

非構造型 P2P ネットワークを用いたユーザ生成コンテンツ配信サービスプラットフォームの研究

研究代表者 朝 香 卓 也 首都大学東京 システムデザイン学部 教授

1 はじめに

近年、UGC (User Generated Content) が人気を集めており、様々な UGC を配信するサービスが提供されている [1-7]. その一つとして、ニコニコ動画というサービスが挙げられる. ニコニコ動画では、利用者が動画を見ながらコメントを投稿することができ、他の利用者がその動画を視聴したときにそれが動画上に表示される. 本サービスにおけるコメントは動画への付加情報と位置づけることができる. ユーザから付与される付加情報は、将来的に大容量化する可能性がある. その際に問題となるのは、クライアント-サーバ方式を用いた通信では、トラフィック量の増大に伴い、Web サーバの負荷が高くなることである. したがって、P2P (Peer to Peer) ネットワークを利用して、大容量化した付加情報をピア間で共有し、Web サーバ負荷を減少させるシステムの構築が必要である.

本研究では、情報追加型 UGC における付加情報を P2P ネットワークを利用してピア間で共有する ECTI (Effectively Collect with Time Information) 方式を提案する. ECTI 方式は、クエリに時刻情報を記録してフラッディングさせることで、記録された時刻よりも後に投稿された付加情報だけを選択的に発見し、収集の際のトラフィック量を削減する. 本稿では、シミュレーションによって、ECTI 方式の評価を行い、その有効性を示す.

2 従来の研究

Web サーバによってコンテンツの配信を行うサービスについては、多数のユーザによる Web サーバへのアクセス集中を軽減する技術が研究されている. 例えば、Web キャッシュサーバは、本来の Web サーバが保持するコンテンツのキャッシュを蓄積し、そのコンテンツへの要求がユーザからあったときに、本来の Web サーバに代わってコンテンツ配信を行う技術である [8]. しかし、膨大な数のユーザがキャッシュにアクセスした場合、アクセス集中によって Web キャッシュサーバ自体がダウンするという問題がある. 一方、本研究においては、特定のサーバを必要としないピア型 P2P ネットワークを利用するため、サーバダウンという問題は起こらない.

しかし、ピア型 P2P ネットワークは、ネットワーク上を流通するデータを一元管理できないという欠点があり、著作権問題に抵触する使われ方をされる傾向がある [9]. これに対処するために、例えば、中央サーバを利用して、DRM (Digital Rights Management) 技術を P2P ネットワークに組み込むことで、流通するコンテンツを管理する手法が提案されている [10]. しかし、これらの手法は、コンテンツのピア間の移動からローカルでの移動まで、逐一把握することを目的とするため、各ピアは常に中央サーバと通信する必要があり、中央サーバの負荷やスケーラビリティに問題がある. 一方、付加情報と位置づけられるコンテンツには、著作権という概念がないため、本研究においては、著作権問題を考慮する必要はない.

ピア型 P2P ネットワークで、コンテンツを検索する際には、フラッディングという手法が用いられる. これは、検索元ピアが隣接ピアに検索クエリの転送を依頼し、そのピアはさらに自身の隣接ピアに検索クエリの転送を依頼するという手法である. 一般に、クエリには TTL (Time To Live) が設けられ、転送回数が制限される. しかし、クエリの転送に制限を設けると、コンテンツのヒット率が低下するという問題が発生する. これに対処するために、保持するコンテンツが類似するピアと隣接関係を結ぶようにオーバーレイポロジを構築する手法 [11] [12] や、コンテンツ収集の際に、検索パス上のすべてのピアにコンテンツのレプリケーションを配置する手法 [13] [14] [15] が提案されている. しかし、前者の手法では、同じピアに何度もクエリが転送されてしまうクリッピングという問題が生じる. また、後者の手法では、キャッシュを配置する回数が多くなることで、必要なストレージやネットワークリソースが大きくなるという問題が生

