

## P2P モデルにおけるキャッシュ機構の有効性の評価および提案

2H-4

渡里 雅史<sup>†</sup> 植原 啓介<sup>‡</sup> 村井 純<sup>†</sup>慶應義塾大学 環境情報学部<sup>†</sup> 慶應義塾大学大学院 政策・メディア研究科<sup>‡</sup>

## 概要

本論文では、Peer to Peer(P2P)モデル通信における制御トラフィックの軽減について考察する。既に、文献[2]においてキャッシュ機構がトラフィック軽減に有効であることが確認されている。しかし、文献[2]はキャッシュの不整合を考慮しておらず、十分な考察とは言いがたい。本論文においては実トラフィックにおけるキャッシュの不整合を調査し、より最適なキャッシュ機構を提案する。

## 1 はじめに

P2P モデルは従来のサーバ・クライアントモデルとは異なり、不特定多数のクライアントが相互につながりネットワークを形成する。個々のクライアントは形成された P2P ネットワークを通じて、様々な情報を共有する。

本論文では、この P2P モデルを対象とする。それに適したキャッシュの方法を考察するために、P2P モデルである Gnutella[1]を用いた。

## 1.1 Gnutella

P2P では情報を集約するサーバが存在しない。そのため、情報の検索を行う際に、クライアント自身が P2P ネットワーク上に存在する他のクライアントに対して検索を行う必要がある。

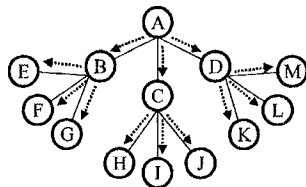


図 1: Gnutella モデルにおける検索方式

Gnutella でのファイル検索の仕組みを、図1を用いて説明する。ユーザ A は B, C, D と接続されている。B は E, F, G と、C は H, I, J と、D は K, L, M と接続されているとする。A が検索者、また M が目的のファイルを所有しているとする。まず、ユーザ A は B, C, D にリクエストを送信する。B, C, D は接続されている他のユーザにリクエストを伝達する。そして、M → D → A という経路を通して、レスポンスが伝達される。

The evaluation and proposition for the effective cache mechanism on p2p architecture  
Masafumi WATARI<sup>†</sup> Keisuke UEHARA<sup>‡</sup> Jun MURAI<sup>†</sup>  
Faculty of Environmental Information, Keio University<sup>†</sup>  
Graduate School of Media and Governance, Keio University<sup>‡</sup>  
Keio University Shonan Fujisawa Campus  
5322, Endo, Fujisawa, Kanagawa 252, Japan  
E-Mail: watari@sfc.wide.ad.jp

## 1.2 Gnutella におけるキャッシュ機構

1.1 節で述べた Gnutella における検索では、検索結果を受信したにも関わらず、検索要求が拡散する可能性がある。検索要求が拡散するにつれて加速度的に大量のトラフィックが発生するため、Gnutella ではの検索の効率化が重要となる。

文献[2]における研究では、検索にキャッシュ機構を用いることにより、検索を効率化し、トラフィックを軽減している。この機構では、他のクライアントが保持している情報をキャッシュすることにより、検索要求の拡散を防ぐことが可能である。図1では、たとえば D がキャッシュ持つことによって、K, L 無駄なトラフィックが流れない。

## 2 本研究における問題意識

本論文ではキャッシュの整合性に関して考察を行い、キャッシュ機構の有効性を確認し、さらに P2P モデルに特化したキャッシュ機構を示す。この機構は以下を考慮する。

- キャッシュの整合性  
キャッシュ機構はトラフィック軽減のために有効であるが、無駄なキャッシュを持つ事により、実在する情報との整合性に問題が生じる。
- 短期間存在型クライアント  
P2P においてはダイヤルアップ接続等のような短期間存在型クライアントの存在も考えられるため、これらのクライアントが持つファイルに対するキャッシュの整合性の維持も重要となる。

本論文では、これらを検証するために以下の項目を評価する。

- キャッシュ機構の有効性  
キャッシュが不整合となった理由を調べることにより、P2P モデルにおけるキャッシュ機構の有効性を求める。
- キャッシュの有効期間  
キャッシュの有効期間を調べる事により、P2P におけるキャッシュの最適な有効期間を求める。

## 3 実験手法

実験環境としては、Gnutella を用いてネットワークを流れるトラフィックを観測した。文献[2]と同様の手法を用いて、Gnutella ネットワークにクライアントを設置しパケットの観測を行った。また、設置したクライアントを経由して流れるトラフィックから、リクエストとリプライが一致したものをキャッシュとして保持し、一定の期間継続してパケットを監視を続けた。作成したキャッシュの整合性の確認は、キャッシュされているファイルを持つクライアントに直接リクエストを行いその結果を用いた。

## 4 結果

## 4.1 キャッシュした情報の整合性

図2は、ファイルと所有するクライアントの情報をキャッシュした場合、キャッシュの整合性が継続して確認できた時間を、横軸に時間をとりとし、その時間で整合性が確認できたファイル数を縦軸で示す。また、キャッシュの作成時は、ファイルの存続を 100% と仮定する。

