

Web 高速化技術のキャッシュ量削減方式

Cache size reduction in Fingerprint cache system

石丸 大† 庄野 篤司† 佐藤 英昭†
Dai Ishimaru Atsushi Shono Hideaki Sato

1. はじめに

近年、インターネットの多様化がすすみ、CGIやASPを利用して動的にコンテンツを作成する技術も多くのサイトで使用されるようになっていいる。一方、Webを見る端末もPCに限らずPDAや携帯電話といったモバイル機器に広がっている。ところでWeb高速化手段の1つである従来のプロキシキャッシュではアクセスの度に生成される動的コンテンツには有効でなかった。そこで、我々は動的コンテンツをキャッシュの対象としてデータ転送量を削減するフィンガープリントキャッシュ[1] (以後FPキャッシュ)を開発してきた。さらに比較的遅い回線を使用しているモバイル機器上でのFPキャッシュの実現を検討してきた[2]。モバイル機器はデータメモリ容量も比較的少ない。そこで本稿ではFPキャッシュに有効な、新しいキャッシュコントロール方法を提案する。実際にこのキャッシュコントロール方法を用いたFPキャッシュを、サーバ側プロキシとクライアントとなるPDA上に実装して、その評価を述べる。最後に、今後の課題について言及する。

2. FP キャッシュ

FP キャッシュは、WEB アクセスを高速化するために開発されたキャッシング技術である。従来のプロキシキャッシュでは、URL とコンテンツを一対一で対応させて保存している。既に保存されている URL に対するリクエストがあると、その URL に対応するコンテンツをキャッシュから返すことでデータの転送量を削減させている。この方法は、静的コンテンツには非常に有効であるが、URL 対コンテンツが一意に定まらない動的コンテンツには有効でなかった。

一方、FP キャッシュでは、コンテンツを一意に識別するフィンガープリント (以後 FP) とコンテンツの対を保存する。FP とはコンテンツのハッシュ値であり、コンテンツに対して一意に定まる識別子である。本技術を用いるには、Web サーバ側・Web ブラウザ側にそれぞれに FP キャッシュを実装した装置が必要である(図 1)。Web サーバ側の装置 (装置 1) と Web ブラウザ側の装置 (装置 2) の両方に FP キャッシュを保持し、もし Web サーバから以前取得したコンテンツが装置 1 に届くとコンテンツの代わりに FP だけを装置 2 に送る。装置 2 は該当 FP と対応付けられているコンテンツを Web ブラウザに返すことで装置 1 ~ 装置 2 間のデータ転送量を削減 (FP 転送) する。また、以前取得したコンテンツと類似したコンテンツの場合は、装置 1 が保存しているコンテンツの FP とそのコンテンツとの差分情報だけを装置 2 に送ることでデータ転送量を削減する (差分転送)。この場合、装置 2 は差分情報と FP か

らコンテンツを復元して、Web ブラウザにコンテンツを返す。

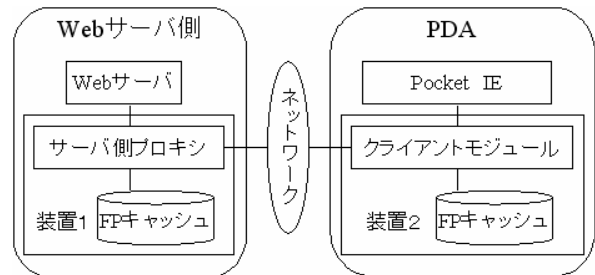


図 1 FP キャッシュの基本構成

我々は FP キャッシュ技術を PDA 上に実装した[2]。しかし、PDA はデータメモリ容量 (補助記憶領域) の制限があるという問題が生じた。そこで、少ないキャッシュ容量を効率よく使用する仕組みが必要となった。

3. 圧縮率を用いたキャッシュコントロール

3.1 従来のキャッシュコントロール方式

従来の FP キャッシュでは、差分転送した全てのコンテンツを差分伸長後に保存していた。そしてキャッシュ領域が一杯になると古いキャッシュの削除が行われていた。このため、装置 2 が PDA のようにキャッシュ容量が十分でないマシンの場合、差分転送を行おうとした時に、差分のベースとして再利用しようとしたコンテンツが既にキャッシュから削除されている確率が高くなるという問題があった。

3.2 提案するキャッシュコントロール方式

差分転送における差分データ量は本来のコンテンツサイズより小さい。ここで圧縮率を、

圧縮率 = $1.0 - \text{差分データ量} / \text{コンテンツサイズ}$ と定義する。圧縮率が高いコンテンツとは、差分情報が少ない、即ち、キャッシュ内にそのコンテンツと類似したデータを保持しており、そのデータからコンテンツのより多くの部分を復元することが可能であることを示す。このようなコンテンツを保存すると、キャッシュ内に類似したデータが重複することになり、効率が悪いと考えられる。そこで、ある閾値を設定して、その閾値より圧縮率が高いコンテンツは保存しないこととする。この判定は装置 1 に行い、判定結果を装置 2 に伝える。装置 1 および装置 2 では、この判定結果情報によりコンテンツを保存するか否かが決定される。このように圧縮率の高いコンテンツを保存せず、圧縮率の低いコンテンツだけを保存することでキャッシュ領域の有効利用を行う。

† 株式会社 東芝 研究開発センター,
Toshiba Corp. Research and Develop Center

4. 評価と考察

実際にサーバ側に圧縮率判定部を実装し、クライアントに PDA を用いたシステムを構築し、本キャッシュコントロール方式の評価を行なった。評価は、評価用コンテンツにおいて、設定する閾値の違いによる利用キャッシュ量と圧縮率の変化を測定するという方式で行なった。また一般サイトを巡回した時の圧縮率を測定し、その結果についても考察を述べる。