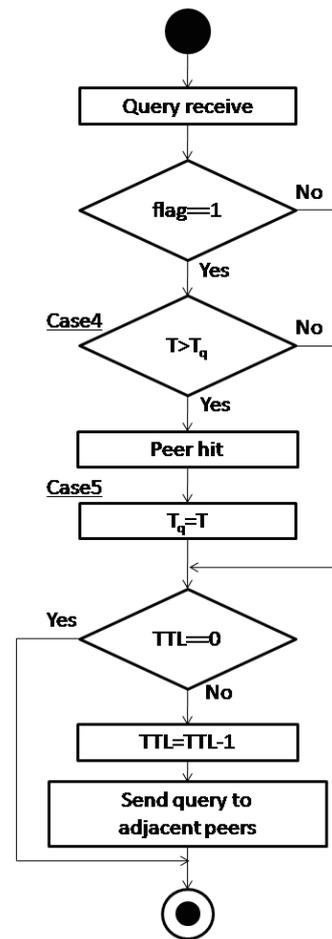
じる。一方、本研究で扱うコンテンツは付加情報であり、ある程度のヒット率の低下は許容されることが期待できるので、それらの手法を用いる必要はなく、前述した問題は発生しない。また、付加情報を対象とする研究も行われている。過去に視聴した動画における印象に残るワンシーンを再度視聴したいという場合は、現在のシークバー、早送りボタンや巻き戻しボタンでは、目的とするシーンを発見することが困難である。これに対処するために、付加情報が多く投稿されているシーンを印象に残るシーンと仮定することで、複数のサムネイルを作成し、ユーザが効率よく目的であるシーンを発見できるシステムが検討されている[16][17]。本研究は、付加情報の利用によって動画共有サイトにおけるユーザビリティの改善を目的としたシステムの検討ではなく、P2P ネットワークにおける付加情報共有方式の検討であるという点で、前述した研究とは異なる。

3 提案方式

提案方式であるECTI方式の利点は、クエリに時刻情報を記録してフラッディングさせることにより、記録された時刻よりも後に投稿された付加情報だけを選択的に発見でき、収集の際のトラフィック量を、従来方式であるEC方式に比べて、大きく削減できる点である。

まず、付加情報を検索する場合について説明する。検索元ピアが、過去に一度視聴した動画を再度視聴する場合は、過去に収集した時点での、その動画に対応する付加情報群が、利用した端末にキャッシュとして保持されている。そのような場合は、検索元ピアは、求める付加情報群のID情報に加えて、クエリに最終付加情報投稿時刻を記録してフラッディングさせる。ここで、最終付加情報投稿時刻とは保持されている付加情報群の中で最も投稿時刻が新しい付加情報の投稿時刻である。また、過去に一度も視聴していない動画を初めて視聴する場合は、対応する付加情報群は保持されていないため、検索元ピアは、クエリに初期時刻0を記録してフラッディングさせる。時刻情報をクエリに記録してフラッディングさせることによって、自身が保持する付加情報群の付加情報よりも新しい付加情報だけを、選択的に収集することが可能となる。付加情報群の転送は、クエリの転送ツリーを利用して1ホップずつ多段的に行い、同一時刻に投稿された付加情報は、転送を中継するピアで随時集約する。ただし、ECTI方式では、フラッディングされたクエリに記録された時刻以降に投稿された付加情報だけを転送する。

ECTI方式における時刻情報を利用した検索システムのフローを右の図に示す。この方式の検索でヒットするピアは、自身が保持する付加情報群の最終付加情報投稿時刻が、フラッディングされたクエリに記録された時刻よりも新しいピアである。例えば、時刻10が記録されたクエリがフラッディングされてきた場合、検索元ピアが求める付加情報群を保持しており、かつ、その付加情報群の最終付加情報投稿時刻が時刻10以降であるピアだけがヒットする。たとえ、求められた付加情報群を保持していたとしても、その最終付加情報投稿時刻がフラッディングされたクエリに記録された時刻よりも古ければ、その付加情報群を保持するピアはヒットしない。クエリに時刻情報を記録せずにフラッディングさせたときにヒットするピアの中から、古い付加情報しか持たないピアを除外することで、ヒットするピア数を減らしながらも、新しい付加情報を余すことなく収集でき、トラフィック量の削減が図れる。しかし、検索が終了するまで、終始同じ時刻がクエリに記録されフラッディングされるとすると、同一時刻に投稿された付加情報を何度も発見してしまう可能性があり、効率的ではない。そこで、フラッディングされたクエリに記録する時刻を、ピアがヒットするたびに更新する。つまり、前述のように時刻10が記録されたクエリがフラッディングされ、適切なピアがヒットした場合、次にクエリに



