

検索効率と汚染コンテンツ抑制の 非構造化オーバーレイに対する評価指標の検討

山中 広明^{†1} 岡村 真吾^{†2} 藤原 融^{†1}

コンテンツ検索の効率化と汚染コンテンツダウンロード抑制の両性能を両立するための非構造化オーバーレイ構築手法が提案されている。より良い構築手法を設計するには、両性能のオーバーレイに対する評価指標の設計が重要である。既存研究での評価指標は、所望コンテンツをダウンロードできる確率や、汚染コンテンツを提供する確率が異なるピアが混在する状況での汚染コンテンツ抑制効果の評価には不十分である。そこで本研究では、オーバーレイに対して両性能を統合的に評価するため、他ピアが所望のコンテンツを保持する確率、正当コンテンツを提供する確率、検索問い合わせメッセージが到達する確率、メッセージ到達に要するオーバーレイ上での距離の4つを用いた評価指標を提案する。そして、正当コンテンツ提供を行う確率がピアにより異なる状況でのシミュレーションを行い、提案指標と両性能との相関関係を検証し、有効性を確認した。

A Study on Metrics of Search Efficiency and Prevention of Polluted Contents for Unstructured Overlay

HIROAKI YAMANAKA,^{†1} SHINGO OKAMURA^{†2}
and TORU FUJIWARA^{†1}

Construction methods of unstructured overlay are known for improvement in search efficiency and prevention of downloading polluted contents. To design better construction method, we need a good metric of the two performances for overlays. Existent metrics are insufficient for measuring the probability of downloading desired contents and prevention of polluted contents in the situation where there are peers whose probabilities of providing polluted contents are diverse. In this paper, a metric of the two performances for overlays is proposed. The metric is computed from probabilities of holding desired contents, providing honest contents, and arrival of search query messages and required distance on overlay in the message arrival. By simulations, we confirm that proposed

metric is effective contents sharing in the situation where there are peers whose probabilities of providing polluted contents are diverse.

1. はじめに

近年、P2P コンテンツ共有システムの研究開発が盛んである。その中でもよく知られている、Gnutella や KaZaA などにおけるピアの論理ネットワーク（オーバーレイ）は、非構造化オーバーレイに分類される。一般に非構造化オーバーレイでは、リンク形成を各ピアが自由に行い、曖昧検索を含む様々な種類の検索に対応できるようになっている。コンテンツ検索では、問い合わせ元ピアを起点に、オーバーレイ上の論理リンクを経由して問い合わせメッセージが転送される。そのため、検索問い合わせにどのようなピアが応答するかはオーバーレイ上でのピアの配置に大きく依存する。ピアが保持し、提供するコンテンツは、そのピアの興味によって偏りがあり、また、P2P システムでは、誰でもピアとして参加できるため、ウイルスを混入させたり改変されたりした汚染コンテンツを提供するピアの存在が懸念される。そこで、非構造化オーバーレイに関する既存研究では、コンテンツ検索の効率化と汚染コンテンツのダウンロード抑制の両立を目的としたオーバーレイ構築手法が提案されている¹⁾。より良いオーバーレイ構築手法を設計するには、コンテンツ検索および汚染コンテンツダウンロード抑制を計測する、オーバーレイに対する評価指標の設計が重要である。

検索効率の評価指標は、検索にかかる通信量の少なさや、ピアからコンテンツをダウンロードしたとき、それが所望のコンテンツである確率の高さを反映していることが望ましい。また、汚染コンテンツダウンロード抑制の指標としては、コンテンツをダウンロードしたとき、それが汚染コンテンツである確率の低さを反映していることが望ましい。しかし、既存研究におけるオーバーレイに対する指標では、コンテンツ検索に関しては、問い合わせ元ピアと検索対象コンテンツを保持するピア間の距離により通信量を計測している²⁾が、所望のコンテンツ入手に関しては、隣接ピアの保持コンテンツについてのみ計測¹⁾しており、実際のダウンロードで所望のコンテンツを得られる確率の計測には不十分である。また、汚染コンテンツダウンロード抑制に関しては、汚染コンテンツ提供ピアとオーバーレイ上の他ピア

