

ウェブ共有環境におけるコンテンツ間の 相関性を考慮したキャッシングについて

中村 聡史[†] 塚本 昌彦[‡] 西尾 章治郎[‡]

[†] 大阪大学大学院工学研究科情報システム工学専攻

〒 565-0871 大阪府吹田市山田丘 2-1

E-mail: nakamura@ise.eng.osaka-u.ac.jp

[‡] 大阪大学大学院情報科学研究科マルチメディア工学専攻

〒 565-0871 大阪府吹田市山田丘 2-1

E-mail: {tuka,nishio}@ist.osaka-u.ac.jp

ウェブコンテンツ共有システムでは、ユーザが蓄積したコンテンツを P2P 型のネットワークで共有する。コンテンツを交換する際には、パケツリレー式にコンテンツを転送するため、コンテンツ取得までの待ち時間短縮が重要な課題となる。本システムで共有されるウェブコンテンツ間には相関性が存在するため、相関性を考慮したコンテンツキャッシングが有効であると考えられる。そこで本稿では、ウェブコンテンツ共有システムにおいて、コンテンツ内のインラインオブジェクトや、リンクされているコンテンツなどを効果的にキャッシングし、コンテンツ要求の待ち時間を短縮する方式を提案、実装する。実現する方式では、コンテンツを保持しているピアがコンテンツ要求を受けたときにコンテンツの内容を解析し、インラインオブジェクトやリンクページなどをコンテンツ要求ピアに向け送り出すことで、コンテンツ要求ピアにおける待ち時間を短縮する。

Caching Methods Based on Data Correlation in a Web Contents Sharing System

Satoshi NAKAMURA[†]

Masahiko TSUKAMOTO[‡]

Shojiro NISHIO[‡]

[†]Dept. of Information Systems Eng., Graduate School of Eng., Osaka University

[‡]Dept. of Multimedia Eng., Graduate School of Information Science and Technology, Osaka University

A web contents sharing system shares the web contents which users store in a P2P (peer to peer) network. In case of exchanging contents between two peers, it is important to shorten the waiting time for contents acquisition because the system uses the "bucket brigade" to transmit contents. Since there is several kinds of correlation between the web contents shared, it is effective to cache the contents based on such correlation. Then, in this paper, we realize the caching system which caches the inline objects in contents and the linked contents in order to shorten the waiting time for contents acquiring. In this system, when the peer which holds contents receives a contents request, the peer sends a requested content and analyzes it to get information of inline objects and the linked contents of the content. In addition, the peer sends the inline objects and the linked contents to the request peer.

1 はじめに

近年、コンピュータの低価格化・高性能化、ネットワークインフラの整備などにより、WWW (World Wide Web) 上で公開されているウェブコンテンツの量が飛躍的な速度で増加しつつある。その数は、JP ドメインに所属するサーバで公開されているコ

ンテンツだけでも約 2 億 7421 万ファイルにのぼり、コンテンツの総サイズは 10 テラバイトに達すると推計されている [4]。ユーザは検索サービスなどを利用することにより、これら膨大なコンテンツの中から必要な情報を効率よく探し出すことができる。しかし、ユーザがオフライン状態にあるときや、サーバやネットワークが不調な時はコンテンツに辿り着

くことができない。また、コンテンツは随時更新されていくものであり、ユーザが必要とするときにはコンテンツが別のものになっている可能性もある。さらに、コンテンツの数があまりに多すぎるため、新しいコンテンツに対する検索などは困難である。

こうした問題を解決するため、本研究グループではウェブコンテンツ共有システムを提案・実装している [1]。ウェブコンテンツ共有システムでは、ユーザがあらかじめ WWW 上からローカルディスクに保存しておいたウェブコンテンツを、Gnutella[3] のような PureP2P 型のネットワークで共有する。ウェブコンテンツを保存する際に、取得元の URL や、ウェブページのタイトル、ファイルサイズ、更新日時、取得日時などのメタ情報もまとめて保存しておくことで、メタ情報を利用した柔軟な検索を可能とする。共有されるコンテンツの種類は多岐にわたり、すでに公開されていない過去のコンテンツなども共有できるため、ユーザにとって有用なシステムであるといえる。また、ユーザがコンテンツを保存したときにそのコンテンツが検索対象となるため、新しいコンテンツに対する検索も可能となる。さらに、同じ興味をもつユーザ同士でコンテンツ共有ネットワークを構成することにより、効果的にコンテンツ検索および共有ができる。このウェブコンテンツ共有システムから構成される P2P ネットワークをウェブコンテンツ共有ネットワークと呼んでいる。ウェブコンテンツ共有ネットワークではパケツリレー式にコンテンツを転送する [2] ため、コンテンツ取得までの待ち時間の短縮が重要な課題となる。

