

アドホックネットワークにおける データの複製配置を考慮した Top-k 検索手法

佐々木 勇和^{†1} 原 隆浩^{†1} 西尾 章治郎^{†1}

アドホックネットワークでは、必要なデータのみを効率的に取得するため、端末が何らかの値（スコア）によって順序付けされたデータの上位 k 個以内のものを検索する Top-k 検索を用いることが有効である。ここで、端末の移動によりネットワークが分断した場合、全ての端末にアクセスできないため、検索精度が低下する。そこで、検索精度の低下を防ぐためにデータの複製を配置することが有効である。しかし、複製を考慮せずにデータの返信を行うと、同じデータが複数回返信されたり、検索クエリ発行端末から遠い端末がデータを返信する可能性が高く、トラフィックが増加してしまう。そこで、本稿では、アドホックネットワークにおいて、データの複製を考慮した Top-k 検索手法を提案する。提案手法では、検索クエリメッセージに各端末のデータ（複製）保持情報を添付することにより、同じデータが複数回返信されることを防ぐとともに、クエリ発行端末に近い端末がデータを返信することを可能にする。

A Top-k Query Method considering Data Replication in Mobile Ad Hoc Networks

YUYA SASAKI,^{†1} TAKAHIRO HARA^{†1}
and SHOJIRO NISHIO^{†1}

In mobile ad hoc networks(MANETs), to acquire only necessary data items, it is effective that each mobile node retrieves data items using a Top-k query, in which data items are ordered by the score of a particular attribute and the query-issuing node acquires data items with the k highest scores. When network partitioning occurs, the query-issuing node cannot connect to some nodes, and thus, the accuracy of the query result may decrease. To solve this problem, it is effective to replicate data items on mobile nodes. However, if each node sends back its own data items(replicas) responding to a query without considering replicas held by others, same data items are sent back to the query-issuing node more than one through long paths, which results in increase of traffic. In this paper, we propose a top-k query processing method considering data replication in MANETs. In this method, each node attaches a query message with the information on its own data items, and thus, duplicate transmissions of same data items through long paths can be suppressed.

1. 序 論

近年、無線通信技術の発展と計算機の小型化や高性能化に伴い、ルータ機能をもつ端末のみで一時的な無線ネットワークを形成するアドホックネットワークへの関心が高まっている。アドホックネットワークにおけるデータ検索では、複数の端末に限られた通信帯域を共有するため、膨大なデータの中から必要なデータのみを効率的に取得する必要がある。特に各端末に限られた資源を割り当てる場合や関連性の高い情報のみを収集する場合、検索条件とデータの属性値で決定する何らかの値（スコア）によって順序付けられたデータの上位 k 個のものを検索する Top-k 検索を用いることが有効である。

ここで、アドホックネットワークでは、端末の移動によりネットワークポロジが動的に変化し、ネットワークが分断される可能性がある。この場合、分断された一方のネットワークに属する端末は、他方のネットワークに属する端末のもつデータにアクセスできなくなる。アドホックネットワークにおいて Top-k 検索の時に、ネットワークが分断している場合、検索クエリを発行した端末は、自身が属していない部分ネットワークに属する端末が保持するデータを取得できない。そのため、検索精度が低下してしまう可能性がある。ネットワーク分断時の検索精度を向上させるアプローチとして、各端末が他の端末の保持するオリジナルデータの複製を作成し、ネットワークが分断してもデータにアクセスできるようにすることが有効である。

これまでに筆者らの研究グループでは、文献 3) において、アドホックネットワーク上のトラフィックの削減と検索精度の維持を目的とする Top-k 検索手法、および文献 2) において Top-k 検索のための複製配置手法を提案した。ここで、文献 2) の手法により複製を配置した場合、スコアの高いデータの複製は多くの端末に保持される可能性が高い。このような状況で、文献 3) の手法で Top-k 検索を行うと、同じデータ（複製）が複数回返信されたり、検索クエリの発行端末から遠い端末よりデータが返信されたりすることで、大きなトラフィックが発生してしまう。