^{†1} 大阪大学 大学院情報科学研究科

Graduate School of Information Science and Technology, Osaka University

^{†2} 奈良工業高等専門学校 情報工学科

Information Engineering, Nara National College of Technology

との平均距離を計測³⁾しているが、ピアのコンテンツ提供に関するモデルとして、「常に正当コンテンツ提供」と「常に汚染コンテンツ提供」の2つしか考慮していないため、様々な確率で正当コンテンツを提供するピアの混在が考えられる P2P システムには不十分である。

そこで本研究では、コンテンツ検索の効率性と、正当コンテンツ提供を行う確率が異なるピアが混在する状況での汚染コンテンツダウンロード抑制を統合的に評価する、オーバーレイに対する指標を提案する。提案指標では、検索効率を評価するため、他のピアが所望のコンテンツを保持する確率および十分な確率でのメッセージ到達に要する転送経路の距離を用いる。また、汚染コンテンツ抑制を評価するため、ピアが正当コンテンツを提供する確率を用いる。さらに、両者の評価のため、問い合わせ元ピアからの問い合わせメッセージが到達する確率も用いる。そして、非構造化オーバーレイ上でのコンテンツ共有のシミュレーションを、正当コンテンツを提供する確率がピアにより異なる状況で行い、提案した指標と、コンテンツ検索効率性、および汚染コンテンツダウンロード抑制効果との相関性を検証し、提案する指標の有効性を確認する。

以下、2章では、非構造化オーバーレイ上でのコンテンツ授受について述べる。3章では、オーバーレイに対する既存評価指標の問題点と提案する評価指標の設計方針を述べる。次に、4章では評価指標を提案し、5章ではその有効性を検証する。最後に6章でまとめと今後の課題を述べる。

2. 非構造化オーバーレイ上でのコンテンツ共有

本章では、非構造化オーバーレイ上でのコンテンツ共有システムについて述べる。本研究ではオーバーレイを、ピアをノード、論理リンクをエッジとする無向グラフ G として考える。全ピアの集合を P 、 G 上でピア $i \in P$ とのエッジが存在するピアの集合を B_i とする。ピアは、以下で述べるようにコンテンツを所望し、そのコンテンツを入手するため、オーバーレイ上でのコンテンツ授受を行うことで、コンテンツ共有システムが運用される。

2.1 ピアの所望コンテンツ

ピアが所望するコンテンツは、そのピアが何に興味を持つかによって決まるものと考えられる。そこで、文献 4) と同様、ピアの所望コンテンツを、カテゴリとよばれるコンテンツの属性によりモデル化する。各正当コンテンツには少なくとも1つのカテゴリが与えられており、それらカテゴリの集合を D とする。ピアの興味はカテゴリによって表現され、ピアの所望するコンテンツは、それに基づくものとする。 d_i を、ピア $i \in P$ が興味を持つカテゴリを示す $|D|$ 次元ベクトルとする。 d_i の要素は全て非負実数であり、そのうち少な

くとも1つは正数、全要素の和は1とする。 d_i の n ($1 \leq n \leq |D|$) 番目の要素 $d_i(n)$ は、ピア i のカテゴリ $n \in D$ に対する興味の大さを示す尺度であり、カテゴリ n のコンテンツ検索を行う確率とする。

2.2 コンテンツ授受

コンテンツ授受は、オーバーレイ上でのコンテンツ検索およびダウンロードにより行われる。まずピアはコンテンツ検索、すなわち所望のコンテンツを保持しているピアの問い合わせを行なう。本研究では問い合わせ法として、フラッディング⁵⁾を取り上げる。

フラッディングでは、まず、コンテンツを入手しようとする問い合わせ元ピア $i \in P$ が問い合わせメッセージを、 G 上の全隣接ピアに転送する。メッセージを受信したピアの1つである $j \in P$ は、問い合わせに回答する場合はピア i に直接回答し、受信メッセージに対する処理を終了する。回答しない場合は、メッセージに含まれる非負整数変数 t を1減じる。減じた結果0になった場合、受信メッセージに対する処理を終了する。減じた結果が正だった場合、メッセージの送信元ピアを除く全隣接ピアにメッセージを転送する。なお、問い合わせメッセージには、問い合わせ元ピア i が乱数生成器によって生成したメッセージ ID も含まれており、ピア j はその ID が過去に処理したことのあるメッセージに含まれている場合、メッセージに対する処理は一切行わずメッセージを破棄するものとする。これは、 G 上の閉路が原因で、重複して問い合わせメッセージを受け取った場合の冗長な処理を回避するためである。問い合わせメッセージを受け取った全てのピアは、上述のピア j のように処理を行う。なお、問い合わせ元ピアが変数 t に設定する値は Time-To-Live (TTL) と呼ばれ、問い合わせメッセージを受信するピアの、問い合わせ元ピアからの G 上での最短距離の上限を決めるパラメータである。