T_q : Time which recorded on query
 T : Time which the last additional information was posted

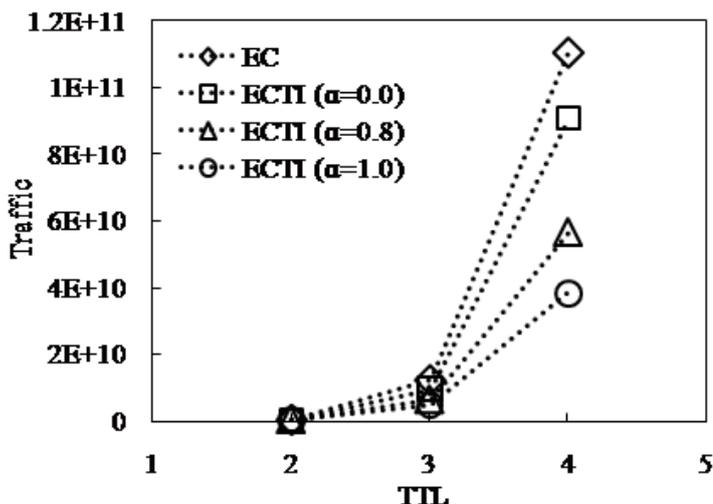
記録されフラッディングされる時刻は時刻10ではなく、ヒットしたピアの保持する付加情報群の最終付加情報投稿時刻に更新される。時刻情報の更新を行えば、同一時刻に投稿された付加情報が何度も発見されることが少なくなり、より効率的に新しい付加情報のみを見つけることができる。

4 シミュレーション評価

ECTI方式の有効性を検証するために、従来方式であるEC方式を比較方式として、シミュレーションによる性能評価を行った。表1にシミュレーション条件を示す。ここで、キャッシュ容量を超える付加情報群をダウンロードした際には、FIFO (First In First Out) [18]によりキャッシュ置き換えを行う。初期状態で各ピアがキャッシュとして保持する付加情報群の数は、べき乗則に従う。付加情報を検索し収集するピアと、そのピアが検索する付加情報群はランダムに選択される。選ばれたピアは、収集を完了した時点で新たに一つ付加情報を、自身が収集した付加情報群に追加するものとする。また、本シミュレーションでは、ピアの参加、離脱は考慮しない。

評価結果を示す。次図に、ピア数を1000、付加情報群の数を5000でそれぞれ一定とし、TTLを2, 3, 4と変化した際のトラフィック量を示す。ここで、収集の際にピア間で受け渡しされる付加情報の総数をトラフィック量と定義した。ECTI方式におけるトラフィック量はEC方式におけるトラフィック量より常に小さく、TTLが大きくなるほどその差は大きくなる。TTLが3のとき、EC方式でのトラフィック量に対して、ECTI方式 ($\alpha = 0.0$)でのトラフィック量は17%減、ECTI方式 ($\alpha = 0.8$)でのトラフィック量は39%減、ECTI方式 ($\alpha = 1.0$)でのトラフィック量は51%減であった。また、TTLが4のとき、EC方式でのトラフィック量に対して、ECTI方式 ($\alpha = 0.0$)でのトラフィック量は18%減、ECTI方式 ($\alpha = 0.8$)でのトラフィック量は49%減、ECTI方式 ($\alpha = 1.0$)でのトラフィック量は65%減であった。

EC方式とECTI方式 ($\alpha = 0.0$)では付加情報の収集方法は同じであるが、時刻情報を記したクエリをフラッディングさせる点が異なる。ECTI方式 ($\alpha = 0.0$)がEC方式よりもトラフィック量を減少させる理由は、EC方式では求める付加情報群を保持するピアがすべてヒットするが、ECTI方式ではクエリに記録された時刻より投稿時刻が新しい付加情報を保持するピアだけがヒットするからである。また、ECTI方式において、重み付け変数が1.0に近づくほど、トラフィック量が減少することが確認できる。これは、付加情報収集の基準となる時刻が、 α が1.0に近づくにつれて新しくなり、古い時刻の付加情報が収集されなくなるためである。