そこで本稿では、文献 3) の手法を拡張し、アドホックネットワークにおけるデータの複製配置を考慮した Top-k 検索手法を提案する。提案手法では、同じデータが複数回返信されることを抑止するために、各端末が自身のもつデータのうち上位 k 個に入る可能性のあるデータの識別子をクエリメッセージに添付する。これにより、クエリの伝搬経路上で、自身よりクエリ発行端末に近い端末が保持するデータを把握することができる。データの返信時に各端末は自身よりクエリ発行端末に近い端末が保持しているデータを返信しない。また、異なる経路上を伝搬したクエリメッセージを受信したとき、その経路上に同じデータを

^{†1} 大阪大学
Osaka University

保持する端末が存在した場合、その端末の方がクエリ発行端末に近ければ、そのデータを返信しない。

以下では、2章で関連研究を紹介する。3章で想定環境について述べ、4章で先行研究を紹介する。5章で提案手法について述べ、6章でシミュレーション実験の結果を示す。最後に7章で本稿のまとめと今後の課題について述べる。

2. 関連研究

近年、さまざま分野において Top-k 検索に関する研究が盛んに行われているが、アドホックネットワークにおける Top-k 検索に関する研究は、筆者らの先行研究を除いてほとんど行われていない。そこで、端末が自身や周辺端末情報のみを用いて自律分散的に動作するという観点でアドホックネットワークと類似している非構造型 Peer to Peer(P2P) ネットワークに着目し、Top-k 検索について紹介する。非構造型検索 P2P ネットワークでは、検索クエリは検索クエリ発行端末から隣接端末にブロードキャストされ、これを受信した端末からさらにその隣接端末に転送されるフラッディングという方法で処理される。文献 5) では、端末がクエリ応答の要求数 k を含めた検索クエリをフラッディングし、これを受信した端末が自身のもつデータの中から上位 k 個のスコアをもつものを返信する手法を提案している。また文献 9) では、1 回の検索クエリの送信で 1 個のデータを取得し、この処理を k 個の検索結果が取得できるまで繰り返す Top-k 検索手法を提案している。これらの手法では、返信されるデータ数は少ないものの、検索結果の取得に時間がかかるため、ネットワークポロジが動的に変化するアドホックネットワークでは有効ではない。また、複製を考慮せずに返信しているため、余分なトラヒックが発生してしまう。

アドホックネットワークの分野において、データの可用性を向上させることを目的とした複製配置に関する研究が盛んに行われている。特に、筆者の所属する研究グループでは、これまでに、文献 4) において、端末の記憶領域に制限のある環境を想定し、各端末が各データのアクセス頻度とネットワークポロジを考慮して周期的に複製を配置する手法を提案した。この研究では、データアクセスの際、まずクエリをネットワーク内にフラッディングし、アクセス対象のデータの所持端末を調べた後、クエリ発行端末から最も近い所持端末にアクセスすることを想定している。さらに、文献 8) において、モバイルアドホックネットワークにおける消費電力を考慮したデータ複製へのアクセス方式を提案した。この方式では、まずクエリのフラッディングによってデータ所有者を調べた後、経路長や経路上の端末の消費電力を考慮して、データの複製を返信する端末および経路を決定している。これらの研究では、自身または他の端末がもつ 1 つのデータに対してデータの識別子を指定してアクセス要求する環境を想定しており、1 回のアクセス要求で効率的に複数のデータを取得する必要がある本研究とは異なる。

3. 想定環境

本稿では、アドホックネットワークを構成する端末が、他の端末が保持するデータに対して、Top-k 検索を行う環境を想定する。Top-k 検索を行う端末は、検索条件を指定して検索クエリを発行し、ネットワーク内の端末の保持するデータの中から、上位 k 個以内のスコアをもつデータを取得する。データのスコアは、検索条件から何らかのスコアリング関数によって算出される。本研究で想定する Top-k 検索の手順は、スコアリング関数とは独立に動作するため、これを特に限定しない。

ネットワーク内には m 個の端末(識別子: M_1, M_2, \dots, M_m)が存在し、各々が自由に移動する。各端末は c 個のオリジナルデータを保持している。また、自身のオリジナルデータを格納する記憶領域以外に、データ h 個分の複製を配置する記憶領域を保持している。簡単化のため、すべてのデータや複製のサイズは等しいものとする。

