

P2P ネットワーク情報の拡散及び消滅特性の解析

辻 琢真† 小林 郁太郎‡

†東京大学 工学部 ‡東京大学 大学院工学系研究科

1. はじめに

インターネットの持つ低コストで広範な検索や情報伝達能力に起因して、昨今ではロングテールに絡むビジネスが活況を呈している。将来的にインターネットが更なる発展を遂げるためには、例えば、難病の医者検索や、数人規模の研究者・趣味人の世界的連携といった、極めて利用頻度の低い膨大種類の情報を蓄積・検索・提供可能か否か、すなわち、ロングテールの末端を何処まで活性化できるかが重要になる。

2. 研究概要

P2P ネットワークにおけるアクセス数の少ない情報(以下、マイナー情報と表記)の成長に関する先行研究にて、十分な時間が経過した後においては、図1のようにネットワーク内の全ての情報がネットワーク内の情報蓄積数において対等となることが示されている。マイナー情報検索後の情報取得には、それが希少であるが故に多数のクエリー転送を要し、毎回の検索で多くのキャッシュがネットワーク内に蓄積されることが原因である¹⁾。先行研究ではマイナー情報に対する一定の恒常的アクセスや、Time to Live(以下、TTLと表記)に制限がないことを前提としており、上述した極めて利用頻度の低い情報の、現実的な条件のもとでの伝搬・消滅に議論が及んでいない。

そこで本研究においては、極小アクセスやTTL制限、情報源ピアや情報を所有するピアの消滅等の条件を考慮して、膨大種類のマイナー情報が

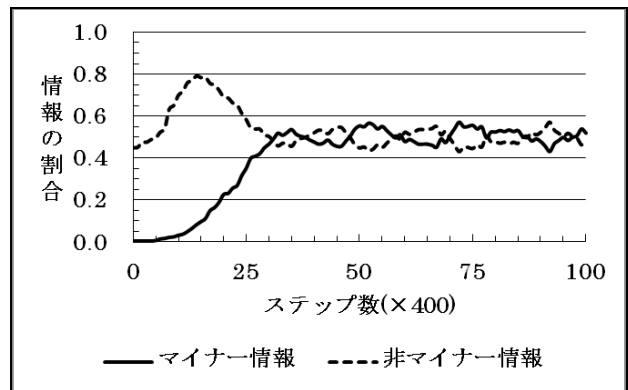


図1 マイナー情報数の時間推移

P2P ネットワーク内に蓄積されるメカニズムについて解析を行い、社会の情報インフラとして機能するための必要条件について検討している。

具体的には、シミュレーションと解析を行うことでアクセス頻度やTTL制限等と消滅確率との相関を導出し、アクセス頻度が極めて低く、消滅確率の高い情報をネットワーク内に残存蓄積させるために必要な諸要素について提案を行う。

3. シミュレーション

マイナー情報と他の情報間競争における情報数変化を、各ピアの保持情報数に上限を定めて再現した。P2P ネットワークにおける検索では、バケツリレー式に検索クエリーを転送し、経路ルート上のピアに取得情報のキーワード(情報の内容と所在に関するメタ情報)を残すという仕組みを採用している²⁾。マイナー情報が拡散するのは、情報取得時に経路上のピアに情報を残置することによる。(1)式においても情報取得までの経路が長いマイナー情報ほど、キャッシュ数変化が大きいことを示すことができる。

$$\frac{dC}{dt} = A \times L \quad (1)$$

(C: キャッシュ数 A: アクセス数 L: 経路長)

図2は、ネットワーク規模を400、TTLの上限を5と定めた際に得られたマイナー情報数の時間

P2P ネットワーク情報の拡散及び消滅特性の解析

Analysis of Information Penetration and Termination Characteristics in P2P Network

