡ 庄野 篤司† 村松 孝治‡
佐藤 英昭† 岩崎 孝夫‡
† (株) 東芝 研究開発センター ‡東芝ソリューション (株)

1 はじめに

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 ASP・サー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 (動的コンテンツ) が多い．

インターネット上の Web トラフィックを削減する手段として Web キャッシュが広く用いられているが，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

従来の Web キャッシュは，URL 対コンテンツを一対一で対応させることで，同じ URL に対するリクエストに対しキャッシュからレスポンスを返してデータ転送量削減を実現している．しかし，動的コンテンツの場合は同じ URL でコンテンツ内容が書き換わるので一意に定まらず，キャッシュ効果が発揮できない．

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 (以降 FP) 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．FP 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．これにより，リクエスト先の URL とは関係なく，コンテンツが一致すればキャッシュ効果が出るようになる．

また，FP キャッシュを使った差分転送方式も独自に開発し，類似コンテンツにアクセスした際もデータ転送量を削減することができる．

本技術を用いるためには，Web サーバ側とクライアント側それぞれに FP キャッシュを実装した装置が必要である．我々はこれまでに，Web サーバ側・クライアント側ネットワークにそれぞれプロキシサーバを設置するデュアルプロキシモデル [1] と，Web サーバ側にプロキシサーバ，クライアント側は Web ブラウザのプラグインとして組み込むプラグインモデル [2] の 2 つのモデルに関して実装してきた．本稿では新たに実装したクライアント端末の中にクライアント側プロキシサーバを組み込んだ方式 (クライアントモジュールモデル) について説明する (図 1).

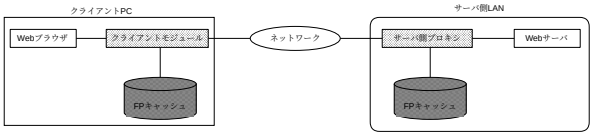

図 1 クライアントモジュールモデルの FP キャッシュ機構

サーバ側に配置する FP キャッシュは，フォワード型あるいはリバース型プロキシサーバとして設置される．クライアント側に配置する FP キャッシュ装置はクライアント PC 内に配置する．この構成はクライアントが携帯端末を経由してアクセスする場合に有効である．サーバ/クライアント側両方に設置する FP キャッシュ装置はそれぞれ FP キャッシュを保持し，以降利用者から同じ内容のコンテンツのリクエストを処理する際に，サーバ側 FP キャッシュ装置から下流へのレスポンスではコンテンツは送らずに FP だけを返す．クライアント側 FP キャッシュ装置では受け取った FP から対応するコンテンツを取り出して Web ブラウザに返す．これにより両 FP キャッシュ装置間のデータ転送量を削減できる (FP 転送).

また，以前取得したコンテンツと類似したコンテンツの処理の場合は，サーバ側 FP キャッシュ装置がベースとなるコンテンツ (再利用されるコンテンツ) の FP と差分情報だけをクライアント側に返すことによって両装置間のデータ転送量を削減できる (差分転送)．クライアント側 FP キャッシュ装置は，差分情報と FP からコンテンツを復元・生成し．Web ブラウザに返す．

クライアントモジュールモデルは，デュアルプロキシモデルのクライアント側 FP キャッシュの配置が変わる以外は同じ構成である．我々は，クライアントモジュールモデルに対して SSL 通信に対して本技術を対応する点と，コンテンツ書き換え型シングルサインオンシステムに対応する点に対するモデルの改良と検討を行った．

3 コンテンツ書き換え型シングルサインオンシステムへの対応

3.1 コンテンツ書き換え型シングルサインオンサーバ

シングルサインオンは，1 度の認証で複数のシステムのサービスを受けるようにできる機能である．これは，複数の Web サーバに対する認証を 1 箇所で行うようにしており，この機能を持つアプリケーションには，複数のサーバを一つの仮想サーバのサブディレクトリとして見せて提供しているものもある．そのために，シングルサインオンサーバでは以下のような処理を行っている．

1. Web ブラウザからのリクエスト中の URL を本来のアドレスパスに書き換える
2. Web サーバからのレスポンス中に本来のアドレスが含まれる場合，仮想の URL へ書き換える

Development of Fingerprint-Cache for Single Sign-On system.
Yuichi Koba †, Yasuhiro Kimura †, Hiroshi Nakayama ‡, Atsushi Shono †, Kouji Muramatsu ‡, Hideaki Sato † and Takao Iwasaki ‡.
† Corporate R & D Center, Toshiba Corporation.
‡ Toshiba Solutions Corporation.