時間の経過に伴ってデータやスコアリング関数が更新されることでスコアが変化する。この際、簡単化のため、すべてのデータおよびそのスコアはある一定の周期 T で変化するものとする。なお、各データが更新されるタイミングは一致するとは限らないものとする。

4. 先行研究

これまでに筆者らのグループでは、先行研究として、アドホックネットワーク上の Top-k 検索手法³⁾、および Top-k 検索のための複製配置手法²⁾を提案した。本章では、これらの先行研究について概説する。

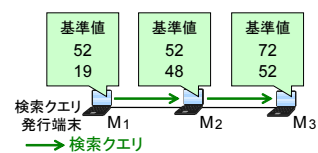
4.1 Top-k 検索手法

文献 3) の提案手法では、アドホックネットワーク上のトラヒックの削減と検索精度の維持を目的としている。この手法では、検索結果の候補となるデータを絞り込むため、検索クエリを発行した端末が N 個の基準値を添付して隣接端末に送信する。ここで、基準値は検索クエリの発行端末が保持するデータの上位 $\frac{k}{N}$ 個ごとのスコアに設定されるため、 i 個目の基準値は、それぞれ $\frac{k}{N}i$ ($1 \leq i \leq N$) 個のデータを取得できることを保証する。

次に、検索クエリを受信した各端末は、各自の保持するデータのスコアに基づいて基準値を逐次更新する。具体的には、基準値を、受信した検索クエリに添付された基準値と、自身が保持するデータの上位 $\frac{k}{N}$ 個ごとのスコア(計 $2N$ 個)の中で、上位 N 個以内のスコアに更新する。また、検索クエリにメッセージ送信端末とその親端末の識別子を添付することで、検索クエリを発行した端末を根とする木構造における自身の親と子、検索クエリの発行端末から自身までの経路、および親以外の隣接端末の検索クエリの発行端末からのホップ数を把握する。

検索クエリの転送時に更新した基準値を用いて、各端末は検索結果の候補となるデータを絞り込み、クエリ応答の返信では、この候補となったデータのみを自身の親へ返信する。具

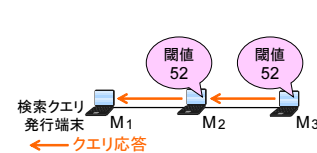
(1) 検索クエリの転送



各端末のもつデータのスコア

端末	M ₁	M ₂	M ₃
順位 1	85	77	92
2	70	67	84
3	61	56	80
4	52	48	72
5	44	46	60
6	39	40	59
7	25	32	37
8	19	28	30
...

(2) クエリ応答の返信



各端末の返信データ・検索結果

M ₁ の検索結果	M ₂ の返信データ	M ₃ の返信データ
92	92	92
85	84	84
84	80	80
80	77	72
77	72	60
72	67	59
70	60	
67	59	

図 1 Top-k 検索の処理例

体的には、 N 個目の基準値が k 位のスコアに対応しているため、このスコア以上のデータをクエリ応答として返信する。さらに、このスコアを閾値として、クエリ応答に付加する。クエリ応答を中継する端末が、受信した返信データと閾値から、自身のもつデータのうち、検索結果に入る可能性のあるもののみを返信し、それに伴い、クエリ応答に付加する閾値を更新する。また、端末がリンク切断を検出した場合、他の隣接端末を経由する返信経路を構築し、構築した経路に沿ってクエリ応答を送信する。

図 1 を用いて、文献 3) の Top-k 検索手法の処理例を説明する。この図は、 M_1 が 2 個の基準値 ($N = 2$) を用いて上位 8 個以内のデータを検索する例 ($k = 8$) を示している。図中の (1) における表は各端末の保持するデータのスコアを、(2) における表は各端末の返信データと、検索クエリの発行端末が取得した検索結果を示している。なお、(2) における表中の青字は、検索結果または返信データ内で、自身が保持していないデータのスコアを表している。また、(1) における四角吹き出しは各端末の基準値を、(2) における丸吹き出しは各端末の閾値を表している。例えば、 M_2 は、 M_1 から受信した検索クエリメッセージに含まれる基準値リスト 52、19 と自身の $\frac{k}{N}$ 個ごとのスコア 48、28 の中から上位 2 個のスコア 52、48 を M_2 の基準値とし、検索クエリの転送を続ける。検索クエリの転送後、子がない M_3 は、2 番目の基準値 52 を閾値とし、52 以上のスコアとそのデータを M_2 に返信する。 M_2 は自身の閾値を、 M_3 から受信したメッセージに含まれる閾値 52 と自身の基準値 48 の大きい方に更新し、受信した返信リストと自身の保持するデータの中で、スコアが閾