ウェブコンテンツ共有ネットワークでは、中継ピアにおいてコンテンツをキャッシュをすることによりコンテンツ要求の待ち時間を短縮する。しかし、これまでは中継したコンテンツを中継ピアで一定期間キャッシュしているだけであったため、有効なキャッシング方式であるとはいえなかった。コンテンツ要求に関する待ち時間を短縮するには、コンテンツ間の相関性を考慮したコンテンツのキャッシングが有効であると考えられる。

そこで本研究では、ウェブコンテンツ共有システムにおけるコンテンツ間の相関性を考慮したキャッシングに関する考察を行い、コンテンツ要求における待ち時間を短縮するキャッシングシステムを提案・実装する。また、実際にウェブコンテンツ共有システム上でキャッシングシステムを運用し、キャッシ

ングシステムの有効性について考察を行う。

以下、2 章ではウェブコンテンツ共有システムについて説明し、3 章ではキャッシング手法について議論する。4 章ではキャッシングシステムの実装について述べ、5 章では実現したキャッシングシステムを用いた実測実験を行う。6 章でキャッシングシステムに関する考察を行い、最後に 7 章にてまとめと今後の課題について述べる。

2 ウェブコンテンツ共有

ウェブコンテンツ共有システムは一般的な P2P ファイル共有システムとは異なり、共有されるファイルがネットワーク帯域に比べ非常に小さく、膨大であるという特徴がある。また、ウェブコンテンツであるため、複数のコンテンツが連続してアクセスされるという特徴がある。

ウェブコンテンツ共有ネットワークで共有されるウェブコンテンツは、それ自身である HTML ページの他に、CSS などのスタイルシート、JPEG や PNG などの画像などのインラインオブジェクトから構成されている。このインラインオブジェクトはウェブコンテンツの閲覧において欠かせないものである。ウェブコンテンツ共有ネットワークにおいてコンテンツを取得する場合には、まずコンテンツの母体となる HTML ページを要求し、HTML ページを取得し、解析した後、順次インラインオブジェクトを要求し、取得するという手間が必要になる。そのため、インラインオブジェクトの量が多い場合には要求および応答の数も増加してしまい、結果としてコンテンツ取得における待ち時間が長くなってしまふ。また、コンテンツ要求ピアとコンテンツ保持ピアの間のホップ数が大きい場合には、何度も応答が行き来する必要があるため、コンテンツ中継ピアの負荷も増大してしまう。

図 1 は、ウェブコンテンツ共有ネットワークにおいて、コンテンツ取得要求ピア A から、ピア B、ピア C を中継して、コンテンツ保持ピア D に対してコンテンツ取得要求を行っている様子である。まず始めに母体となるコンテンツを要求し、コンテンツを取得した後、コンテンツを解析し、インラインオブジェクトを要求している。インラインオブジェクトは一気に要求され、コンテンツ保持ピアから順次