問い合わせメッセージへの回答は、受信したピアが、問い合わせに該当するコンテンツを保持しているか否かと、正当コンテンツを提供する確率により決まるものとする。ピアがある確率に基づき正当コンテンツ、汚染コンテンツのいずれかを提供するモデルを考える。ピア $j \in P$ が正当コンテンツを提供する確率を f_j 、汚染コンテンツを提供する確率を $(1 - f_j)$ とする。ピア j は、問い合わせメッセージを受信したとき、問い合わせに該当するコンテンツを保持していれば、必ず回答する。一方、保持していなければ、汚染コンテンツを提供するため確率 $(1 - f_j)$ で回答する。これら以外の場合回答しない。

そして、問い合わせ元ピア i は、回答したピアからランダムに選んだ1つのピア $k \in P$ にコンテンツ提供を要求し、そのピアが提供するコンテンツをダウンロードする。

3. 評価指標の設計

2章で述べた非構造化オーバーレイ上でのP2Pコンテンツ共有における、検索効率と汚染コンテンツ抑制のオーバーレイに対する評価指標を設計する。検索にかかる通信量が平均的に小さく、また、ピアからコンテンツをダウンロードしたとき、それが所望のコンテンツである確率が高いとき、検索効率が良いといえる。また、コンテンツをダウンロードしたとき、それが汚染コンテンツである確率が低いとき、汚染コンテンツのダウンロードの抑制効果が高いといえる。これらがどの程度満たされているか、統合的に評価する評価指標を設計する。

3.1 オーバレイに対する既存評価指標

まず、検索にかかる通信量の評価指標として、文献2)では、問い合わせ元ピアと検索対象コンテンツを保持するピアのうち、 G 上での距離が最短のピアまでの距離を計測している。この距離と同時に、コンテンツの検索のために全てのピアが転送した問い合わせメッセージ数も計測しており、最短距離が短ければ、転送メッセージ数も少なく、通信量において効率が良い結果になっている。また、文献1), 6), 7)では、検索において全リンク上を転送された問い合わせメッセージ数により通信量を計測しているが、これはピアが転送した問い合わせメッセージ数²⁾と同じ数値である。

次に、ダウンロードしたコンテンツが所望のコンテンツである確率については、文献1)で、自ピアと隣接ピアの保持コンテンツの類似度を計測している。類似度が高ければ、興味も類似しており、自ピアの所望するコンテンツを保持する可能性は高いが、2.2節で述べたフラッディングでは、隣接ピアが以外のピアにも問い合わせメッセージが到達する可能性があるため、隣接ピアのみの計測では不十分である。なお、文献8)では、検索結果、すなわち問い合わせに回答したピアが提供予定のコンテンツについて、適合率を評価しているが、2.2節で述べたように、ダウンロード先ピアをランダムに選択する場合、適合率はダウンロードしたコンテンツが所望のコンテンツとなる確率である。

汚染コンテンツのダウンロード抑制に関して、文献1)では、自ピアと隣接している汚染コンテンツ提供ピア数を指標としている。しかし、2.2節で述べたフラッディングでは、隣接ピア以外のピアも回答する可能性があるため、隣接ピアのみの計測では不十分である。文献3), 9)では、1つのピアと G 上の他ピアまでの平均距離を指標としている。特に文献3)では、回答したピアに占める汚染コンテンツ提供ピアの割合も同時に計測している。平均距離が大きくなる変化と、回答した汚染コンテンツ提供ピアの割合が小さくなる変化が連動しており、指標の有効性が分かる。このような連動がみられるのは、汚染コンテンツ提供ピア

と問い合わせ元ピアの G 上での距離が長くなることで、問い合わせメッセージが到達しにくくなるためである。ただし、コンテンツ提供に関しては、「常に正当コンテンツ提供」と「常に汚染コンテンツ提供」の2種類しか考えられておらず、様々な確率で正当コンテンツを提供するピアが混在している場合には、対応できない。