値 52 以上のデータとそのスコアを M_1 に返信する。

4.2 複製配置手法

文献 2) の複製配置手法では、文献 6) で提案されたパス複製法に従って、各端末が自身のクエリ応答を親に返信する際に、返信データの複製を配置する。このアプローチは、複製の配置によって追加のトラフィックが発生しないため、通信帯域が限られているアドホックネットワークにおいて有効であると考えられる。ここで、一般的に端末の記憶領域には制限があり、端末が作成できる複製の数は限られている場合が多い。このとき、すべての端末が単純にスコアの高いデータの複製を優先して配置すると、端末間で保持する複製の種類が類似してしまう。相互接続している端末間で複製の重複が多いと、これらの端末全体で保持できる複製が少なくなる。その結果、ネットワークが分断した場合に、検索結果に入るべきデータを取得できない可能性が高くなる。そこで、文献 2) の手法では、高順位のデータを優先しつつ、配置するデータの種類の多様化するように配置する。具体的には、各データのスコアと乱数からなる評価値を定義し、その評価値が大きいデータの複製を配置する。

また、データの更新によるスコアの変化が発生する環境では、データの有効期限を考慮せずに複製を配置すると、多くの端末が有効期限までの残り時間の短い複製を配置してしまう可能性がある。その結果、多くの複製が短時間のうちに無効になってしまうため、複製配置の効果が小さくなる。そこで、文献 2) の手法では、データの更新によるスコアの変化が発生する環境では、スコアが高く有効期間が長いデータの複製を優先的に配置する。

5. 提案手法

本稿で提案する Top-k 検索手法は、基本的には文献 3) と同様の手順により Top-k 検索を実行する。しかし、文献 3) の手法では、複製を考慮せずに基準値の決定や返信を行っているため、1 章で述べたように、同じデータが複数回返信されたり、クエリ発行端末から遠い端末によるデータ返信が行われたりすることで、大きなトラフィックが発生してしまう。そこで提案手法では、文献 3) の手法を拡張して、複製の存在を考慮して基準値および返信データを決定する。具体的には、各端末が自身のもつデータのうち上位 k 個に入る可能性のあるデータの識別子をクエリメッセージに添付する。これにより、クエリの伝搬経路上で、自身よりクエリ発行端末に近い端末が保持するデータを把握することができる。データの返信時に各端末は自身よりクエリ発行端末に近い端末が保持しているデータを返信しない。また、異なる経路上を伝搬したクエリメッセージを受信したとき、その経路上に同じデータを保持する端末が存在した場合は、その端末の方がクエリ発行端末に近ければ、そのデータを返信しない。提案手法の詳細を以下に示す。

5.1 基準値の設定

提案手法では、文献 3) の手法と同様に、 N 個のスコアを基準値として設定する。ここで、端末 M_i における N 個の基準値 $B(i, j) (1 \leq j \leq N)$ はそれぞれ、検索クエリを発行した端

末が、 M_i までの経路上の端末から $\frac{k}{N}j$ ($1 \leq j \leq N$) 個のデータ取得できることを保証するスコアとなる。

各端末 M_i は、自身が保持するデータ（複製）に対して、基準値を以下のように設定する。

$$S(i, \frac{k}{N}j) \quad (1 \leq j \leq N) \quad (1)$$

ここで、 $S(i, j)$ は、端末 M_i が、自身が保持する全データ（複製）のスコアを降順に並べたときの j 番目のスコアを示す。さらに、各端末は、受信した検索クエリメッセージに添付された基準値と自身がもつデータの基準値のうち、上位 N 個を自身の基準値 $B(i, j)$ とする。なお、クエリ発行端末の基準値（初期値）は、その端末がもつデータの上位 $\frac{k}{N}$ 個ごとのスコアとなる。