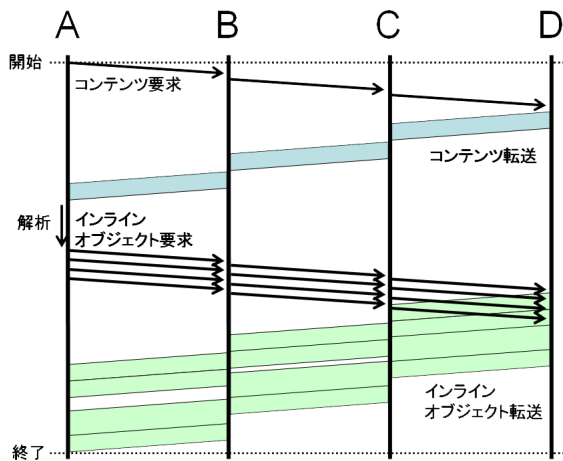


図 1: コンテンツ取得要求

コンテンツが転送されている。

このコンテンツのやり取りにおいて、コンテンツ間の相関性を考慮してコンテンツを効果的にキャッシングするとコンテンツ取得までの待ち時間を短縮できると考えられる。

3 相関性を考慮したキャッシング

3.1 コンテンツ間の相関性

ユーザがブラウザを利用して WWW 上にあるウェブコンテンツを閲覧する場合、ウェブコンテンツと、そのウェブコンテンツ内に含まれる画像やスタイルシートなどのインラインオブジェクトは必ず連続してアクセスされる。また、そのウェブコンテンツからリンクとして接続されている他のウェブコンテンツは、連続してアクセスされる可能性が高い。さらに、ディレクトリサービスなどで同一のカテゴリに分類されているようなウェブコンテンツや、検索エンジンを利用して同一キーワードで取得されるウェブコンテンツも連続してアクセスされる可能性が高い。

以上のように、あるコンテンツ閲覧システムにおいて複数のコンテンツが存在し、それらのコンテンツが連続してアクセスされる場合、コンテンツ間に相関性があるという。この相関の度合いを相関度と呼び、アクセスされる頻度が高ければ高いほど相関度は高いものとする。

コンテンツ共有システムにおいて、コンテンツ間の相関性は大きく分けて以下のように分類される。

従属関係： コンテンツ A がコンテンツ B をコンテンツの一部として含む場合、コンテンツ B はコンテンツ A に従属しているという。主コンテンツがアクセスされた際、従属しているコンテンツは必ず連続してアクセスされる。

接続関係： コンテンツ A からコンテンツ B への接続が存在する場合、コンテンツ B はコンテンツ A に対して接続関係にあるといえる。接続関係にあるコンテンツは連続してアクセスされる可能性が高い。

類似関係： コンテンツ A とコンテンツ B がある条件下において同一カテゴリに分類される場合、コンテンツ A とコンテンツ B は類似関係にあるといえる。類似関係にあるコンテンツは連続してアクセスされる可能性が高い。

無関係： コンテンツ A とコンテンツ B の間に関係が存在しない場合、コンテンツ A とコンテンツ B は無関係であるといえる。無関係にあるコンテンツが連続してアクセスされる可能性は低い。

このコンテンツ間の相関性をウェブコンテンツ共有に適用すると、以下ようになる。

従属関係： コンテンツと、コンテンツ内部に含まれるインラインオブジェクトとの関係。

接続関係： コンテンツと、コンテンツからリンクされているコンテンツとの関係。

類似関係： コンテンツと、同一キーワードで検索されるコンテンツとの関係。

無関係： コンテンツと、まったく関係のないコンテンツとの関係。

上記の相関性は、コンテンツの種類や問合せの内容により強弱が異なる。例えば、リンクメニューしかないようなコンテンツの場合、そのコンテンツから接続関係にあるコンテンツへとアクセスされる可能性は高いが、テキストがメインでリンクは付加的なものである場合、そのコンテンツから接続関係にあるコンテンツへとアクセスされる可能性は低い。一方、類似関係については、同一の検索キーワードによりリストアップされても、類似度が高ければ連

続いてアクセスされる可能性は高いが、類似度が低い場合、連続してアクセスされる可能性は低い。しかし、ウェブコンテンツである以上、従属関係にあるコンテンツは相関度が最も高く、無関係のものは相関度が最も低い。そこで、本稿では簡単のため、コンテンツ間の相関度を従属関係、接続関係および類似関係、無関係の3段階に分け考える。

これらの相関性を考慮したコンテンツのキャッシングがコンテンツ取得における待ち時間短縮において有効であると考えられる。

3.2 能動的キャッシング

ウェブコンテンツ共有ネットワークにおいては、コンテンツ要求ピアからコンテンツ保持ピアまでのホップ数がコンテンツ取得の待ち時間に大きく影響する。そのため、相関度の高いコンテンツをコンテンツ要求ピアの近くのピアでキャッシュしておくことにより、コンテンツを取得するまでの待ち時間を大幅に短縮することができる。例えば、コンテンツ A の取得要求が発生した場合に、コンテンツ A と従属関係にあるコンテンツ B や、直接接続関係にあるコンテンツ C を、コンテンツ要求ピアの近くにキャッシングしておくことで、コンテンツ取得にかかる待ち時間を短縮できると考えられる。そこで、本研究ではコンテンツ取得要求発火時にコンテンツ保持ピアが能動的にコンテンツを転送し、中継ピアでのキャッシングを行う、能動的キャッシング手法を実現する。

