

移動計算機の情報発信環境における 階層型データのためのキャッシュシステム

稲田 文武[†] 野口 大介^{†,*} 田頭 茂明[†]
最所 圭三[†] 福田 晃[†]

現在、画像や動画といったマルチメディアデータの利用が急速に増えており、それらを世界に向けて発信する手段の一つとして WWW (World Wide Web) が普及している。我々は WWW を用いて移動計算機からマルチメディアデータを発信するためのシステムを提案している。本研究では移動計算機の通信の問題である帯域の狭さと切断に対処するため、階層構造を持ったデータに対し、WWW サーバ上でその階層ごとの優先度に応じた蓄積を行なえるキャッシュシステムを提案する。これにより、分断時や帯域が狭い場合でも最低限の品質を保証するデータの中心部分を利用者に提供できる。

A Cache System for Layered Data on the Information Announcement System in Mobile Computing Environment

FUMITAKE INADA,[†] DAISUKE NOGUCHI,^{†,*} SHIGEAKI TAGASHIRA,[†]
KEIZO SAISHO[†] and AKIRA FUKUDA[†]

Today, the WWW (World Wide Web) can be used to transmit multimedia data to the world. We have proposed the system which uses the WWW to transmit them from mobile computers. In this paper, we propose a new cache system which can cache layered multimedia data on the WWW server in order to deal with low bandwidth or disconnection between a mobile computer and the server. The cache system has the capability of providing with users the part of data which has the lowest quality even if a mobile computer is connected through low bandwidth communication pathes or disconnected.

1. 背景と目的

WWWはインターネットとともに一般に普及し、画像や動画等のマルチメディアを発信する一つの手段となっている。また、半導体技術等の進歩により移動計算機の小型化、軽量化が進み、将来的にはより携帯性が高まると予想される。そして、これらを組み合わせた移動計算機を用い、マルチメディアデータを送受信するための手法も提案されている¹⁾²⁾。また、移動計算機で取り込んだマルチメディアデータを WWW を用いて発信することも可能であり、我々はそのためのシステムを開発中である³⁾。

現在、移動計算機で通信を行なう場合、移動性を考

慮して PHS や無線 LAN 等を用いることが多い。一般にこれらの通信回線は、切断が起りやすく、比較的帯域が狭い。また、電源等の関係上、移動計算機自身も常に通信可能な状態であるとは限らない。このため、WWW を用いて移動計算機から情報発信する場合、通信帯域の狭さと切断という2つの問題を考慮しなければならない。

通信帯域の狭さの問題を解決する一般的な手法の一つとして、キャッシュやミラーを用いる方法があり、移動計算機においてもこの手法を用いることが可能である。この場合、移動計算機の情報を管理するサーバ(ここでは WWW サーバと呼ぶ)を置き、そこに移動計算機から発信されるデータを全てミラーする方法や、過去に発信したことがあるデータをキャッシュする方法等が考えられる。しかし、単にミラーする方法では、移動計算機が保持するデータと同じ容量のディスク資源が WWW サーバ上に必要となる。この場合、WWW サーバがミラーの対象とする移動計算機の台

[†] 奈良先端科学技術大学院大学 情報科学研究科
Graduate School of Information Science, Nara Institute
of Science and Technology

^{*} 現在、株式会社デンソー

Presently with Denso Co. Ltd.

数が少ない場合、WWWサーバはそれらが保持するデータを全て保存することが可能であるが、移動計算機の台数が多くなった場合、不可能となる。一方、WWWサーバでキャッシュを行う場合は、キャッシュに全てのデータが常時存在するわけではないので、移動計算機の切断時に提供できないデータが発生する。

データが完全な形で提供されない場合でも、そのデータの一部でも提供できれば、利用者にとって有益な場合がある。どの部分が必要かはデータや利用者によって異なる。完全な形でのデータが必要な場合もある。画像データや動画データ等のマルチメディアデータの場合、データの一部でも利用者にとっては有益な情報となる。これらのデータは品質を落とすことにより、データ量を削減できる。更に、第3.2節で示すように、階層的なマルチメディアデータに関しては、低位の階層である中心部分をWWWサーバにキャッシュすることにより多くの移動計算機上のデータをキャッシュでき、移動計算機の通信回線が切断された場合でも利用者に低品質であるが意味を持ったデータを提供できる。また切断状態でない場合でも、WWWサーバにキャッシュしている部分を移動計算機から転送しなくて済むため、その分の転送量を削減できる。