5.2 検索クエリの転送

提案手法では、検索結果の候補となるデータを絞り込むため、各端末が自身の基準値を検索クエリに添付して隣接端末に送信する。また、これを受信した端末が基準値を更新することで、絞込みの効率を向上させる。以下では、端末 M_p が Top-k 検索クエリを発行する際の各端末の動作について説明する。

- (1) クエリ発行端末 M_p は検索条件、および要求データ数 k を指定する。さらに、5.1 節の方法に従って、基準値の初期値を設定する。
- (2) 端末 M_p は、隣接する端末に検索クエリメッセージを送信する。このメッセージには、検索クエリ発行端末 M_p の識別子、要求データ数 k 、検索条件、端末 M_p の基準値リスト $B(p, j)$ ($1 \leq j \leq N$)、 N 番目の基準値 $B(p, N)$ 以上のスコアをもつデータの識別子リスト、ホップ数が含まれる。ホップ数は、クエリ発行端末までのホップ数を示し、この場合 0 となる。また、クエリ応答の返信先を把握するため、メッセージ送信端末とその親端末の識別子が含まれる。この場合、メッセージ送信端末は M_p となり、親端末には何も設定されない。
- (3) 検索クエリメッセージを初めて受信した端末 M_q は、メッセージ送信端末を自身の親とし、ホップ数とメッセージ送信端末の組、およびデータの識別子リストを記録する。受信したデータの識別子リストを調べ、自身が同じ識別子のデータを所持している場合、そのデータは基準値の計算には用いない（所持していないものとして扱う）。端末 M_q は、手順 (1) と同様に、自身のもつデータ、および受信したクエリメッセージに含まれる基準値リストから、5.1 節の方法に従って、自身の基準値 $B(q, j)$ ($1 \leq j \leq N$) を決定する。
- (4) 端末 M_q は、メッセージ送信端末と親端末の識別子をそれぞれ M_q と自身の親端末の識別子とし、基準値リストを自身の基準値 $B(q, j)$ ($1 \leq j \leq N$) とする。データの識別子リストは、受信したクエリメッセージ中のデータの識別子リストに対して、自身が所持し、かつ、受信したリストに含まれないデータのうち、 N 番目の基準値 $B(q, N)$ 以上のスコアをもつデータ識別子を追加したものとする。また、ホップ数を

- 1 増やし、これらの情報を含めた検索クエリメッセージを自身の隣接端末に送信する。
- (5) 同じ検索クエリメッセージを再び受信した端末 M_r は、メッセージ送信端末の親端末が自身の場合、メッセージ送信端末を自身の子とする。一方、メッセージ送信端末が自身の子でない場合、手順 (3) と同様に、メッセージに含まれるホップ数とメッセージ送信端末の組を記録する。受信した端末と自身の端末のクエリ発行端末からのホップ数を比較して、自身の方が大きい場合、または、受信した端末と自身の端末のクエリ発行端末からのホップ数が同じ、かつ、送信端末の受信した端末の識別子を比較して自身の方が小さい場合、データの識別子リストを記録する。さらに、受信したデータの識別子リストを調べ、自身が同じ識別子を所持している場合、そのデータを所持していないものとして、5.1 節に従って、自身の基準値 $B(r, j)$ ($1 \leq j \leq N$) を再計算する。これは、複数の経路上から同じデータが返信されるのを防ぐためである。まず、ホップ数を比較することにより、クエリ発行端末により近い端末が、データを返信することを可能としている。ここで、送信端末と受信端末は基本的に通信範囲内にいるので、互いが送信したメッセージを相手を受信できていることを把握できる。そのため、上記の手順により、両方がデータ返信を取りやめることはなく、一方（より遠い端末）のみが返信を取りやめることを可能とする。

検索クエリの転送において、基準値を更新することで検索結果の候補となるデータの絞込みができる。さらに、クエリの伝搬経路上において、自身よりクエリ発行端末に近い端末のデータの保持状況を把握できるため、無駄なデータ返信を防ぐことができる。また、各端末は、検索クエリメッセージに含まれるメッセージ送信端末とその親端末の識別子により、検索クエリ発行端末を根とする木構造における自身の親と子把握できる。ホップ数とメッセージ送信端末の組から、親子関係ではない隣接端末、自身、および検索クエリ発行端末の位置関係を把握できる。