3.2 設計方針

検索効率、汚染コンテンツダウンロード抑制のオーバーレイに対する指標を設計する。検索にかかる通信量は、文献2)より、問い合わせ元ピアから問い合わせメッセージに回答するピアまでの最短距離に依存する。ただし、十分な確率で問い合わせメッセージが到達しなければ、実用的ではないので、十分な確率で到達するときの転送経路の距離を考える。また、所望のコンテンツ入手と汚染コンテンツ抑制には、文献1), 3)より、回答するピアが所望のコンテンツを保持する確率と正当コンテンツを提供する確率が重要だと考えられる。

以上を踏まえ、オーバーレイに対する評価指標を提案する。提案する指標では、他のピアが所望のコンテンツを保持する確率(4.1節)、他のピアへ問い合わせメッセージが到達する確率(4.2節)、十分な確率でのメッセージ到達に要する転送経路の距離(4.3節)、正当コンテンツを提供する確率を用い、指標を定義する(4.4節)。

4. 評価指標の提案

4.1 所望のコンテンツを保持する確率

ピア $x \in P$ が、あるピア $i \in P$ の所望のコンテンツを保持する確率を求める。まず、各カテゴリ $n \in D$ について、 n に属するコンテンツ全体のうちピア x が保持しているコンテンツの割合を $|D|$ 次元ベクトル \mathbf{h}_x で示す。ベクトル \mathbf{h}_x の n ($1 \leq n \leq |D|$)番目の要素 $\mathbf{h}_x(n)$ ($0 \leq \mathbf{h}_x(n) \leq 1$)を次式のように定義する。

$$\mathbf{h}_x(n) = \frac{(x \text{ が保持する } n \text{ のコンテンツ数})}{(n \text{ の全てのコンテンツ数})}$$

ピア i の所望のコンテンツは、2.1節で述べたベクトル \mathbf{d}_i で表す。 \mathbf{d}_i は、ピア i がどのカテゴリのコンテンツをどれだけ確率で入手しようとするのかを示している。したがって、式(1)に示すように、ベクトル \mathbf{d}_i と \mathbf{h}_x の内積でピア x がピア i の所望のコンテンツを保持する確率 $ds(i, x)$ を求める。

$$ds(i, x) = \langle \mathbf{d}_i, \mathbf{h}_x \rangle \quad (1)$$

4.2 問い合わせメッセージの到達確率

ピア i からピア x ($i, x \in P, i \neq x$)まで、問い合わせメッセージが、距離 u ($1 \leq u \leq \text{TTL}$)

以下の経路で到達する確率 $al(G, i, x, u)$ を求める。まず、ピア i とピア x が隣接している ($x \in B_i$) 場合、 i から x へメッセージを直接送信するので必ず到達する。したがって、 $al(G, i, x, u) = 1$ となる。一方、隣接していない場合、距離 1 の経路では到達できないので、 $x \notin B_i$ かつ $u = 1$ のとき、確率 $al(G, i, x, 1) = 0$ になる。

これら以外の場合、 ix 間の距離 2 以上 u 以下の経路で到達できる可能性があるが、そのときの到達確率は、経路途中のピアの問い合わせメッセージを転送する確率により決まる。なお、2.2 節より、ピアは同一問い合わせメッセージを 2 回以上処理しないので、ここで考える経路は閉路を含まない。経路途中のピア $j (\neq i, x)$ が、メッセージを転送するのは、応答しないときである。2.2 節より、問い合わせに該当するコンテンツを保持している場合は必ず応答し、保持していない場合は確率 $(1 - f_j)$ で応答する。これとピア j がピア i の問い合わせに該当するコンテンツを保持する確率は $ds(i, j)$ なので、応答する確率は $ds(i, j) + (1 - ds(i, j))(1 - f_j)$ である。したがって、ピア j がピア i からの問い合わせメッセージを転送する確率 $fw(i, j)$ は、次式 (2) のように求められる。

$$fw(i, j) = f_j(1 - ds(i, j)) \quad (2)$$

このように、経路上の各ピアがメッセージを転送する確率は求めることができるが、ピア i から x まで距離 u 以下の経路で到達する実際の確率を求めようとすると、部分経路が等しい複数の経路が存在する場合が考えられ、計算が複雑になる。ただし、1 つの経路上を転送される確率は、その経路途中のピア j の転送確率 $fw(i, j)$ の積により求めることができ、その 1 つしか経路が存在しない場合も含め、これは、実際の確率の下界である。そこで、ピア ix 間の距離 u 以下の全ての閉路を含まない各経路について、経路上を転送される確率を求め、より実際の確率に近い下界を求めるため、求めた確率の中で最大の確率を ix 間でのメッセージが転送される近似的な確率とする。