本稿では、階層構造を持つマルチメディアデータを対象とし、その階層毎のキャッシュをWWWサーバに用意することにより、移動計算機が切断されている場合、完全な形でデータをキャッシュする方式に比べ、多くのデータを提供でき、またアクセス頻度が高いデータに対しては高品質なデータを提供できるようなキャッシュシステムを提案する。

2. 移動計算機におけるWWW情報発信環境

現在、我々は移動計算機からの情報発信のためのシステムを構築している³⁾。このシステムでは、利用者に移動計算機の場所や接続状態を意識させないで、移動計算機からの情報を提供することを目的としている。

図1に本システムの概要を示す。このシステムではWWWサーバは移動計算機の接続の状態、位置等の管理情報及び、移動計算機上の発信の対象となっている全てのデータの複製を保持している。WWWサーバに管理情報を保持することにより、データの受け手である利用者は、移動計算機の場所や発信形態を意識することなしに、移動計算機上と同じデータ又は少し古いデータを得る事ができる。

移動計算機の状態として、接続状態と分断状態がある。またデータを、蓄積することで再利用可能な蓄積型リソースと、リアルタイムデータのように一度しか

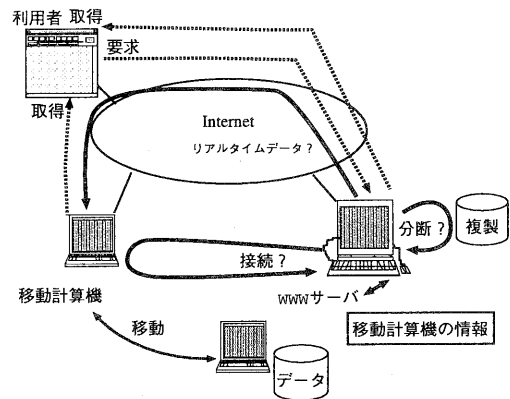


図1 情報発信システムの概要

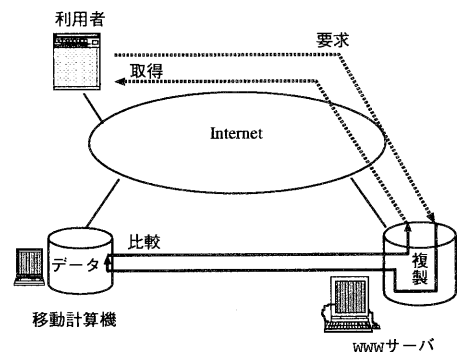


図2 接続時蓄積型情報発信

利用できない非蓄積型リソースに分類している。本研究では扱うデータとして特に、キャッシュが有効である蓄積型リソースを対象とする。

本システムでは移動計算機の接続状態により、蓄積型リソースの発信形態が異なる。まず、接続時の蓄積型情報発信の概要を説明する。図2にその時の制御及びデータの流れを示す。情報の受け手である利用者からWWWサーバに移動計算機上のデータの要求が到着した場合、移動計算機上のデータとWWWサーバ上の複製を比較する。移動計算機上で更新が行われていれば、WWWサーバは移動計算機からデータを取得し、複製を更新した後、利用者にデータを送信する。更新されていなければ、WWWサーバ上の複製を利用者に送信する。また、WWWサーバ上の複製の更新方式として他に、移動計算機上のデータが更新された際にWWWサーバ上の複製を更新する方式も用意している。

次に、分断時の蓄積型情報発信の概要を説明する。図3に示すように、分断時は移動計算機からデータを

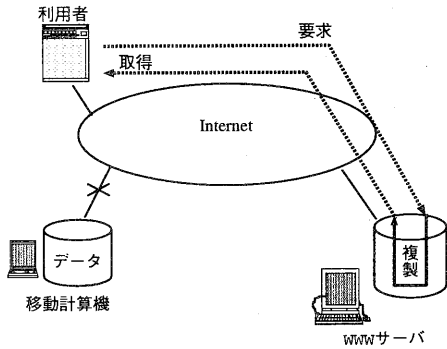


図3 分断時蓄積型情報発信

発信することはできないので、WWWサーバは接続時に作成された移動計算機上のデータの複製を発信する。

移動計算機の移動に対処する機構として、代表的なものは MobileIP⁴⁾ である。これは IP 層で計算機の移動を考慮した機構であり、移動の際の分断状態時に何らかの情報を発信するということはない。これに対して、本 WWW 情報発信システムはアプリケーション層で移動に対処しており、分断時にもアプリケーションに適した処理が可能となる。データが蓄積型リソースの場合、WWWサーバにデータを蓄積することで利用者に対してデータを提供し続けることができる。