能動的キャッシング手法では、コンテンツ保持ピアがコンテンツ取得要求を受信したときに、要求されたコンテンツをコンテンツ要求ピアに向け送信すると同時に、コンテンツの内容を解析する。解析により従属関係にあるインラインオブジェクトや接続関係にあるリンク先のコンテンツをリストアップし、相関度にあわせてコンテンツ要求ピアに向けコンテンツを転送する。このやり取りにより、連続してコンテンツにアクセスすると予想されるコンテンツ要求ピアの近くで、対応するコンテンツをキャッシュすることが可能となり、コンテンツ取得にかかる待ち時間を短縮できると考えられる。

図 2 上は、コンテンツ取得の際に能動的キャッシング手法を利用している様子である。コンテンツ

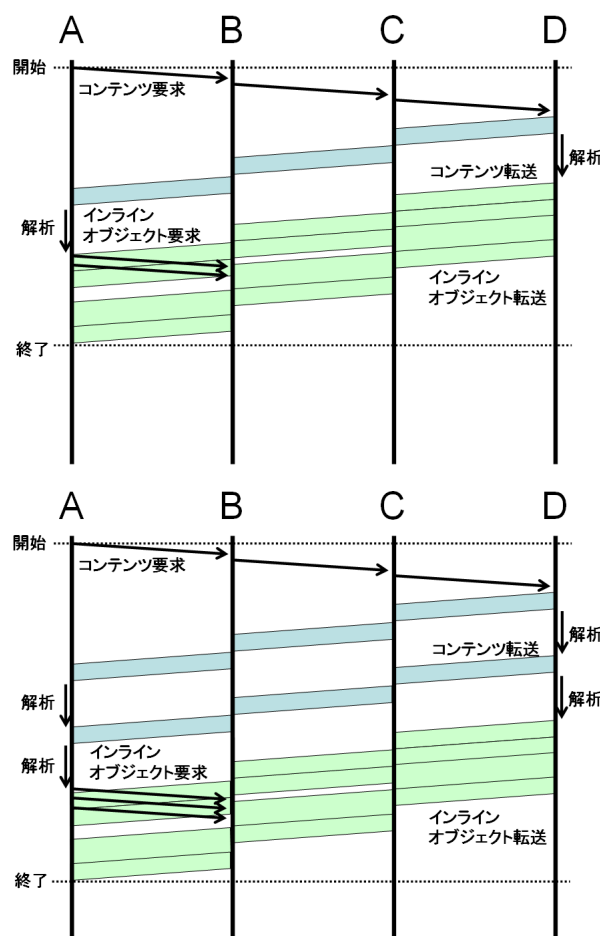


図 2: 能動的キャッシングを利用したコンテンツ取得

取得要求ピア A から、ピア B, ピア C を中継して、コンテンツ保持ピア D に対してコンテンツ取得要求を行っている。まず始めに母体となるコンテンツを要求し、コンテンツを取得した後、コンテンツを解析し、インラインオブジェクトを要求している。このときに、ピア D 側でもコンテンツの解析を行い、関連するインラインオブジェクトを転送しているため、ピア B-ピア C 間、ピア C-ピア D 間の無駄なやり取りを省くことに成功しており、結果としてコンテンツ取得までの待ち時間が短縮されている。図 2 下は、同様にコンテンツ能動的キャッシング手法を利用してコンテンツを取得している様子である。ここでは、フレームページなど HTML ページの中に HTML が含まれるようなコンテンツを要求し、取得している。本来は母体となるコンテンツ、フレームページ、インラインオブジェクトと 3 回の

コンテンツ取得要求を行う必要があるが、コンテンツ保持ピアで全ての処理を行うことにより、応答を省くことができている。

3.3 分散キャッシング

P2P ネットワークで共有されるすべてのウェブコンテンツを各ピアが保持していれば、コンテンツ取得における待ち時間を大幅に短縮することができるが、保存できるディスクには制限があるうえ、ウェブコンテンツの量は無数であるため現実的ではない。また、コンテンツをネットワーク上に理想的に配置できればコンテンツ取得における待ち時間を短縮できると考えられるが、P2P 型のネットワークを利用しているため、トップダウンでキャッシュの再配置を行うことはできない。