ファイルの検索を行った結果、5分後にそのファイルを得られた割合は、全体の 15% に過ぎなかった。また、10

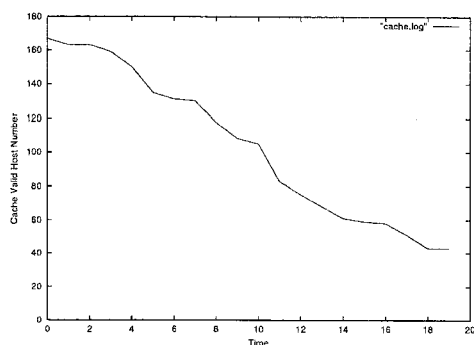


図 2: ファイルと所有するクライアントの整合性

分を過ぎても継続して情報が得られた割合は、全体のわずか 7% だった。

#### 4.2 キャッシュの不整合と分析

ファイルを所有するクライアントにリクエストを行った結果、そのファイルを確認できなかった結果を以下にまとめた。表 1 は、キャッシュの不整合が生じた理由と、その総数と割合を示す。

表 1: キャッシュの整合性

不整合の理由	ファイル数 3796	割合 (%)
File Not Found	175	4%
Host Unreach	940	25%
Host Refused Connection	2556	67%
File Found	125	4%

- File Not Found  
理由としては、ファイルの移動、削除、またはファイルの共有を許可していないことなどが考えられる。
- Host Unreach  
理由としては、クライアントがインターネットへの接続が切れたからだと考えられる。
- Host Refused Connection  
理由としては、Gnutella クライアントが落ちているか、クライアントがファイアウォールの中にいることが考えられる。

## 5 考察

### 5.1 Gnutella におけるキャッシュ機構

P2P モデルのように、クライアント同士で形成されたネットワークにおいて、キャッシュ機構を用いると、キャッシュの整合性とののが困難になる。

これは P2P モデルの特徴の一つである、クライアントがネットワークに短期間の間しか存在しないことなどが挙げられる。例として、ダイアルアップクライアントのような常時接続でないユーザが挙げられる。

4 節の結果から明らかであるように、キャッシュとの不整合の多くは、クライアントがネットワークから離れることが原因である。クライアントがネットワークに存在するにも関わらず、ファイルが得られない割合は少ない。

以上から Gnutella のネットワークにおけるキャッシュは、整合性の問題から長期間に渡って有効ではないことを実証した。キャッシュする情報を 5 分以上所有した場合の整合性は、わずか 1 割だった。

### 5.2 P2P モデルに適したキャッシュ機構の提案

P2P モデルでは、トラフィックを軽減するために、キャッシュ機構が有効であるが本研究の研究からキャッシュの整合性にまだ問題があることが実証された。

P2P モデルでは、整合性を保つために、有効期間を加えるなどした最適なキャッシュを行う必要がある。

4.1 節から一度情報をキャッシュした場合は、5 分以上所有しないことで情報の整合性を図ることが望ましい。

4.2 節から、P2P モデルでは短期間存在型クライアントが多く存在するため、これを考慮しなければならない。以上のことを踏まえて、P2P モデルに適したキャッシュ方法を提案する。

- クライアントによる重み付け  
ダイアルアップクライアントのような、短期間しか存在しないクライアントを、区別する。これらのクライアントはネットワークから頻繁にはずれ、キャッシュに不整合が生じる可能性が高くなる。クライアントは起動時にダイアルアップであることを通知し、他の Gnutella クライアントはこれらのクライアントをキャッシュをしないことでキャッシュの最適化を図る。
- 接続時間による重み付け  
接続時間別に重み付けを行うことで、キャッシュする優先順位を付ける。優先度の高い、長期間ネットワークに存在するクライアントから情報をキャッシュする。

## 6 おわりに

本論文では、P2P モデルは、クライアントが自由に入りできるネットワークであることを踏まえて、ファイルの取得可能時間を評価した。また、クライアントがファイルを取得できない問題は、ファイルを持つクライアント自体に到達できないことが原因であることを示した。

P2P モデルにおいて、キャッシュを用いた場合、キャッシュした情報の整合性は、短期間での間しか有効でないことが実証した。

今後の課題としては、キャッシュの整合性とトラフィック量の両側面からとらえ、P2P モデルにおける最適なキャッシュ機構の構築が挙げられる。

## 7 謝辞

本論文を書くにあたって、最後まで指導して頂いた慶應義塾大学徳田・村井・楠本・中村・南研究会の皆様にご感謝します。

## 参考文献

- [1] <http://gnutella.wego.com>
- [2] Kunwadee Sripanidkulchai. The popularity of Gnutella queries and its implications on scalability. Carnegie Mellon University  
<http://www.cs.cmu.edu/kunwadee/research/p2p/gnutella.html>
- [3] Gnut version 0.4.25 <http://www.gnutelliums.com/linux/unix/gnut/>
- [4] The Gnutella Protocol Specification v0.4 Document Revision 1.2 Clip2 <http://www.clip2.com>