本研究におけるキャッシュシステムは、この WWW 情報発信システムにおける複製機構に代わるものである。

3. 階層型データのためのキャッシュシステム

3.1 マルチメディアデータの発信

第2章で述べた WWW 情報発信システムを用いて通信を行う場合、分断が生じた時でも、WWWサーバは利用者に対し何らかの情報を提供する。データが蓄積型リソースの場合、分断する前やデータの更新時に、移動計算機上にあるデータの全てを WWWサーバに送っておく。そのために、移動計算機から発信するデータを保存するための領域と同じ容量の領域を WWWサーバに用意しなければならない。この方法の利点として、分断時でも全てのデータを WWWサーバから利用者に対して提供できることがあげられる。しかし、WWWサーバには移動計算機上のデータ量と同じだけの容量のデータ保存領域が必要となる。

また、他に考えられる方法として、WWWサーバに用意する保存領域を限定しておき、移動計算機の情報発信者がその容量に収まるように、移動計算機から

発信するデータ量を抑える方式が考えられる。そして、分断時に発信するデータを WWWサーバに送る。この方法は WWWサーバに用意できるデータ保存領域が限られている場合に有効である。しかし、情報発信者に対して発信できる情報の量を制限する等の制約を課す。

そこで3番目の方法として、データの品質を落として、WWWサーバに格納することを提案する。発信されるデータには、プログラムファイルや圧縮ファイル等のように変化させることができないデータと、静止画や動画等のマルチメディアデータのように品質を低下させることでデータ量を少なくすることが可能なデータに分別できる。例えば、静止画像データの場合、解像度を小さくしたり、使用する色数を少なくするなどして画質を落としても、利用者はその画像の意味を認識することが可能である。また動画データ場合、画質を落としたり、フレームを一定間隔で省くことで、データ量を削減することが可能である。一方、品質を利用者で変化させることが不可能なデータに関しては、従来通り、複製を用意する。

この方法により、WWWサーバでのデータの保存領域を削減できる。そして分断時には品質は落ちるけれどもデータを提供することができる。これは、マルチメディアデータが増加している状況を考慮すると、有効な手段である。

3.2 階層型マルチメディアデータ

本稿で提案するキャッシュシステムを説明する前に、対象とするマルチメディアデータの構造について説明する。マルチメディアデータとしては動画データや画像データ等が挙げられるが、それぞれに性質が異なる。本研究ではこれらのデータのうち、以下のような階層構造を持ち、それぞれの階層を分離できるものを扱う。

- 最も低い階層は、単独でそのデータの中心部分である低品質な部分を提供できる。
- これに上位の階層を順々に加えていくことにより、より高品質な部分を与えることができる。
- 最上位層まで加えると完全な情報を与えることができる。

この性質を利用し、要求された品質に応じた階層までのデータを利用者に提供することで、データ量も削減できる。プログレッシブ JPEG で圧縮された画像データは、この典型例である。このようなデータに対して、低い階層ほど高い優先度を与え、処理や転送等を優先して行うことでより多くのデータを提供できる。

図4の場合、階層1がデータの中心部分であり、最も高い優先度を持つ。また、階層4が最も低い優先

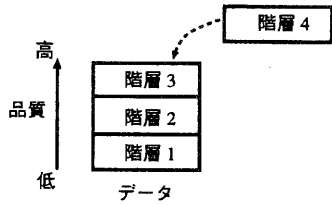


図4 階層型データ

度となる。そして現在、階層3までが存在しており、それに階層4のデータを付け加えることにより、このデータは完全なデータとして再現できる。

このような構造を持つデータを以降、階層型データと呼ぶ。

3.3 階層型データのためのキャッシュシステム

3.3.1 キャッシュの構成

第3.2節で述べた階層型データを効率よく利用者に提供するために、データの階層構造を利用したキャッシュシステムを提案する。分断時にできる限り多くのデータを利用者に与えるようにするためには、階層の低いデータをより多くWWWサーバに蓄えておかなければならない。