ウェブコンテンツ共有システムにおけるローカルディスクへのコンテンツ保存スタイルとしては、ユーザがデータとして明示的にローカルディスクに保存するタイプと、中継時にキャッシュとして一時的に保存するタイプがある。ユーザが明示的にコンテンツをローカルディスクに保存している場合は、ユーザが削除処理を行わない限りコンテンツが消えてしまうことはないが、キャッシュに保存されているコンテンツは、キャッシュの状態に応じてシステムに自動的に削除されてしまう。この保存スタイルと、コンテンツの相関性を考慮することにより、キャッシュの分散配置を行う。

実現する分散キャッシング手法では、キャッシュサイズが限界に近づくと、キャッシュ中のコンテンツに優先度をつけ、優先度の低いコンテンツから先にキャッシュから削除する。キャッシュに格納されたばかりのコンテンツやアクセスされたばかりのコンテンツは優先度を高く、キャッシュに長い間格納されておりアクセスされていないコンテンツは優先度を低く設定する。また、隣接ピアがコンテンツをデータとして保存している場合には、そのコンテンツが削除される心配がないため、キャッシュ中のコンテンツの優先度を低く設定し、隣接ピアがキャッシュのみにコンテンツを保持している場合は、コンテンツ削除の可能性を考慮し、優先度を中程度に設定する。隣接するどのピアも同じコンテンツを保持していない場合は、そのコンテンツに対する優先度

を上げ、なるべくキャッシュから削除されないようにする。一方、相関度の高いコンテンツは同時にアクセスされる可能性が高く、相関度の高いコンテンツは一つのピアで保持していた方がコンテンツ問合せの効率が良くなるため、従属関係にあるコンテンツはあわせて保存しておくようにする。

以上の仕組みにより、隣接するピア同士で極力コンテンツを保持することが可能となるため、隣接ピアごとにキャッシュが最適化されると考えられる。この分散キャッシングにおいては、各優先度の重み付けの設定が性能に大きくかかわってくると考えられるため、十分に考察する必要がある。

4 実装

4.1 能動的キャッシング

能動的キャッシング手法では、コンテンツ要求に対応しているコンテンツを保持しているピアがコンテンツ要求を受信すると、以下のような処理を行う。

1. 要求されたコンテンツをコンテンツ要求ピアへ向け転送する。このとき、ピアにおける計算負荷が大きいときは終了する。計算余裕があるときには解析中のフラグを追加し、手順 2 へすすむ。
2. コンテンツを解析し、従属関係にあるインラインオブジェクトや接続関係にあるリンク先コンテンツを相関度順に転送候補リストに格納。
3. コンテンツ検索時に利用された Query-ID をもとの、同一 Query で発見されたコンテンツを相関度順に転送候補リストに格納。
4. 転送候補リストに同一のコンテンツが格納されている場合は、相関度の低いアイテムをリストから削除。
5. 転送候補リストから相関度順にコンテンツ保持の有無を確認し、コンテンツを保持している場合は、転送範囲を指定してコンテンツを転送。

コンテンツ間の相関度が高ければ高いほど、そのコンテンツをコンテンツ要求ピアの近くに配置する必要があるため、コンテンツの転送範囲はコンテンツの相関度およびコンテンツ要求ピアからコンテ

コンテンツ保持ピアまでのホップ数に応じて設定する。この、コンテンツの転送範囲 (*TTL*) は、コンテンツの相関度およびコンテンツ要求ピアからコンテンツ保持ピアまでのホップ数 (*Hops*) に応じて以下のように決定する。

$$TTL = Hops \times \text{相関度}$$

相関度は、0 から 1 までで定義される値で、相関度が高ければ 1 に近づき、相関度が低ければ 0 に近づく。本稿では簡単のため、従属関係にあるコンテンツの相関度を 1、接続関係、類似関係にあるコンテンツの相関度を 0.25、無関係のコンテンツの相関度を 0 としている。*TTL* は小数点以下を切り捨てる。この *TTL* を転送するコンテンツに付加することにより、転送範囲の制限を行う。