†辻 琢真 東京大学 工学部

Takuma TSUJI

Faculty of Engineering, The University of Tokyo

‡小林 郁太郎 東京大学 大学院工学系研究科

Ikutaro KOBAYASHI

Graduate School of Engineering, The University of Tokyo

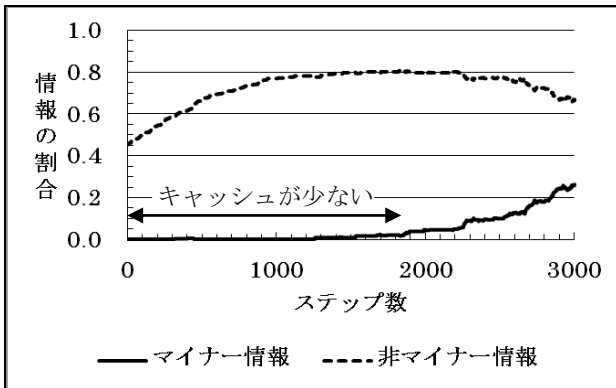


図2 TTL制限を加えた際のマイナー情報の時間推移

推移を示したものである。情報が消滅する可能性としては、図中の矢印で示したキャッシュがほとんどない段階における情報源の消滅や、キャッシュを持つピアのネットワークからの脱退等が考えられる。P2Pネットワークではピアの参加及び脱退が自由であるため、ピアの入れ替わりに伴い、脱退したピアが所有していた情報はネットワーク内から一時的に失われることとなる。本研究では、この現象について、一定時間毎にネットワーク内のピアの一部が新しいピアに入れ替わる仕組みを導入することで簡略化した。

4. 情報消滅の確率モデルと考察

シミュレーションによる解析の結果、一定数のキャッシュがネットワーク内に蓄積された以降は、時間経過と共に情報が拡散していくことを示すことができた。つまり、情報の消滅に関して、ネットワーク内に一定数のキャッシュを得ることができるか、といったことが重要な要素となる。

そこで、ネットワーク内におけるマイナー情報数の推移を、ネットワーク規模 N ・接続次数 λ ・TTL(上限 T)・検索対象率 t (情報が検索される確率)・情報取得率 r (検索情報を取得する確率)を変数とし、確率的に計算を行った。ただし、検索対象率及び情報取得率を表す関数は、図3に示す形状を取るものとする。

ステップ毎にネットワーク内のピアが自身の保有しない情報を検索するものとし、その過程で上述した情報の更新に伴う情報数変化が生ずる。このとき、ステップ数 n における情報数を A_n とすれば、増加因子 I_n 及び減少因子 D_n を用いて、情報数の漸化式を(2)式のように表現できる。

$$A_{n+1} = A_n + I_n - D_n \quad (2)$$

(2)式において、 $f \equiv A_{n+1} - A_n \geq 0$ を満たすと

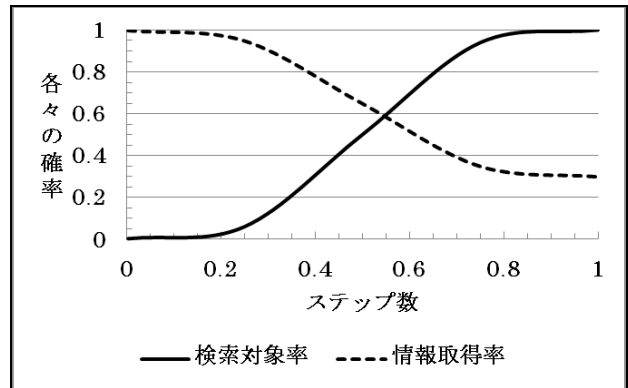


図3 検索対象率及び情報取得率

き、マイナー情報がネットワーク内に拡散することを意味する。接続次数 λ 及び TTL の上限 T としたときの検索クエリーの可能な探索ルート数 $R(\lambda, T)$ を用いると、 f の値は(3)式のように近似できる。

$$f = \frac{1}{N} \sum_{k=1}^T R(\lambda, T) \cdot k - t \cdot r \cdot R(\lambda, 1) \quad (3)$$

ここで、 λ を固定すれば、 f の第2項はほぼ定数となるため、 f の符号は第1項の値に依存することとなる。つまり、ネットワーク規模に対する TTL 上限及び接続次数に依存することとなる。

計算の過程で、近似に伴う多少の誤差が生じる可能性は否めないが、マイナー情報がネットワーク内に拡散する条件について、ネットワーク規模に対して、適切な TTL 上限及び接続次数が存在し得ることを示すことができた。また、その閾値として、検索対象率及び情報取得率が重要なパラメタとなっている。

5. 展望

将来的に、ロングテールの末端を占める極小検索頻度のマイナー情報がネットワークの主体となり、インターネットの真価となることが予想される。

今後、膨大種類のマイナー情報を維持蓄積するために必要な諸要素の検討に基づく提案を行い、インターネットが社会情報インフラとして発展する一助としたい。

参考文献

- 1) 青木真也, "クエリー転送型ネットワークの情報拡散と収束・競合", 修士論文(2009)
- 2) Michael Miller, "P2P コンピューティング入門", 翔泳社(2002)