

# データサイズを考慮したキャッシュ置き換え アルゴリズムの提案と検証

三野敦史† 岸本克巳‡ 長田智和‡ 谷口祐治§ 玉城史朗†

†琉球大学工学部情報工学科 ‡琉球大学理工学研究科 §琉球大学総合情報処理センター

## 1 はじめに

WWW 関連プロトコルのトラフィック量の増大は、ネットワークトラフィック量の増加、WWW サーバへのアクセス集中による過剰負荷、それに伴うレスポンスタイム低下などの問題を引き起こしている。これらの問題の原因として、同じページを複数の人が閲覧することにより生じる WWW ページデータの重複が挙げられる。この重複を減らす方法には、ページデータの保持・再利用が可能な WWW キャッシュサーバの利用が有効であると考えられる。

本研究では、WWW キャッシュサーバでページデータの格納・管理に利用されるキャッシュ置き換えアルゴリズムに着目し、より効果的な手法を提案する。

## 2 既存手法

### 2.1 WWW キャッシュサーバ

WWW キャッシュサーバとは、WWW サーバとクライアントとの間に存在し、WWW データ通信を仲介するサーバである。キャッシュサーバはページデータを保持（キャッシュ）し、それを利用することにより外部ネットワークへのトラフィック量の軽減、レスポンスタイムの向上を実現している。その動作を図 1 に示す。

キャッシュサーバとして動作するプログラムには、squid[1], apache, CERN サーバ, DelGate サーバなどがある。

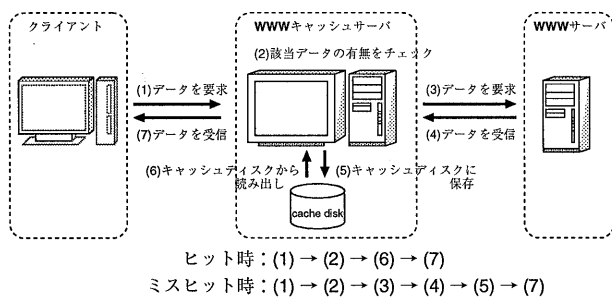


図 1: WWW キャッシュサーバの動作

A Proposition and Verification of "Cache Replacement Algorithm" Considering of the Data-size of WWW Page  
 Atsushi MINO† Katsumi KISHIMOTO‡ Tomokazu NAGATA‡ Yuji TANIGUCHI§ Shirou TAMAKI†  
 †Department of Information Engineering, Faculty of Engineering, University of the Ryukyus  
 ‡Graduate School of Science and Engineering, University of the Ryukyus  
 §Computing and Networking Center, University of the Ryukyus

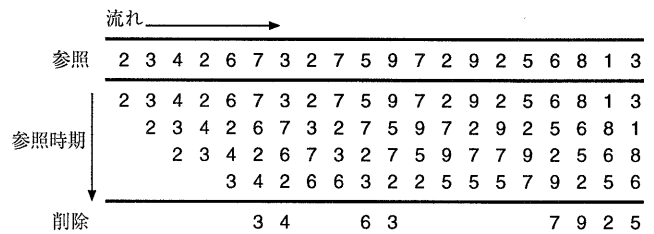


図 2: LRU の動作

### 2.2 LRU(Least Recently Used)

WWW キャッシュサーバでは、一般的にキャッシュ置き換えアルゴリズムに LRU(Least Recently Used) が用いられる。LRU は参照時期を基準にし、データ削除は参照されてから最も時間が経過しているものを対象に実行される。その動作を図 2 に示す。

### 2.3 性能評価基準

WWW キャッシュサーバの性能評価には、クライアントが要求した WWW ページがキャッシュに存在する確率を表すヒット率が用いられるが、本研究ではバイトヒット率を提案し、評価に利用する。

## 3 提案手法

### 3.1 既存手法の問題点

LRU は WWW ページのデータサイズを考慮しないが、外部ネットワークへのトラフィック量の軽減度合やレスポンスタイムの向上度合はデータサイズによって変化するため、ヒット率で判断しにくいという問題点がある。例として、ヒット率が同じ場合、外部ネットワークへのトラフィック量の軽減度合は、ヒットした総データサイズが大きい方が上まわるということが挙げられる。またミスヒット時に、データサイズの大きなページやネットワーク的に遠いサーバへ取得しに行くこと、レスポンスタイムが急激に悪化するという問題点もある。

最初の問題点の解決策として、本研究では性能評価基準にバイトヒット率を用いる。バイトヒット率とは、クライアントからの全要求データサイズに占めるヒットした WWW ページの総データサイズの割合を表し