コンテンツを受信したピアは、まず受信したコンテンツをキャッシュ内に格納する。また、転送されてきたコンテンツを *TTL* が 0 より大きい場合、*TTL* を 1 つ減らした後、受信したコンテンツをコンテンツ要求ピアへ向けて転送する。*TTL* が 0 のときはコンテンツの転送を終了する。なお、コンテンツを中継するピアが転送されてくるコンテンツを保持している場合、ヘッダだけを受信し、コンテンツ自体は受信しない。受信したコンテンツが未解析で、計算余裕がある場合はここで解析を行い、相関度の高いコンテンツの要求を行う。

中継ピアが従属関係や接続関係にあるコンテンツに対するコンテンツ取得要求を受信した場合には、現在受信中のコンテンツ、現在ローカルディスク (キャッシュを含む) に保存されているコンテンツを順にチェックし、受信中であればコンテンツが保存されるのを待ち、キャッシュ内のコンテンツをコンテンツ取得要求ピアへと転送する。もし、コンテンツを保持していない場合は、コンテンツ保持ピアへ向け、コンテンツ取得要求を転送する。

以上のやり取りにより、コンテンツ取得までの待ち時間を短縮する。

4.2 プロトコル

ウェブコンテンツ共有ネットワークにおける相関性を考慮した能動的キャッシング手法を実現するため、プロトコルを拡張する。

まず、コンテンツ要求ピアからコンテンツ保持ピアまでのホップ数が必要となるため、コンテンツ取得問合せを以下のように拡張する。

```
GET <取得するコンテンツの URL>
Queryhit-ID: <Queryhit で使用した ID>
Hops: <コンテンツ要求ピアからのホップ数>
```

Hops はコンテンツ要求ピアからのホップ数であり、問合せ転送時に 1 つずつ加算される。この Hops は、コンテンツの転送範囲を計算する際に利用する。

GET メッセージに対応するコンテンツを保持しているピアは、対応するコンテンツをコンテンツ要求ピアへと転送する。コンテンツは GET メッセージをもとに、コンテンツ要求ピアに順次転送される。このコンテンツ取得要求に対するコンテンツ転送を、明示的に PUT という形で行うため、コンテンツ転送のプロトコルを以下のように変更する。

```
PUT <転送するコンテンツの URL>
Queryhit-ID: <Queryhit で使用した ID>
Date: <現在の日時>
Last-Modified: <コンテンツの最終更新日>
Content-Type: <コンテンツのタイプ>
Accept-Ranges: <レンジタイプを指定>
Content-Length: <コンテンツのサイズ>
TTL: <コンテンツの転送回数制限>
Hops: <コンテンツ保持ピアからのホップ数>
Status: <処理ステータス>
Received: from <中継ピアの IP アドレス>
:
```

TTL では転送するコンテンツを何ホップ先まで転送することを許可するかを制限する。*TTL* はコンテンツが他のピアへ転送される際に 1 つずつ減らし、*TTL* が 0 になると転送を中止する。また、*Status* にてコンテンツを解析しているかどうかを指定する。もし、コンテンツを解析している場合は *Analyzed* を、解析していない場合は *None* を指定する。コンテンツはこの PUT ヘッダの後で送信する。

中継ピアに計算の余裕があり、コンテンツの処理状態が *None* になっている場合は、PUT ヘッダの *Status* を *Analyzed* にしてコンテンツを次のピアへ転送する。その後、コンテンツを解析し、コンテンツ保持ピアに対してコンテンツ転送要求を行う。

```
GET <取得するコンテンツの URL>
Queryhit-ID: <Queryhit で使用した ID>
Hops: <コンテンツ要求ピアからのホップ数>
Correlation: <コンテンツの相関度>
Referrer: <相関するコンテンツの URL>
```