5.3 クエリ応答の返信

5.2 節の手順により、各端末は検索結果の候補となるデータを絞り込めるため、クエリ応答では、この候補となったデータのみを返信する。また、クエリ応答を中継する各端末は、受信した情報から、自身の返信予定のデータの中で返信の必要がないデータを検出した場合、その返信を行わない。以下では、5.2 節の手順の後、各端末がクエリ応答を返信する動作について説明する。

- (1) 自身の子がいない端末 M_s は、クエリ応答メッセージを自身の親に返信する。このメッセージには、検索クエリ発行端末 M_p 、閾値、および返信リストが含まれる。閾値は推測した k 位のスコア $B(s, N)$ である。返信リストには、自身が所持し、かつ、受信した識別リストに含まれないデータのうち、閾値以上のスコアをもつものが含まれる。
- (2) クエリ応答メッセージを受信した端末 M_t は、自身の閾値を受信したメッセージに含まれる閾値と $B(t, N)$ の大きい方に更新する。そして、メッセージに含まれる返信リ

ストから閾値未満のデータを削除する。また、自身が所持し、かつ、クエリ転送時に受信したデータの識別子リスト、および、返信リストに含まれないデータのうち、閾値以上のスコアをもつものを返信リストに追加する。返信リストに含まれるデータ数が k より大きい場合、閾値を返信リストに含まれるデータの k 番目のスコアに更新し、閾値未満のデータを返信リストから削除する。

すべての子からのクエリ応答を受信した端末は、自身の親に更新したクエリ応答メッセージを送信する。

クエリ応答の返信では、閾値を用いて、検索結果に入らないと推測されたデータを返信しないことで、トラフィックを削減する。また、データの識別子リストを参照し、自身よりクエリ発行端末に近い端末がデータを返信したり、他の経路上の端末が返信するデータを返信しないことで、さらにトラフィックを削減する。

6. 評価結果

本章では、提案手法の性能評価のために行ったシミュレーション実験の結果を示す。本実験では、ネットワークシミュレータ Qualnet4.0⁷⁾を用いた。

6.1 シミュレーション環境

600[m]×600[m]の2次元平面上の領域に、50台の端末(識別子: M_1, M_2, \dots, M_{50})が存在する。各端末はランダムウェイポイント¹⁾に従い、5[m/秒]の一定の速度で移動する。また、各端末はIEEE 802.11bを使用し、伝送速度11[Mbps]、通信伝播距離が100[m]程度となる送信電力でデータを送信する。各端末は1[KB]のサイズのデータ(オリジナル)を100個保持するものとし、各端末のもつデータのスコアは0から100の範囲の一様乱数に従う。検索クエリメッセージは、クエリ発行端末、送信端末、要求データ数 k 、検索条件、基準値、データの識別子リストから構成されている。各要素は、基本的に1[B]であり、基準値は、識別子、スコア、生存時間から構成されているため、3[B]である。したがって、検索クエリメッセージは、提案手法では、(4 + 識別子リストの要素数 + 基準値数 × 3) [B]となり、文献3)の手法では、(4 + 基準値数 × 3) [B]となる。

複製の配置は、文献2)の手法に基づいて決定する。ここで、文献2)の複製配置手法では、データ更新の有無により、複製の配置方法が異なる。データ更新がある場合、更新間隔も考慮に入れて配置する複製を決定する。6.2節では、データ更新がないものとし、6.3節では、各データの更新間隔は2000秒とし、各データが最初に更新されるタイミングはランダムに決定した。

本実験では、各端末が所持できる複製数を0から100の間で変化させた。以上のシミュレーション環境において、各端末は、端末 M_1 から端末 M_{50} まで、順番に1回ずつ20個の基準値を用いて、上位100個 ($k = 100$) 以内のデータを要求する検索クエリを60[秒]間隔で発行する。Top-k 検索手法には、提案手法と文献3)の手法を用い、以下の二つの性能