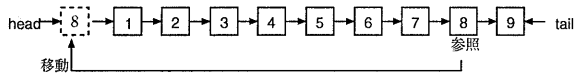


図 3: LRU のリスト構造

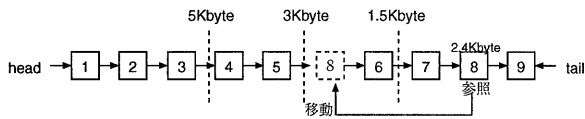


図 4: 提案手法 1 のリスト構造

たものである。また次の問題点の解決策として、次の二つの手法（提案手法 1・提案手法 2）を提案する。

### 3.2 提案手法 1

LRU では、取得してきた WWW ページとキャッシュヒットした WWW ページは無条件に参照順リストの先頭に位置付けられる（図 3）。これはページの管理が同じ優先順位の下で行われていることによるものである。そこで本手法では、ページのデータサイズと WWW サーバからのレスポンスタイムを基準に優先順位を付け、それに基づき参照時の位置決めを行う（図 4）。優先順位は、ページのデータサイズが大きく、サーバからのレスポンスタイムが遅いほど高くなるように設定する。さらに参照時には、その優先順位が高いページほど参照順リストの先頭の方に位置するようにし、リストに滞在する期間を調整する。

### 3.3 提案手法 2

クライアントからのページデータ取得要求を受け取り、処理を行う外部プログラムを提案する。要求を受け取ると、その要求をキャッシュサーバに渡すとともに自身も WWW サーバに要求を出しヘッダー情報を取得する。次に、その情報からデータサイズの大きさを取り出し、要求頻度とともに統計データとして保持し優先順位付けを行う。この優先順位に基づきデータの再要求を行うことで、該当ページがキャッシュサーバに滞在する期間を調整する。

### 3.4 比較

提案手法 1・2 ともページデータに優先順位付けを行い、キャッシュディスクに占める優先順位の高いページの割合を大きくし、レスポンスタイムの向上を図る手法である。手法 1 は、データサイズとレスポンスを基準に優先順位付けを行い、実装はキャッシュサーバプログラムの改造により行う。それに対し手法 2 はデータサイズと要求頻度を基準にし、キャッシュサーバから独立して動作する外部プログラムを作成することで行う。

## 4 WWW ページデータ

附属小・中学校の授業におけるアクセスログ [2] からの WWW ページとデータサイズの分布図を図 5 に示す。クライアントから要求のあった WWW ページの平均データサイズは約 5.5KByte であり、そのうちデータ

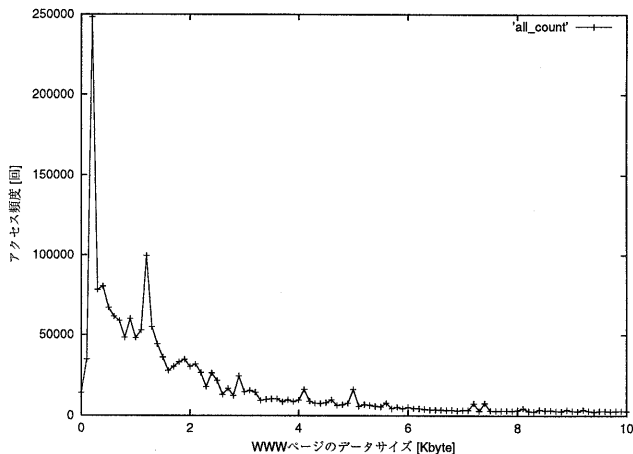


図 5: WWW ページデータサイズ分布

サイズが 1.5KByte 以下のページが全体の 50% を占め、さらに 5KByte 以下では 80% を占めた。残りの 20% は、5KByte から最大で 31MByte まで分散したページ要求があった。以上の結果より附属小・中学校で検証する場合は、外部プログラムの優先順位付け基準値を WWW ページのデータサイズが 5KByte 以上とすることで、全体的なレスポンスタイムの向上を効果的に行える。

## 5 まとめ

本稿では、データサイズを考慮したキャッシュ置き換えアルゴリズムを提案した。

今後は、次項について研究を進める。

- 提案手法の検証
- 附属小・中学校の授業におけるアクセスログを元にした改良
- 先読み機能の追加
- アクセスの少ない時間帯における一括先読み機能の追加
- 各種パフォーマンス評価
  - － 外部ネットワークへのトラフィック量の軽減度合
  - － レスポンスタイムの向上度合

## 参考文献

- [1] Squid Web Proxy Cache, <http://www.squid-cache.org/>
- [2] Katsumi KISHIMOTO, Tomokazu NAGATA, Yuji TANIGUCHI, Shinji KONO, Shiro TAMAKI "Mobile Agent based Web Cache Proxy System," THE FIRST IEEE ELECTRO/INFORMATION TECHNOLOGY CONFERENCE, 2000