

## アドホックネットワークにおける ヒストグラムを用いた Top-k 検索手法

佐々木 勇和<sup>†1</sup> 萩原 亮<sup>†1</sup> 原 隆浩<sup>†1</sup>  
篠原 昌子<sup>†1</sup> 西尾 章治郎<sup>†1</sup>

アドホックネットワークでは、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ため、端末が何らかの値（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。本稿では、アドホックネットワークにおいて、トラフィックの削減と検索結果の取得精度の低下を抑止を目的する Top-k 検索手法を提案する。提案手法では、各端末がヒストグラムを作成し、ネットワーク全体の  $k$  番目のスコアを推定する。次に、推定した  $k$  番目のスコア以上のスコアをもつデータを返信することで、できる限り検索結果に入るデータのみを返信する。また、返信先の端末とのリンク切断を検出した端末は、他の隣接端末にクエリ応答を返信することで、検索結果の取得精度の低下を抑止する。

### A Top-k Query Method using Histogram in Mobile Ad Hoc Networks

YUYA SASAKI<sup>,†1</sup> RYO HAGIHARA<sup>,†1</sup>  
TAKAHIRO HARA<sup>,†1</sup> MASAKO SHINOHARA<sup>†1</sup>  
and SHOJIRO NISHIO<sup>†1</sup>

In mobile ad hoc networks, 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 mobile node acquires data items with the  $k$  highest scores. In this paper, we propose a query processing method for top-k query for reducing traffic and also keeping high accuracy of the query result. In this method, each node constructs a histogram and estimates the  $k$ -th score in the entire network. When transmitting a reply, each mobile node sends back only data items whose scores are larger than the estimated  $k$ -th score to reduce traffic as much as possible. Moreover, when a mobile node detects the disconnection of a radio link, it sends the reply to another neighboring node to keep high accuracy of the query result.

### 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、複数の端末が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、Top-k 検索を実現する単純な方法として、端末が検索クエリをネットワーク全体にフラッディングし、これを受信した端末が自身のもつデータの中からスコアの高いものを固定数返信する方法が考えられる。各端末の返信するデータの数が多の場合、検索クエリを発行した端末は、ネットワーク全体の上位  $k$  個のデータ（検索結果）を取得できる可能性が高い。しかし、検索結果に入らないデータまで返信されるため、不要なトラフィックが発生する。一方、各端末の返信データ数が少ない場合、検索結果に入らないデータが返信される可能性は低くなるが、検索結果に入るデータが返信されず、検索結果の取得精度が低下する。また、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トポロジが動的に変化する。Top-k 検索の実行中に端末間のリンクが切断された場合、検索クエリを発行した端末が検索結果に含まれるデータを取得できず、検索結果の精度が低下する可能性がある。例えば図 1 において、左端の看護師が血圧の高い 3 人の被災者を検索する場合、各看護師が自身の管理情報から血圧の高い 3 人の被災者の情報を返信すると、必要以上の被災者の情報が返信されてしまう。一方、各看護師が最も血圧の高い被災者の情報のみを返信した場合、血圧が 3 番目に高い被災者 H の情報が返信されない。また、右下と中央の看護師のもつ端末間のリンクが切断された場合、最も血圧の高い被災者 J の情報が返信されない。

そこで本稿では、アドホックネットワークにおいて、トラフィックの削減と検索結果の取得精度の低下を抑止を目的とする Top-k 検索手法を提案する。提案手法では、各端末が自身のもつデータのスコアからヒストグラムを作成し、ネットワーク全体の  $k$  番目のスコアを推定する。このとき、 $k$  番目のスコアを正確に推定するために、各端末は検索クエリやクエリ応答にヒストグラムを添付して送信する。次に、各端末は推定した  $k$  番目のスコア以上のスコアをもつデータのみを返信することで、できる限り検索結果に入るデータのみを返信する。また、返信先の端末とのリンクが切断した端末は、他の隣接端末にクエリ応答を返信