また、接続時においても、全てのデータを移動計算機から利用者に提供するか、キャッシュに残存するデータのみを提供するかを選択できるようにすることで、通信量を削減することが可能である。この選択を利用者の要求に応じて行ったり、移動計算機が接続されている通信回線の速度に応じて行うなどすることにより、より柔軟な情報提供が可能となる。

このキャッシュシステムにおいては図5に示すような階層毎のキャッシュを持つ。階層の数や階層毎の容量はWWWサーバの状況や画像の品質の提供方法によって設定できるようにする。階層毎のデータ量が一定ではないので必要な階層毎のキャッシュの容量を特定することはできないが、階層型データの特性上、ある階層のデータは最下層以外、その階層より低い階層がなければ意味を持たないので、その性質を考慮する必要がある。

またキャッシュの更新をする際に、アクセスの頻度で各階層のキャッシュを更新する場合は、アクセスの多いデータほどより高品質なデータが提供される確率が高くなる。

3.3.2 キャッシュにおける階層の割り当て

キャッシュの階層の割り当て法として以下の3つの方法が考えられる。

- (1) 階層を考慮しないでキャッシュする。
- (2) 全てのデータに対して平等にキャッシュする。

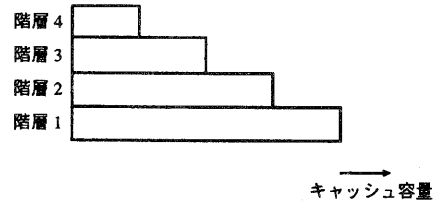


図5 階層型データのためのキャッシュシステム

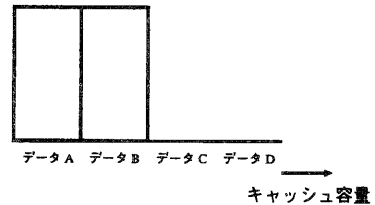


図6 階層の数を一つに設定

- (3) 下の階層ほどキャッシュを多くする。

これらの方法の得失を以下のデータ及びキャッシュ容量を用いて説明する。

- 4つのデータA, B, C, Dがある。
- 全てのデータは同じ大きさを持つ。
- 各階層は同じ大きさである。
- 各データは4つの階層からなる。
- キャッシュは2つのデータを完全に保存できる容量を持つ。

(1) の場合、一般のキャッシュと同じ性質を持ち、完全なデータを提供するか、全く提供しないか、のどちらかになる。図6の場合、データA, Bは最高品質のものを提供できるが、データC, Dは提供できない。この方法の利点として一つのデータに関しては最高品質で蓄えることができるが、欠点として蓄えることのできるデータの数が少ないことがある。

(2) の場合、全てのデータを収容できる階層までキャッシュするものである。図7では階層1, 2のみのキャッシュ容量を割り当てている。この方法の利点としてキャッシュに蓄えることのできるデータの数が多くなるが、欠点として全てのデータに関してデータの低い階層のみを蓄えることとなり、キャッシュでは低品質のデータしか提供できない。

(3) の場合、データの低い階層がより多くキャッシュに残るようになる。より多くアクセスされるデータは、その頻度に応じてキャッシュに残り、より高品質な部分までキャッシュに残る。そして最も低い階層が全て収まるようにキャッシュの容量を設定することで、データである中心部分を常に提供できる。