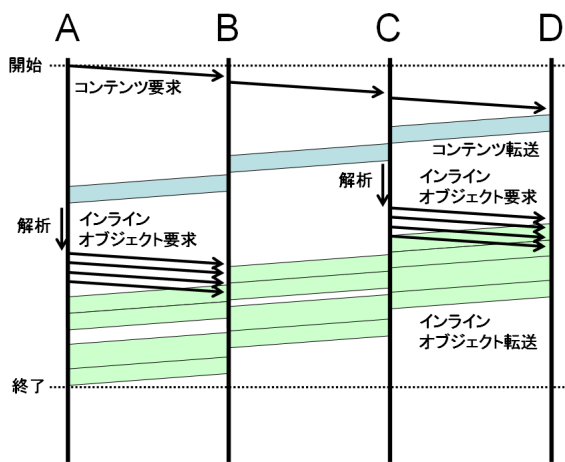


図 3: 中継ピアでの解析例

ここで Correlation は解析結果に基づくコンテンツの相関度を、Referrer は相関関係にあるコンテンツの URL を指定する。このやり取りにより、中継ピアにおけるコンテンツの能動的キャッシングも可能となる。図 3 は、中継ピアにおけるコンテンツ能動キャッシングの例である。

5 簡易実測実験

能動的キャッシングシステムを利用したプロトタイプシステムの有用性を示すため、実機を用いた簡易実測実験を実施した。

実験では、中継ピアの数によりどの程度コンテンツ取得にかかる時間が変化するかを調べるため、コンテンツ要求ピアと、コンテンツ保持ピアをそれぞれ両端として、直線状にピアを相互接続する。中継ピアの台数は、実験に応じて変化させる。また、2,000 個のウェブページ（ファイル数は約 17,000）のコンテンツを用意し、各コンテンツについて、コンテンツ要求ピアがコンテンツ（ウェブページ）取得要求を送信し、コンテンツを取得完了するまでの時間を計測する。2,000 個すべてのウェブページについて同様の応答を行い、応答時間を記録する。この実験を、従来システムと、能動的キャッシングを用いた今回のシステムで比較する。なお、今回の実験では分散キャッシングは用いず、中継した全てのコンテンツを各ピアがキャッシュできるようにしている。また、コンテンツ検索は行わないため、

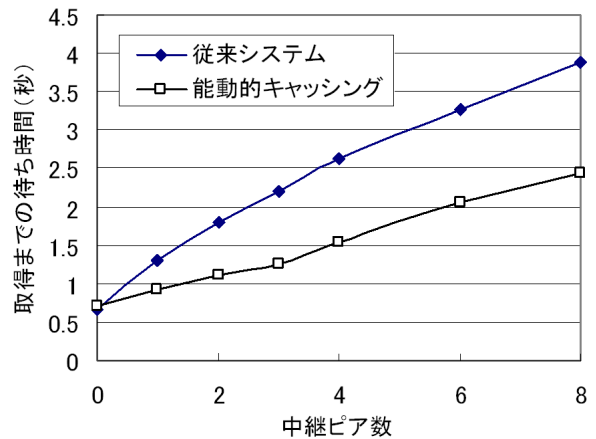


図 4: 実験結果

類似に関する相関性を利用したコンテンツのキャッシングは利用しない。さらに、実験の都合上、6 台目以降の計算機を別のポートを利用して再度端のコンピュータから接続するようにしている。そのため、中継ピアが 4 台以上の実験結果は参考データである。

実験結果は図 4 の通りである。図の横軸は、コンテンツ要求ピアと、コンテンツ保持ピアの間で中継するピアの数を、縦軸はコンテンツ取得完了までにかかった時間の平均を表している。図より明らかに、中継ピア数が 1 以上の時、コンテンツ取得までの待ち時間を短縮できていることがわかる。一方、中継ピアが存在しないときは、従来手法に比べ、悪い結果が出ていることがわかる。これは、コンテンツ保持ピアおよびコンテンツ要求ピアの間で無駄な問合せが発生してしまったことや、コンテンツ保持ピアでコンテンツ解析中にコンテンツ要求ピアからコンテンツ要求が発生してしまい、待ち時間が発生してしまったことが原因と考えられる。

今回実施した実験は、あくまで簡易的な実験であったため、中継ピア数によるコンテンツ取得までの時間しか計測を行わなかった。ウェブコンテンツ共有ネットワークでは、コンテンツ検索の問合せに加え、他のピアもコンテンツ取得要求などを行うため、必ずしも今回のようにスムーズにコンテンツが取得できるわけではなく、キューでの待ち時間も発生すると考えられる。今後は、より現実的な実測実験を実施するための仕組みを実装する予定である。