4 おわりに

本研究では、情報追加型UGCにおけるP2Pネットワークを利用した付加情報共有方式として、付加情報をピア間で共有することを目的としたEC方式とECTI方式を提案した。シミュレーションによってECTI方式の評価を行った結果、ECTI方式を利用した付加情報の検索および収集は、トラフィック量削減の観点でEC方式よりも有効であることが確認できたが、付加情報完成率は低下することが分かった。しかし、その低下はトラフィック量削減率に比べて小さく、対象が付加情報であることを考慮すると大きな問題にはならないと考えられる。投稿時刻別付加情報完成率は、ECTI方式では重み付け変数を用いることで、トラフィック量の増加を抑制しつつ制御できることがわかった。

今後の課題は妥当な付加情報完成率を決定することである。しかし、ユーザが満足する完成率は一概に決定できるものではないため、ニコニコ動画のユーザに対するアンケートなどにより調査を行う必要がある。

【参考文献】

- [1] YouTube, <http://jp.youtube.com/>
- [2] flickr, <http://www.flickr.com/>
- [3] mixi, <http://mixi.jp/>
- [4] twitter, <http://twitter.com/>
- [5] Wikipedia, <http://ja.wikipedia.org/>
- [6] ニコニコ動画, <http://www.nicovideo.jp/>
- [7] ニコニコ DVD, <http://dvd.nicovideo.jp/>
- [8] L. Fan, P. Cao, J. Almeida, and A. Broder, "Summary cache: A scalable wide-area web cache sharing protocol," *Proceedings of ACM SIGCOMM*, pp.254-265, Sep. 1998.
- [9] E. Adar, and B. A. Huberman, "Free riding on gnutella," *Technical report, Xerox PARC*, vol.5, no.10, Oct. 2000.
- [10] T. Iwata, T. Abe, Y. Ueda, H. Sunaga, "A DRM system suitable for P2P content delivery and the study on its implementation," *Proceedings of The 9th Asia-Pacific Conference on Communications*, vol.2, pp.806-811, Jun. 2003.
- [11] M. Li, W. C. Lee, and A. Sibasubramaniam, "Semantic Small World: An Overlay Network for Peer-to-Peer Search." *Proceedings of The 12th IEEE International Conference on Network Protocols (ICNP)*, Dec. 2004.
- [12] A. Crespo, and H. Garcia-Molina, "Semantic Overlay networks for p2p systems," *Technical report, Computer Science Department, Stanford University*, pp.1-13, Oct. 2002.
- [13] E. Cohen, and S. Shenker, "Replication strategies in unstructured peer-to-peer networks," *ACM SIGCOMM'02*, pp.177-190, Oct. 2002.
- [14] V. Gopalakrishnan, B. Silaghi, B. Bhattacharjee, and P. Keleher, "Adaptive replication in peer-to-peer systems," *Distributed Computing Systems*, pp.360-369, 2004.
- [15] A. Mondal, S. K. Madria, and M. Kitsuregawa, "An efficient context and location-based dynamic replication scheme for mobile-P2P networks," *Lect. Notes Comput. Sci.*, vol.4080/2006, pp.390-408, Sep. 2006.
- [16] 中村貴洋, 青木秀憲, 宮下芳明, "マンガ的手法を用いたニコニコ動画ナビゲーション," *情報処理学会研究報告*, vol.10, No.4, pp.103-110, 2008.
- [17] 青木秀憲, 宮下芳明, "ニコニコ動画における映像要約とサビ検出の試み," *情報処理学会研究報告*, vol.2008, No.50, pp.37-42, 2008.
- [18] I. Stoica, S. Shenker, and H. Zhang, "Core-stateless fair queueing: A scalable architecture to approximate fair bandwidth allocations in high speed networks," *IEEE/ACM Transactions on Networking*, vol.11, no.1, pp.33-46, Feb. 2003.
- [19] A.L. Barabasi and R. Albert, "Emergence of scaling in random networks," *SCIENCE*, vol.286, pp.509-512, Oct. 1999.

〈発表資料〉

題名	掲載誌・学会名等	発表年月
情報追加型 UGC における P2P を利用した付加情報共有方式	電子情報通信学会技術研究報告, ns2010-46, pp.47-52	2010年7月
セマンティック P2P ネットワークにおける効率的なキャッシュアルゴリズムを利用したコンテンツ検索方式	電子情報通信学会技術研究報告, ns2011-92, pp.57-62	2011年10月