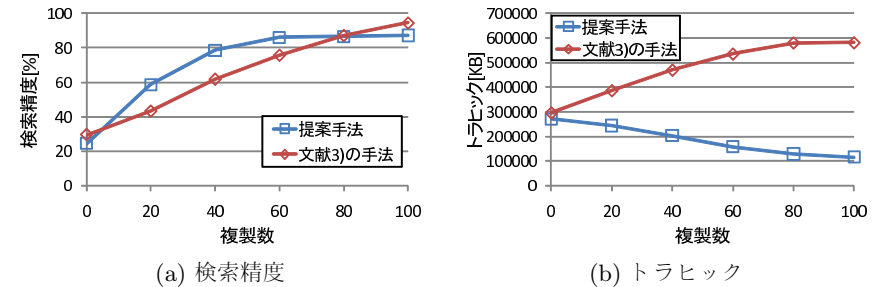


図2 データ更新が発生しない場合における複製数の影響

指標について評価した。

● 検索精度

ネットワーク全体の上位 k 個以内のスコアをもつデータのうち、検索クエリの発行端末が取得したデータ数の割合の全試行における平均。

● トラフィック

1つのクエリに対して、ネットワーク全体で送信されたメッセージ、およびデータの合計サイズの全試行における平均。

6.2 データ更新が発生しない場合における複製数の影響

データ更新が発生しない環境において複製数を変化させたときの評価結果を図2に示す。この図において、横軸は複製数、縦軸は図2(a)では検索結果の検索精度、図2(b)ではトラフィックをそれぞれ表す。

図2(a)の結果から、複製数が増加すると両方の手法で検索精度が向上する。これは、複製数が増加すると上位 k 個に入る可能性のあるデータを各端末が多く保持できるためである。二つの手法を比較すると、提案手法の方がほとんどの複製数の場合で文献3)の手法より高い検索精度を示している。これは、提案手法の方が余分なトラフィックを抑えられることから、パケットロスの影響が小さくなるためである。また、図2(b)の結果から、提案手法では、複製数が多くなるにつれて、トラフィックが下がっていることがわかる。提案手法では、各端末がクエリ中継時にデータ保持情報を添付するため、自身よりクエリ発行端末に近い端末が保持しているデータを把握することができる。これにより、同じデータを複数の端末が保持している場合、クエリ発行端末により近い端末のみがデータを返信できる。複製数が多くなると、よりクエリ発行端末に近い端末のみがデータを返信できるようになるため、トラフィックが小さくなる。文献3)の手法では、他の端末のデータの保持を考慮せずに、データを返信しているため、複製数が多くなるにつれて、同じデータが複数回返信され、トラフィックが大きくなっている。一方、複製数が100の場合、文献3)の手法の方が検索精度

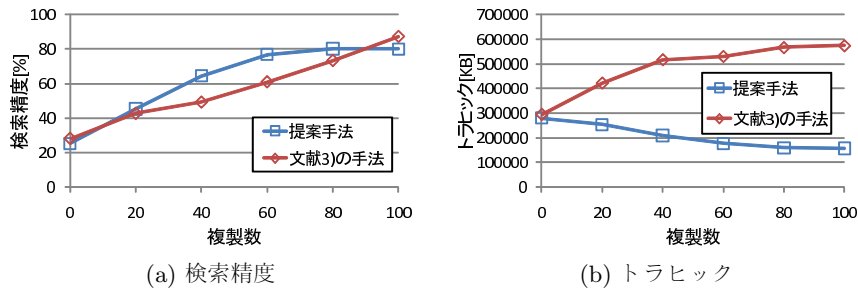


図 3 データ更新が発生する場合における複製数の影響

が高い。これは、文献 3) の手法を用いた場合、データが長い経路上や、複数の経路上を返信されることが多いため、提案手法と比較して、中継端末が複製を作成する機会が多くなるからである。その結果、各端末が上位 100 個のスコアをもつほぼすべてのデータの複製で保持できるため、検索精度が高くなっている。

上述のように、提案手法では、文献 3) の手法に比べて、複製を作成する機会が少ない。そのため、各端末が所持できる複製数が多い場合でも、複製の所持数が上限に達することのない端末が多く存在する。具体的には、各端末が 100 個の複製を所持できる場合でも、シミュレーション終了時で各端末は、平均で 80 個程度の複製しか保持できていなかった。提案手法の検索精度の最大値が 85% 程度となっているのは、ネットワークが分断して、少数の端末からしかデータ返信がない場合、この少ない複製数が大きく影響したものである。