以上より、確率 $al(G, i, x, u)$ は、次式 (3) のようになる。

$$al(G, i, x, u) = \begin{cases} 1 & (x \in B_i) \\ 0 & (x \notin B_i, u = 1) \\ \max \left\{ \prod_{k=2}^v fw(i, p_k) : 2 \leq v \leq u, p \in AC_{i,x}^{G,v} \right\} & (\text{otherwise}) \end{cases} \quad (3)$$

ただし、 $AC_{i,x}^{G,v}$ は、 G 上の閉路を含まない、 ix 間の距離 v の経路を示す頂点列 $p = \{p_1, p_2, \dots, p_{v+1}\}$ ($p_1 = i, p_{v+1} = x, p_k \neq p_l (k \neq l)$) の集合である。

なお、フラッディングでは、問い合わせメッセージが転送される経路の最大距離は TTL

なので、ピア i からピア x までメッセージが到達する確率は、 $al(G, i, x, \text{TTL})$ である。

4.3 メッセージ転送に要する経路の距離

コンテンツ検索の効率性を考える。まず、ピア間の問い合わせメッセージ転送の効率性を考えたとき、距離は小さい方が効率的だが、その距離での到達確率が十分でなければ実用的でない。到達確率は、距離について単調増加 ($al(G, i, x, u)$ は u について単調増加) だが、十分な確率に達する距離が大きいと効率的でない。そこで、式 (4) に示す、ピア i からピア x へのメッセージの到達確率が十分な確率 q_s に達する最短距離 $ad(G, i, x, q_s)$ を、メッセージ転送の効率性指標とする。

$$ad(G, i, x, q_s) = \begin{cases} \min\{u : al(G, i, x, u) \geq q_s\} & (al(G, i, x, \text{TTL}) \geq q_s) \\ \infty & (\text{otherwise}) \end{cases} \quad (4)$$

$ad(G, i, x, q_s)$ は小さい方が、転送効率が良いと考える。

4.4 オーバレイに対する評価指標

ピア $i \in P$ にとってのオーバレイ G の評価指標を定義する。3.2 節より、他のピアへのメッセージの到達確率、他のピアが所望のコンテンツを保持する確率、正当コンテンツを提供する確率の積は大きいと良い。また、問い合わせメッセージの転送に要する経路の距離は小さい方が良い。以上より、ピア i のオーバレイ G に対する評価指標 $util(G, i)$ を次式 (5) のように定義する。

$$util(G, i) = \begin{cases} \left(\sum_{x \in AL_i^G} \frac{al(G, i, x, \text{TTL})}{ad(G, i, x, q_s)} f_x \cdot ds(i, x) \right) / |AL_i^G| & (AL_i^G \neq \emptyset) \\ 0 & (\text{otherwise}) \end{cases} \quad (5)$$

ただし、 $AL_i^G = \{x \in P \setminus \{i\} : al(G, i, x, \text{TTL}) > 0\}$ であり、フラッディングにより、ピア i から、問い合わせメッセージが 0 より大きな確率で到達するピアの集合とする。

5. 評価指標の有効性検証

コンテンツ授受のシミュレーションにより、提案した評価指標の有効性を検証する。シミュレーションでは、オーバレイを提案指標で計測する。同時に、検索にかかる通信量、所望コンテンツおよび汚染コンテンツをダウンロードする確率を計測し、その結果と提案指標の相関性により、有効性を検証する。

5.1 検索効率性と汚染コンテンツ抑制効果の計測

提案指標の有効性を検証するため、コンテンツ検索にかかる通信量、検索により得られた応答ピアから所望コンテンツおよび汚染コンテンツをダウンロードするそれぞれの確率を、別に計測する。あるピアの1回の検索にかかる通信量の計測は、検索においてオーバーレイの全エッジ上での転送されたメッセージ数^{(6),(7)}により行う。ピア i のある1回のコンテンツ検索において、所望のコンテンツおよび汚染コンテンツそれぞれのダウンロード確率 $P_h(i)$, $P_p(i)$ は、応答したピアからランダムに選んでダウンロードするので、それぞれ次式のように求められる。

$$P_h(i) = \begin{cases} \frac{\sum_{x \in RES_i} f_x}{|RES_i|} & (RES_i \neq \emptyset) \\ 0 & (\text{otherwise}) \end{cases}$$

$$P_p(i) = \begin{cases} \frac{\sum_{x \in RES_i} (1 - f_x)}{|RES_i|} & (RES_i \neq \emptyset) \\ 0 & (\text{otherwise}) \end{cases}$$