図8の例ではデータAは階層4までの完全なデータ

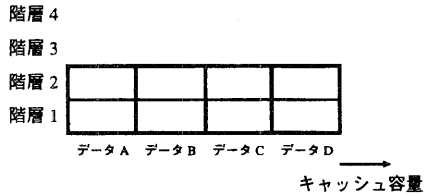


図7 階層1, 2以外容量0に設定

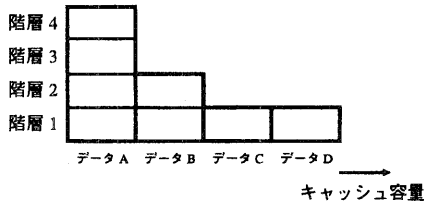


図8 段階的な設定

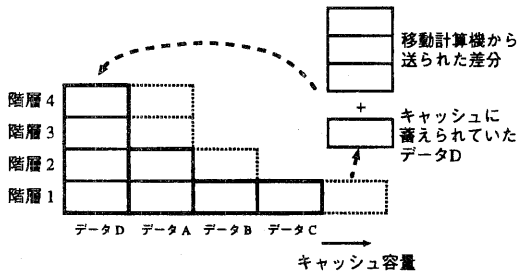


図9 データDがアクセスされた場合

がキャッシュに保持されている。データBは階層2の部分で階層キャッシュに存在し、データC, Dでは最も優先度の高い階層1のみが存在する。

図9は(3)の方式においてデータDが参照された時のキャッシュの更新の例を示している。点線部分は更新された時に各階層でのキャッシュの容量を越え、削除された部分である。この例では、左に位置するデータほど、より新しい参照があったことを示している。

3.3.3 各階層へのキャッシュ割り当て方式

第3.3.2節では各階層のキャッシュの容量を固定していた。この割り当て方を静的キャッシュ割り当てと呼ぶ。この場合、キャッシュできるデータの数に対して柔軟性がないが、上位の階層にも一定の容量が割り当てられるため、アクセス頻度に応じてキャッシュの更新を行う場合、アクセスの多いデータに関しては、より品質の高いデータを提供することが可能である。

これに対して、図10のように、各階層の容量を動的に変化させることも考えられる。キャッシュ容量のみを決めておき、データの数が少なく、キャッシュ容量が余っている場合、上位の階層にも多くのキャッシュを割

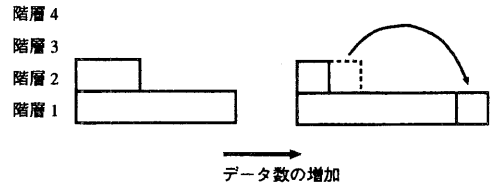


図10 動的キャッシュ割り当て

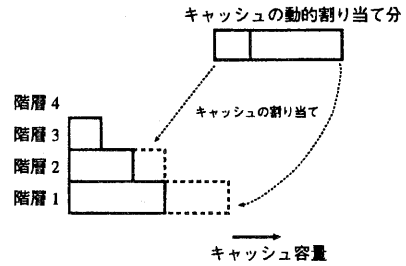


図11 静的・動的キャッシュ割り当て

り当てる。データの数が増えて、下層のキャッシュ容量が足りなくなると、蓄えることのできるデータの数を確保するために上位の階層に割り当てるキャッシュの容量を少なくし、その分下位の階層のキャッシュの容量を増加させる。この割り当て方を動的キャッシュ割り当てと呼ぶ。これにより、データ間や階層間のデータの大きさの違いを意識する必要がなくなる。しかし、全体のキャッシュ容量がデータの量に対して相対的に少なければ、上位の階層に容量が割り当てられる可能性が少なくなる。

静的キャッシュ割り当てと動的キャッシュ割り当てを組み合わせる方式も考えることができる。これはキャッシュに割り当てる容量のうち、ある容量を固定的に各階層に割り当て、残りを動的に割り当てるものである。図11では、各階層の実線部分があらかじめ割り当てられた固定的な容量である。これに動的なキャッシュの割り当て分を各階層に割り当てていく。

この方法では、データの数が少ない場合、高品質なデータのキャッシュにより多く容量を割り当てることが可能である。また、下層のキャッシュに容量が必要な時でも上層のキャッシュには固定的に割り当てた容量は必ず確保される。

3.4 WWWサーバ上でのキャッシュシステム

階層型データのためのキャッシュを、移動計算機からの情報発信のためのWWWサーバに設ける。接続状態で利用者がデータの全てを要求した場合には、図12のようなデータの流れになる。例としてここではデータAは4階層からなり3番目の階層までWWW