6 考察

実測実験の結果より、ウェブコンテンツ共有ネットワークにおけるコンテンツ取得の待ち時間短縮に、コンテンツの能動的キャッシングが有効であることを示した。結果からも明らかのように、中継ピア数が増加すれば増加するほど、従来システムに比べ、取得までの待ち時間の平均が短縮できていることがわかる。しかし、今回実現した能動的コンテンツキャッシング手法では、コンテンツ保持ピアが要求されたコンテンツを解析し、順次インラインオブジェクトを転送していくため、コンテンツ保持ピアにおける処理負荷が増大してしまうという問題がある。コンテンツ解析の負荷については、今回の実験において解析に要している時間を計測したところ、10KB 程度のコンテンツで 0.03~0.04 秒程度、100KB 程度のコンテンツで 0.1 秒程度、100KB を越えるコンテンツでも 0.2 秒程度となっている。このコンテンツ解析に要している時間は、コンテンツ検索における検索処理に要する時間に比べ負荷は小さいといえるが、あまりにコンテンツ要求が頻発すると、大きな問題となる可能性もある。なおこの問題については、コンテンツの解析を中継ピアで行うことにより解決できると考えられるため、作業に優先度を与え、処理に余裕のあるピアで効果的に解析を行う仕組みを実現する必要がある。

今回の実験では、他のピアではコンテンツ取得要求を発生していなかったうえ、各コンテンツ取得要求にはインターバルがあったため、それぞれのコンテンツ取得の影響はなかったが、実環境ではコンテンツ取得が頻発する可能性がある。このとき、従属関係にあるコンテンツは、必ず同時にアクセスされるため問題はないが、接続関係にあるコンテンツや、類似関係にあるコンテンツを転送し、そのコンテンツが利用されなかった場合、無駄にトラフィックを増やしてしまうことに繋がる。この問題については、コンテンツの相関度をより詳細にすることで対応できると考えられる。例えば、同じコンテンツから接続関係にあるコンテンツでも、同サイトにあるコンテンツに比べ、他サイトにあるコンテンツへのアクセスは一般的に低いといえる。ウェブアクセスの特性を考慮し、コンテンツの相関度をコンテンツ内のリンクの数や、リンク先の情報、リンクの表示位置により計算することにより、無駄なコンテン

ツの転送を低減できると考えられる。コンテンツ相関度の計算については今後の課題とする。

分散キャッシングでは、隣接ピア同士のどちらかがコンテンツを保持するような戦略をとるため、ネットワーク全体で見たときに、コンテンツを効果的に分散配置できるようになると考えられる。しかし、現時点では実装完了していないため、その有効性を十分に評価することができない。また、分散キャッシングにおける隣接ピアとのやり取りや、コンテンツの再配置がトラフィックの増加を招く可能性もあるため、今後検証していく必要がある。

7 まとめ

本稿では、ウェブコンテンツ共有ネットワークにおけるコンテンツ取得の待ち時間を短縮するため、コンテンツ間の相関性を考慮した能動的コンテンツキャッシング手法を提案し、キャッシング手法を用いたシステムを実現した。また、本手法を利用した実測実験により、実際にウェブコンテンツ共有ネットワークにおけるコンテンツ取得までの時間を短縮できていることを明らかにした。

今後は、実測実験の結果を元にモデルを作成し、大規模な P2P 環境におけるシミュレーションを実施する予定である。また、トラフィックの問題や、ピア単位での負荷についても調査する予定である。

謝辞

本研究は、文部科学省 21 世紀 COE プログラム（研究拠点形成費補助）、文部科学省科学技術振興調整費「モバイル環境向 P2P 型情報共有基盤の確立」の研究助成によるものである。ここに記して謝意を表す。

参考文献

- [1] 中村聡史, 塚本昌彦, 西尾章治郎, “コンテンツ流通制御を考慮したウェブコンテンツ共有システムの実現,” 情報処理学会論文誌, Vol. 45, No. 1, 2004 掲載予定.
- [2] 中村聡史, 塚本昌彦, 西尾章治郎, “ウェブ共有システムにおけるコンテンツ流通制御のための監視機構,” 情報処理学会シンポジウムシリーズ マルチメディア, 分散, 協調とモバイルシンポジウム (DICOMO2003) 論文集, Vol. 2003, No. 9, pp. 725-728, 2003.
- [3] “Gnutella,” <http://www.gnutella.com/>.
- [4] 総務省, “平成 15 年版情報通信白書,” 2003.