ただし、 RES_i はピア i のある1回のフラッディングによる問い合わせに回答したピア集合である。

なお、提案指標による計測は、シミュレーション開始時にピアごとに行い、全ピアにおける平均値を5.3節では示す。また、その他の指標は、その後行う100回のコンテンツ授受で毎回計測し、各回の全ピアにおける平均値の100回での平均値を5.3節では示す。これは、オーバーレイに対する提案指標による計測結果と、平均的な通信量、所望のコンテンツおよび汚染コンテンツをダウンロードする確率との関係を調査するためである。したがって、コンテンツ授受によりダウンロードしたコンテンツは、それ以降の授受で受信した問い合わせメッセージに対する処理において保持コンテンツとして扱わず、他ピアに提供もしない。提供するコンテンツはシミュレーション開始時に保持しているものだけである。

5.2 シミュレーション設定

2章で述べたコンテンツ共有のシミュレーションを行う。基礎的な検証を行うため、シミュレーションで用いるオーバーレイ G はランダムグラフとする。 G 上の各ノード間について、確率 r でエッジを結び、生成する。 r はランダムグラフ生成時のパラメータである。

各ピアには所望するコンテンツが設定され、コンテンツ授受中はそれらの中でダウンロードしたことがないコンテンツについて、問い合わせを行う。カテゴリごとの所望するコンテンツ数は、2.1節で述べたモデルに基づき、ピアの各カテゴリへの興味の大きさにより設定

表1 シミュレーションの主なパラメータ

Table 1 Significant parameters in the simulation

| | |
|------------------|------|
| 全ピア数 $ P $ | 500 |
| カテゴリ数 $ D $ | 200 |
| 全コンテンツ数 | 2000 |
| TTL | 4 |
| 式(4)のパラメータ q_s | 0.5 |

する。なお、興味を持つカテゴリ ($d_i(n) > 0$ なる n)、および興味を持つカテゴリ内で所望コンテンツに設定するコンテンツは、文献4)と同様、それぞれカテゴリ自体の人気度順位、コンテンツのカテゴリ内での人気度順位に基づく。

シミュレーション開始時、各コンテンツについて P 中からランダムに1ピアを選び、保持させる。コンテンツ授受においてピアが提供するコンテンツは、このとき保持させられたコンテンツのみである。

以上のような設定でシミュレーションを行う。その他の主なパラメータは表1に示す。

5.3 有効性検証

シミュレーション結果を示し、提案指標の有効性を検証する。なお、本節で示す結果はいずれも、同一シミュレーションを5回行った平均値である。

検索効率性との相関

ピア検索にかかる通信量および所望のコンテンツをダウンロードする確率と、提案指標の相関性を調査する。ここでは、全ピアが必ず正当コンテンツを提供する ($f_i = 1$) もとする。

まず、所望コンテンツのダウンロード確率のみが向上する状況での相関性について調査するため、オーバーレイ G 生成時のパラメータ r を $[0.001 : 0.01]$ のいくつかの値に設定して、それぞれのオーバーレイ上でシミュレーションを行った。パラメータ r を大きくすることで、任意のピア間で問い合わせメッセージが到達する確率は高くなり、所望コンテンツのダウンロード確率も高まる。しかし、 r が大きくなることで、 G 上のエッジ数は多くなるため通信量は増加すると考えられる。図1は、各 r によって生成したランダムグラフ上でのコンテンツ授受における提案指標と、所望コンテンツダウンロード確率(図1(a))、通信量(図1(b))を示したものである。図1(a)より r の増加に伴い所望コンテンツのダウンロード確率が向上しているが、同時に、図1(b)より通信量は増加している。図1においては提案指標と所望コンテンツダウンロード確率、通信量との間に相関が見られない。