<sup>†1</sup> 大阪大学  
Osaka University

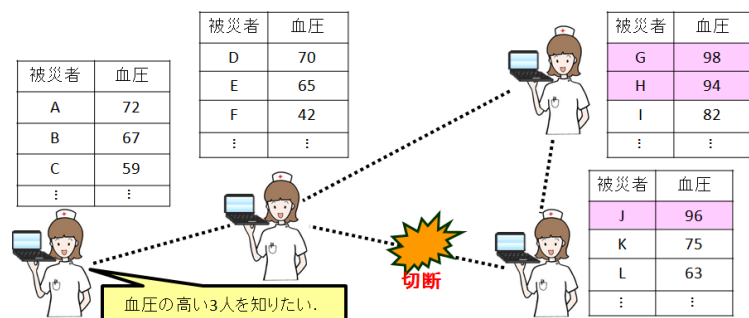


図 1 アドホックネットワークにおける Top-k 検索例

することで、検索結果の取得精度の低下を抑止する。

以下では、2章で関連研究を紹介する。3章で提案手法について述べ、4章でシミュレーション実験の結果を示す。最後に5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

非構造検索トポロジを用いた P2P ネットワークの分野において、Top-k 検索に関する研究が盛んに行われている。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、これを受信した各端末が自身のデータを順位付けし、上位  $k$  個のスコアをもつデータを返信する手法を提案している。この手法では、検索結果に入らない不要データの返信によるトラフィックが多く、1章で述べた問題を解決できない。また文献 2) 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。この手法では、返信されるデータは少ないものの、検索結果の取得に時間がかかるため、ネットワークト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。文献 1) では、検索クエリとクエリ応答を適切な隣接端末に送信することでトラフィックを削減する手法を提案している。またこの手法では、クエリ応答の返信時に隣接端末が存在しない端末は、検索クエリを発行した端末にクエリ応答を直接返信する。この研究では、任意の端末間で直接通信が可能なネットワークを想定している点で本研究と異なる。

これまでに筆者の研究グループは、文献 4) において、アドホックネットワークにおけるトラフィックの削減と検索結果の取得精度の維持を実現する Top-k 検索手法を提案した。この手法では、各端末が検索結果に含まれるデータスコアを推測し、その一部を基準値として用いることで、検索クエリの転送時およびクエリ応答の返信時に検索結果の候補となるデー

タを絞り込む。この手法では、上位  $k$  個に入らないことが確実なデータの返信を防ぐことはできるが、検索結果の絞り込みの効果はそれほど大きくない。一方、本稿の提案手法は、中継端末のもつデータのスコア分布からネットワーク全体の  $k$  番目のスコアを推定するため、絞り込みの効果が大きい。しかし、推定の精度が低い場合、必ずしも上位  $k$  個のデータを取得できない可能性がある。

## 3. Top-k 検索手法

本章では、まず想定環境について述べ、各端末がヒストグラムを作成する手順について述べる。その後、本稿で提案する Top-k 検索手法について説明する。

### 3.1 想定環境

本稿では、アドホックネットワークを構成する各端末が、自身と他の端末のもつデータに対して 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、ネットワーク内の上位  $k$  個のスコアをもつデータを取得する。

ネットワーク内には  $n$  個のデータが存在し、各々が特定の端末に保持されている。簡単化のため、全てのデータのサイズは等しく、複製は作成しないものとする。データのスコアは、検索条件とデータの属性値から決定し、何らかのスコアリング関数を用いて算出されるものとする。

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、ネットワーク全体のデータ数  $n$ 、スコアのとりうる値の範囲を把握しているものとする。また、各端末のもつデータ数はネットワーク全体のデータ数 ( $n$ ) に比べて少ないため、各端末のもつデータのスコア分布は、ネットワーク全体のデータのスコア分布に従っているとは限らない。