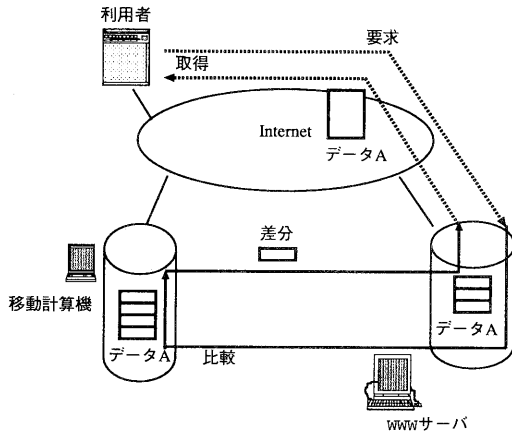


図 12 接続状態の完全なデータの転送

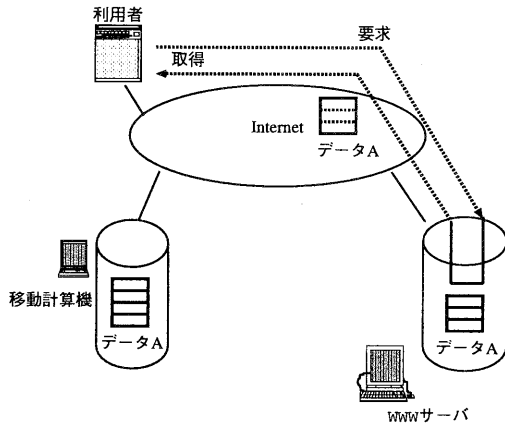


図 13 キャッシュに蓄えられたデータのみ転送

サーバのキャッシュに蓄えられていると仮定する。利用者からの要求により WWW サーバは移動計算機のデータと比較して、その差分である第 4 階層を移動計算機から受け取る。そして、キャッシュに保存されているデータと合成して、完全な形のデータとして利用者へ送信する。

分断状態や接続状態でも利用者がキャッシュに蓄えられているデータのみを要求した場合は、図 13 に示すように、データ A のキャッシュに蓄えられている第 3 階層までを一つのデータとして構成し、それを利用者に対して送信する。利用者では品質の落ちたデータとして内容を確認することが可能である。

このような構成にすることにより、利用者に対し分断状態では品質を落としたデータの一部を提供でき、接続状態ではユーザの選択により完全なデータあるいは品質を落としたデータを提供できる。また、完全な

データを求める場合でも、データの差分のみを送ることにより移動計算機から WWW サーバへの通信量を削減することができる。

4. まとめ

移動計算機の情報発信環境において、マルチメディアデータを発信する際の分断や通信帯域の狭さの問題に対処するために、階層構造を持ったキャッシュシステムを提案した。ここではプログレッシブ JPEG のような階層構造を持ったデータに注目し、その階層の優先度別に WWW サーバにキャッシュする仕組みを提案した。データの階層構造に合わせてキャッシュ自身も階層構造にし、データのなかで中心となる部分をより多くキャッシュに残るようにすることで、マルチメディアデータを発信する際に、移動計算機の通信が分断されたとしても、データの最低限必要な中心部分を利用者へ提供できる。また、移動計算機の通信帯域が狭い場合に、キャッシュに残っている品質の低いデータを利用することで、利用者が速度を優先するか、品質を優先するかを選択できる。

今後は本稿で提案した階層型データのためのキャッシュシステムを実際に発信環境のための WWW サーバ上に実装する予定である。その際、実際にマルチメディアデータを用意し、動作や効率を検証する。この中で、どのような階層型データの条件で各階層でのキャッシュ容量を静的・動的にどれくらい割り当てるべきかを検証していく。更に、このキャッシュシステムでの様々なデータの取り扱いや、受信側への応用等を考察していく。

参考文献

- 1) 辻 順一郎, 小津 浩直, 三浦 教史, 滝沢 直樹: モバイルプロキシサーバシステムの試作, 情報処理学会研究報告 Vol97, No72 pp.67-72, July 1997.
- 2) Armando Fox, Steven D.Gribble, Eric A.Brewer, Elan Amir, "Adapting to Network and Client Variability via On-Demand Dynamic Distillation," ASPLOS VII, pp.160-170, MA,USA, October 1996.
- 3) S. Tagashira, K. Nagatomo, K. Saisho, A. Fukuda, "An Information Announcement System based on WWW for Mobile Computers," IEICE Trans. Fundamentals (採録決定).
- 4) C. Perkins, "IP Mobility Support," RFC2002, October 1996.