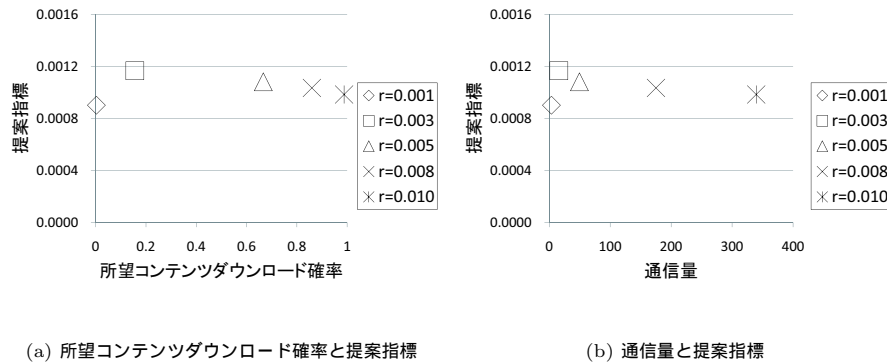


図 1 パラメータ r 変化時の検索効率性と提案指標の相関
Fig. 1 Correlation between search efficiency and proposed metric according to the parameter r

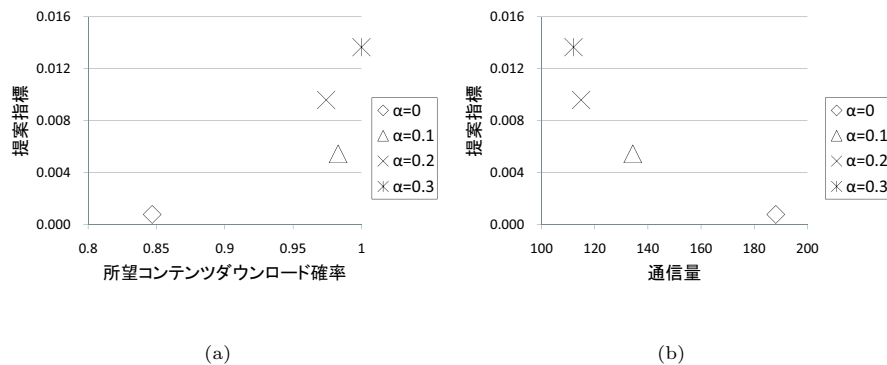


図 2 割合 α 変化時の検索効率性と提案指標の相関
Fig. 2 Correlation between search efficiency and proposed metric according to the portion α

次に、所望コンテンツのダウンロード確率が向上し、通信量が同時に減少し、検索が効率化されるときに相関性を調査するため、 G 上のピアが提供するコンテンツを増加させるシミュレーションを行った。5.2 節で述べたシミュレーション開始前の所望コンテンツおよびコンテンツ提供者の設定直後、各ピアに対して、所望コンテンツの中から一律の割合の数だけランダムに選んだコンテンツを保持させ、コンテンツ授受においてそのピアの提供コンテンツとする。なお、一律の割合を α とする。これにより、各ピアは他ピアの所望コンテンツを保持する確率が高くなり、所望コンテンツのダウンロード確率も高くなると考えられる。また、 G 上のエッジ数は変化せず、ピアが提供するコンテンツは増加するので、通信量は減少すると考えられる。図 2 は、 $r = 0.008$ として生成したオーバーレイ上で、各割合 α に設定して行ったコンテンツ授受における提案指標と、所望コンテンツダウンロード確率 (図 2(a))、通信量 (図 2(b)) を示したものである。図 2 より、 α の増加に伴い所望コンテンツダウンロード確率が向上し、通信量は減少して、検索が効率化されていることが分かる。また、検索の効率化に伴い、提案指標の値が大きくなる相関が見られる。

以上より、提案指標は所望コンテンツダウンロード確率のみ向上しているときは値が増加しないが、通信量も同時に減少しているとき値が増加することから、検索効率性の指標として有効であるといえる。

汚染コンテンツダウンロード抑制効果との相関

ピアの汚染コンテンツを提供する確率による、汚染コンテンツをダウンロードする確率と、提案指標との相関を調査するため、500 ピア中 100 ピアの汚染コンテンツを提供する確率を 0.5 から 1.0 まで 0.1 おきに設定し、それぞれの確率での汚染コンテンツダウンロード確率と提案指標を計測した。100 ピアの汚染コンテンツを提供する確率を $(1 - F)$ とする。その他の 400 ピアは必ず正当コンテンツを提供する ($f_i = 1$)。なお、これら 100 ピアは、全ての正当コンテンツを保持しており、式 (2) より受信した全ての問い合わせメッセージに回答する。これは、ピアの全問い合わせメッセージへの応答等にかかるコストを考えると、あまり現実的ではないかもしれないが、汚染コンテンツを提供する機会が最も多くなる最悪な状況を想定したものである。オーバーレイ G は、パラメータ $r = 0.01$ として生成したランダムグラフである。