### 3.2 ヒストグラムの作成

各端末は自身のもつデータのスコアを用いて、ヒストグラムを作成する。ヒストグラムは、階級と呼ばれる適当な大きさの区間ごとに、値がその区間に含まれるスコアの数数を数えたものであり、スコアの分布状況を表すのに用いられる。本稿では、端末  $M_i$  のもつデータの中で、そのスコアが階級  $c_j$  ( $1 \leq j \leq C$ ) に含まれる個数を数えたものを  $M_i$  のヒストグラム  $H_i$  とする。ただし、階級  $c_j$  は、ヒストグラムの幅の最大値と最小値の範囲  $[MIN, MAX)$  を、大きさの等しい  $C$  個の階級に分割したときの  $j$  番目の階級を示し、その範囲は  $[MIN + \frac{(j-1)(MAX-MIN)}{C}, MIN + \frac{j(MAX-MIN)}{C})$  となる。例えば図 2 は、ヒストグラムの幅の範囲が  $[0, 100)$ 、階級数  $C$  が 10 の場合に、ネットワーク内の端末  $M_1$  から  $M_5$  が作成したヒストグラムおよびネットワーク全体のヒストグラムを示す。

### 3.3 Top-k 検索手法

#### 3.3.1 検索クエリの転送

提案手法では、ヒストグラムに含まれるスコアの数が多いほど、そのスコア分布はネット

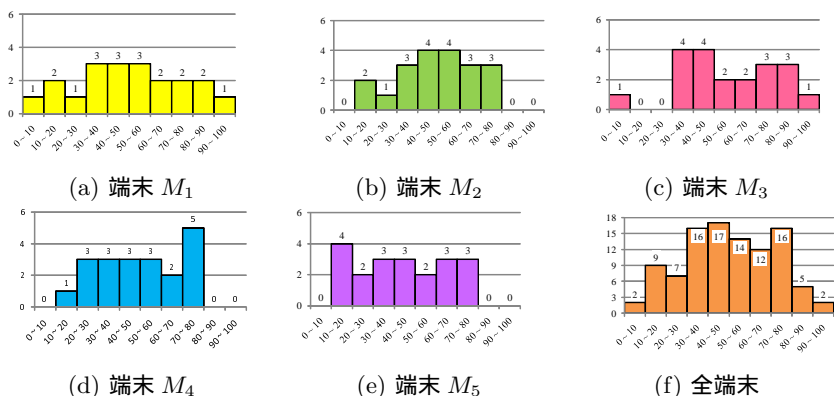


図 2 各端末の作成したヒストグラムとネットワーク全体のヒストグラム

ワーク全体のスコア分布に近づくため、端末は  $k$  番目のスコアをより正確に推定できる。そこで各端末は、ヒストグラムを検索クエリに添付して送信し、検索クエリを中継する端末がヒストグラムを更新しながら転送する。以下では、検索クエリを発行した端末  $M_p$  と検索クエリを受信した端末  $M_q$  の動作について説明する。

- (1) 端末  $M_p$  は検索条件、および要求データ数  $k$  を指定する。また、自身のもつデータのスコアを算出し、3.2 節の方法に従って、自身のヒストグラム  $H_p$  を作成する。
- (2)  $M_p$  は自身の隣接端末に検索クエリを送信する。この検索クエリには、クエリ発行端末  $M_p$  の識別子、検索クエリの識別子、要求データ数  $k$ 、検索条件、経路端末リスト、およびクエリヒストグラムが含まれる。経路端末リストにはクエリ発行端末から自身までの経路上に存在する端末の識別子が含まれ、ここでは  $M_p$  のみとなる。クエリヒストグラムはこれらの経路上の端末のヒストグラムを統合したものであり、ここでは  $H_p$  となる。
- (3) 検索クエリを受信した端末  $M_q$  は、それが初めて受信したものであれば、経路端末リストの末尾に格納されている端末を自身の親とし、経路端末リストに含まれる端末数から、クエリ発行端末から親までのホップ数を調べる。手順 (4) へ進む。  
検索クエリが既に受信したものであれば、手順 (5) へ進む。
- (4)  $M_q$  は、手順 (1) と同様に検索条件から自身のヒストグラム  $H_q$  を作成し、検索クエリに含まれるクエリヒストグラムに  $H_q$  を統合する。また、自身の識別子  $M_q$  を経路端末リストの末尾に追加する。 $M_q$  は、自身の隣接端末に検索クエリを送信し、手順 (3) に戻る。
- (5) 検索クエリを再受信した端末  $M_q$  は、検索クエリに含まれる経路端末リストの末尾