4.1 評価用コンテンツによる評価

評価用コンテンツとして、直前のコンテンツと一部が同じでその他がランダム文字列であるコンテンツ群を用いた。なお直前のコンテンツと一致する割合を類似度と定義する。1コンテンツは 20000Bytes(40文字 x500行) + (ヘッダーなどの付属情報)である。また、類似度に関して 60%、80%の固定値と、60-100%、80-100%の間のランダム値の、合計 4 種類のコンテンツ群を用いた。各コンテンツ群に対してそれぞれ 100 回の読み込みを 1 試行とし、1 試行における平均した圧縮率を Total 圧縮率と呼ぶ。なおキャッシュ容量は十分にあり、1 試行中にキャッシュの削除は行われていない。

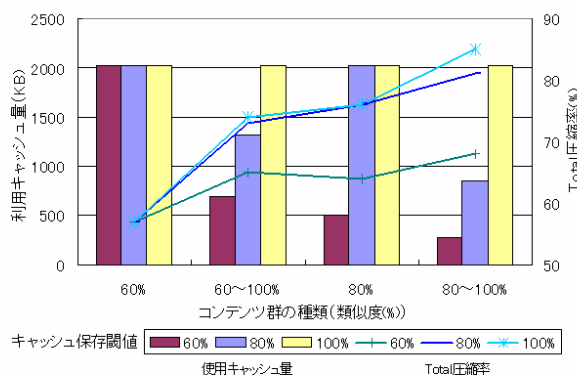


図 2 閾値とキャッシュ量・Total 圧縮率の関係

各コンテンツ群における 1 試行の結果を図 2 に示す。

設定閾値が 100% の時 (従来キャッシュコントロール方式時と同等な状態時) はコンテンツに関らず全て保存するためキャッシュ量が大きくなる。しかし過去のコンテンツをすべてベースコンテンツとして利用できるため、圧縮率は高い値になる。一方、閾値を低く設定すると類似度が閾値より高いコンテンツは保存しないためキャッシュ量は減る。しかし差分転送のベースとして利用できるデータが減り、コンテンツとベースとの相違部分が増えるため、結果として、圧縮率は低い値になる。キャッシュ量を減らして圧縮率を高くすることが理想であるが、これらは設定閾値に対してトレードオフの関係にある。

さらに類似度より閾値が大きい場合、すべてのコンテンツをキャッシュすることから、最適となる閾値はコンテンツの類似度にも依存することが分かる。

本実験結果では、たとえば閲覧するサイトの類似度が 60-100% である場合、閾値を 80% に設定した新キャッシュ

方式を用いることで、圧縮率はほぼ変わらずにキャッシュ量を約 4 割削減できるという優れた効果が得られることが示された。

4.2 一般サイトにおける圧縮率の検証

本キャッシュアルゴリズムを実装した PDA を用いて、一般的なニュースサイトおよびグループウェアサイトを巡回した時の圧縮率について検証する。閲覧は、サイト A 内のコンテンツを巡回した後、サイト B 内を巡回し、さらにサイト C 内を巡回した後、再びサイト A 内を巡回するという手順で行った。閾値は 80% および 100% とした。

まず、閾値を 80% にした場合は、Total 圧縮率は 86.3% であった。巡回中にキャッシュ領域の飽和は起きず、巡回終了時にもキャッシュには空きスペースが残った。

一方、閾値を 100% にした場合は、サイト C を巡回中にキャッシュ領域が一杯になり、Total 圧縮率は 78.7% という低い値になった。

キャッシュに容量が十分にある場合、閾値が低いと圧縮率は下がることが一般的にいえる (前節参照)。しかし、キャッシュ容量に余裕がない場合、閾値を低く設定すると圧縮率が上がることがある。閾値が低いと、保存するコンテンツ数が減って使用するキャッシュサイズが減り、キャッシュが一杯になるまでの時間が延びる。これにより、古いデータが削除されにくくなり、必要に応じて再利用することが可能となるために、コンテンツへのアクセスパターンによっては圧縮率が向上する。

上記の場合では、2 巡目のサイト A 内巡回時に 1 巡目のサイト A のデータを再利用できるかどうか圧縮率に影響を及ぼし、閾値 100% の時よりも閾値 80% の時に Total 圧縮率が高い値になった。またキャッシュ使用量も閾値 80% の時の方が少なくなった。

このようにコンテンツの閲覧パターンに対して適切な閾値を設定できた場合には、キャッシュ使用量が下がり、なおかつ、圧縮率が上がる場合があることが分かった。

5. 今後の課題

本稿では、キャッシュ容量に制限のある PDA で、新キャッシュ削減方式を用いた FP キャッシュの紹介・評価を行い、従来のキャッシュよりも少ないキャッシュ量で、圧縮率は従来とそれほど変わらない効率の良いキャッシュができることを示した。

しかしながら閾値の最適値は閲覧するコンテンツに大きく依存するため、必ずしも閾値 80% の設定が最適解であるわけではない。今後の課題として、閾値をどのように設定するか、たとえば閲覧するコンテンツに合わせて閾値を動的に決定するなど、閾値の設定方法について検討・検証することが挙げられる。

参考文献

[1] 吉井謙一郎他, “フィンガープリントキャッシュと動的コンテンツの配信への応用”, 第 4 回インターネットテクノロジーワークショップ

[2] 石丸大他, “WEB アクセス高速化技術の PDA への応用”, pp. 3-249, 第 66 回全国大会, 情報処理学会