図 3 は、100 ピアが汚染コンテンツを提供する確率 $(1 - F)$ の各値における、汚染コンテンツダウンロード確率と提案指標を示したものである。図 3 より、 $(1 - F)$ が高いほど、汚染コンテンツダウンロード確率は高くなり、同時に提案指標の値も低くなっていることが分かる。これより、提案指標は様々な確率で汚染コンテンツを提供するピアが存在する状況

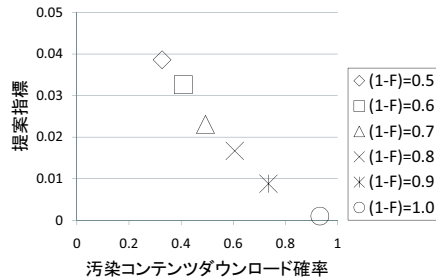


図3 汚染コンテンツダウンロード確率との相関

Fig.3 Correlation between the probability of downloading polluted contents and proposed metric

で、汚染コンテンツをダウンロードする確率を表す指標として有効であることが分かる。

6. おわりに

本研究ではオーバーレイに対して、コンテンツ検索の効率性と正当コンテンツ提供を行う確率が異なるピアが混在する状況での汚染コンテンツダウンロード抑制を統合的に評価するため、他ピアが所望のコンテンツを保持する確率、正当コンテンツを提供する確率、検索問い合わせメッセージが到達する確率、メッセージ到達に要するオーバーレイ上での距離を用いた評価指標を提案した。ランダムグラフ上でのコンテンツ共有シミュレーションでは、提案指標と検索効率、汚染コンテンツダウンロード抑制効果との相関性がみられ、提案指標が有効であることが分かった。今後の課題として、ランダムグラフ以外のオーバーレイ上での有効性検証、提案指標による既存のオーバーレイ構築手法の評価などが挙げられる。

謝辞 本研究の一部は科学研究費補助金(18049050)の助成を受けたものである。ここに記して深謝する。

参考文献

- 1) Pogkas, I., Kriakov, V., Chen, Z. and Delis, A.: Adaptive neighborhood selection in peer-to-peer networks based on content similarity and reputation, *Peer-to-Peer Networking and Applications*, Vol.2, No.1, pp.37-59 (2009).
- 2) Sripanidkulchai, K., Maggs, B. and Zhang, H.: Efficient content location using interest-based locality in peer-to-peer systems, *Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications (INFOCOM 2003)*, Vol.3,

pp.2166-2176 (2003).

- 3) Tian, H., Zou, S., Wang, W. and Cheng, S.: Constructing efficient peer-to-peer overlay topologies by adaptive connection establishment, *Computer Communications*, Vol.29, No.17, pp.3567-3579 (2006).
- 4) Schlosser, M.T., Condie, T.E. and Kamvar, S.D.: Simulating a File-Sharing P2P Network, Technical Report 2003-28, Stanford InfoLab (2003).
- 5) Meshkova, E., Riihijärvi, J., Petrova, M. and Mähönen, P.: A survey on resource discovery mechanisms, peer-to-peer and service discovery frameworks, *Computer Networks*, Vol.52, No.11, pp.2097-2128 (2008).
- 6) Fletcher, G.H., Sheth, H.A. and Börner, K.: Unstructured Peer-to-Peer Networks: Topological Properties and Search Performance, *Agents and Peer-to-Peer Computing, LNCS3601*, pp.14-27 (2005).
- 7) Cohen, E., Fiat, A. and Kaplan, H.: Associative search in peer to peer networks: Harnessing latent semantics, *Computer Networks*, Vol.51, No.8, pp.1861-1881 (2007).
- 8) Zhu, Y., Yang, X. and Hu, Y.: Making Search Efficient on Gnutella-Like P2P Systems, *Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS'05)*, pp.56a-56a (2005).
- 9) Condie, T., Kamvar, S.D. and Garcia-Molina, H.: Adaptive Peer-to-Peer Topologies, *Proceedings of the 4th International Conference on Peer-to-Peer Computing (P2P '04)*, IEEE Computer Society, pp.53-62 (2004).