3.2 問題点と適用方式

従来の FP キャッシュでは，サーバ側 FP キャッシュ装置からクライアント側に送られる差分データは，独自のバイナリ方式であるため，図 2 のように中間にコンテンツ書き換え型シングルサインオンサーバが存在すると，本来書き換えられるべきデータが書き換えられないことになる．更に，差分元のコンテンツのどの部分を使用するかという情報をバイト単位で指定しているため，サーバ側 FP キャッシュ装置で生成した差分データに従ってクライアントモジュール側で復元すると，書き換え後のベースコンテンツを FP キャッシュとして再利用するために，バイト数がずれてしまう．そのため，Web ブラウザ上に正常に表示されない問題が発生してしまう．

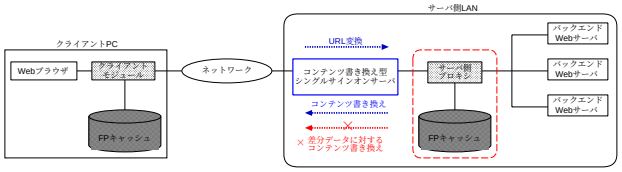

図 2 シングルサインオンサービス環境における FP キャッシュ導入時の問題点

この問題を解決するために，我々は差分データ形式を変更することで対応した．まず，差分データをテキスト化することで，シングルサインオンサーバでのコンテンツ書き換えを可能にする．元々は差分データのコンテント形式は octet-stream/x-fp-compressed という独自バイナリ形式で送っていたが，コンテント形式も元のコンテンツと同じ形式で差分データを送るように変更した．さらに，ベースコンテンツのどの部分を再利用するかの指定を，バイト単位から行単位に変更することで，行の追加・削除がない限り，コンテンツ書き換え後のベースコンテンツで利用しても元のコンテンツに復元できるようにした．

4 SSL 対応

Web システムにおいて，SSL(Secure Sockets Layer) プロトコルを用いて HTTP データを暗号化してセキュアな通信を実現する HTTP over SSL (HTTPS) が事実上の標準になっている．社内システムを外部からアクセスする場合など，HTTPS による Web アクセスが一般的に用いられており，特にシングルサインオン環境ではセキュアなシステムとするために必須となる．

4.1 SSL 対応の問題

HTTPS が用いられている場合は，データは SSL による暗号化が施されている．この時にプロキシサーバによる Web キャッシュを行うためには，SSL で暗号化されたデータを一旦復号し，暗号化前の元の HTTP データを取得して扱うようにしなくてはならない．

サーバ側キャッシュサーバがフォワード型の場合，Web ブラウザと Web サーバ間の通信データをトンネリングするのが一般的な動作なので，通信データを解析して差分情報を生成する我々の技術は使うことができない．

4.2 SSL 対応への適用方式

そこで我々は本技術を適用するための方式を考案した．サーバ側キャッシュサーバはリバース型で動作することを前提とし，クライアントからの SSL 通信はクライアントモジュールが代行する方式である．

この方式で動作させるために，以下のことを行った．

1. Web ブラウザとクライアントモジュール間は http 通信とする
2. クライアントモジュールは，https で通信する必要がある URL のリストを管理し，そのリストに該当する URL へのリクエストを受け取った際には，https で上流のサーバと通信する．
3. クライアントモジュールは，上記 URL リストを参照して，Web ブラウザからのリクエストデータや Web ブラウザへのレスポンスデータ中の URL 変換を行う．これは，リクエストデータの場合，「http」という文字列を「https」と書き換えることで上流サーバと https による通信を可能にする．また，レスポンスデータの場合，データ中に「https」で始める他の Web サーバへのリンクが含まれている場合に，そのリンク先の URL を「http」から始めるリンクに書き換える．
4. Web ブラウザ並びにクライアントモジュールを入れている PC から実際に上流にリクエストが流れる際に SSL 通信で行っていることを利用者に視覚的に示す (アイコン表示等).

上記方式に基づいて実装を行い，コンテンツ書き換え型シングルサインオンシステムの一つに組み込んで動作試験を行ったところ，我々の FP キャッシュ効果を出し，かつ SSL 通信ができることを確認した．

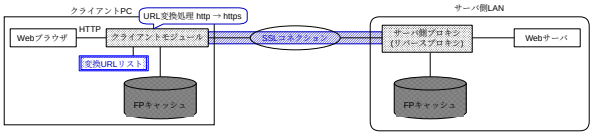

図 3 SSL 通信への対応

5 おわりに

本稿では，我々が開発した開発した FP キャッシュを，コンテンツ書き換え型シングルサインオンシステムに適用する方法を検討し，考案した方式を開発した．今後は本方式の動作検証や性能評価を行い，スケーラビリティや応答性能などを定量的に把握して，性能限界を明確にしていく予定である．

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.

[2] 木場雄一, 木村康浩, 吉井謙一郎, 關俊文, ”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”, FIT(情報科学技術フォーラム) 2002, September 2002.