6.3 データ更新が発生する場合における複製数の影響

データ更新が発生する環境において複製数を変化させたときの評価結果を図 3 に示す。この図において、横軸は複製数、縦軸は図 3 (a) では検索結果の検索精度、図 3 (b) ではトラフィックをそれぞれ表す。

図 3 (a) の結果から、6.2 節と同様に複製数が増加すると両方の手法で検索精度が向上する。データ更新が発生する場合、時間経過に伴い、複製が無効になるため、データ更新が発生しない場合よりも全体的に検索精度が低い。特に、文献 3) の手法において、複製数が 100 の場合で顕著に表れている。これは、配置できる複製数が多い場合、短時間で無効になる複製も多く配置してしまうためである。図 3 (b) の結果から、提案手法は、データ更新がある場合でもトラフィックを小さく抑えることができている。以上の結果より、提案手法は、データ更新がある環境でも有効であることがわかる。

7. 結 論

本稿では、アドホックネットワークにおいて、配置されている複製を考慮した Top-k 検

索手法を提案した。提案手法では、検索クエリメッセージに各端末のデータ保持情報を添付することにより、同じデータが複数回返信されたり、クエリ発行端末から遠い端末によりデータが返信されることを回避することにより、トラフィックを削減できる。

シミュレーション実験の結果から、提案手法により同じデータが複数回返信されることや長い経路上を返信されることを防げるため、トラフィックが削減するとともに、検索精度が向上することを確認した。

複製を多く配置した場合、全ての端末にクエリを伝搬させなくても上位 k 個のデータを取得できる可能性が高くなる。そこで、今後は、検索クエリを中継する端末が、クエリに添付された経路上の端末のデータ保持状況の情報や基準値から、クエリの伝搬を継続するかを判断することを考える。上位 k 個の全てのデータを既に収集可能であると判断した場合、クエリの伝搬を停止することにより、さらなるトラフィックの削減が可能であると考えられる。

参 考 文 献

- 1) T.Camp, J.Boleng, and V.Davies, "A Survey of Mobility Models for Ad Hoc Network Research," *Wireless Communications and Mobile Computing*, Vol.2, No.5, pp.483–502, 2002.
- 2) 萩原亮, 原隆浩, 西尾章治郎, "アドホックネットワークにおける Top-k 検索のための複製配置手法," 第 2 回データ工学と情報マネジメントに関するフォーラム (DEIM2010), Feb. 2010.
- 3) 萩原 亮, 篠原昌子, 原 隆浩, 西尾章治郎, "アドホックネットワークにおける Top-k 検索のためのメッセージ処理手法," *情報処理学会論文誌*, Vol.51, No.1, pp.152–162, 2010.
- 4) 原 隆浩, "アドホックネットワークにおけるデータ利用性向上のための複製配置," *電子情報通信学会論文誌 B*, Vol.J84-B, No.3, pp.632–642, 2001.
- 5) P.Kalnis, W.S.Ng, B.C.Ooi, and K.-L.Tan, "Answering similarity queries in peer-to-peer networks," *Information Systems*, vol.31, no.1, pp.57–72, March 2006.
- 6) Q.Lv, P.Cao, E.Cohen, K.Li, and S.Shenker, "Search and Replication in Unstructured Peer-to-Peer Networks," in *Proc. of Int'l Conf. on Supercomputing (ICS'02)*, pp.53–65 2002.
- 7) Scalable Network Technologies: "QualNet," <http://www.scalable-networks.com>.
- 8) 篠原昌子, 原 隆浩, 西尾章治郎, "モバイルアドホックネットワークにおける消費電力を考慮したデータ複製のアクセス方式," *情報処理学会論文誌*, Vol.48, No.12, pp.3931–3948, 2007.
- 9) W.-T.Balke, W.Nejdl, W.Siberski, and U.Thaden, "Progressive distributed top-k retrieval in peer-to-peer networks," *Proc. Int. Conf. on Data Engineering*, pp.174–185, Apr. 2005.