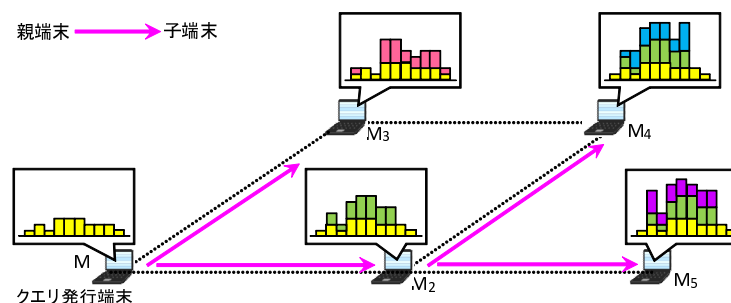


図 3 検索クエリの転送例

の端末を親でない隣接端末とし、経路端末リストに含まれる端末数から、その端末までのホップ数を記録する。また、経路端末リストの末尾から 2 番目の端末が自身の場合、経路端末リストの末尾の端末を自身の子とする。

検索クエリの転送では、各端末がクエリ発行端末から自身までの経路上に存在する端末のヒストグラムを統合したクエリヒストグラムを検索クエリに添付することで、ヒストグラムに含まれるスコア数を増加させる。また、各端末は経路端末リストにより、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、クエリ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クエリ発行端末からのホップ数を把握できる。

図 3 を用いて、 $M_1$  が Top- $k$  検索を行った場合の検索クエリの転送例を説明する。吹出しは各端末が検索クエリに添付したクエリヒストグラムを示し、破線の矢印は端末間の親子関係を示す。検索クエリには、クエリ発行端末からの経路上の端末のヒストグラムを統合したものが添付されているが、統合の状態をわかりやすく示すため、端末  $M_1$ ,  $M_2$ ,  $M_3$ ,  $M_4$ , および  $M_5$  のヒストグラムに相当する部分をそれぞれ黄色、緑色、桃色、青色、および紫色で表す。なお、各端末のヒストグラムは図 2 で表されるものとする。

### 3.3.2 クエリ応答の返信

3.3.1 項で述べたように、提案手法では、ヒストグラムに含まれるスコア数が多いほど、 $k$  番目のスコアをより正確に推定できるため、各端末はヒストグラムをクエリ応答にも添付して送信する。また、クエリ応答を中継する端末は、受信した情報から返信の必要がないと判断されるデータは返信しない。以下では、3.3.1 項において、端末  $M_p$  が検索クエリを転送した後、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) 自身の子がない端末  $M_r$  は、送信した検索クエリに添付したクエリヒストグラムを用いて、 $k$  番目のスコアの推定 (3.3.4 項) を行う。その後  $M_r$  は、クエリ応答を自身の親に送信する。このクエリ応答には、クエリ発行端末  $M_p$  の識別子、検索クエリの識別子、返信リスト、および応答ヒストグラムが含まれる。ここで返信リストに

は、 $M_r$  のもつデータの中で推定した  $k$  番目のスコア以上のスコアをもつデータとそのスコアが含まれる。また、応答ヒストグラムは、自身および  $M_r$  の全ての子孫端末のヒストグラムを統合したものが含まれ、ここでは  $H_r$  となる。

- (2) 全ての子からクエリ応答を受信するか、自身の検索クエリを送信してから一定時間経過した端末  $M_s$  は、受信した全てのクエリ応答に含まれる応答ヒストグラムと、自身が検索クエリに添付したクエリヒストグラムを統合する。このヒストグラムは、クエリ発行端末から  $M_s$  までの経路上の端末、および  $M_s$  の全ての子孫端末のヒストグラムを統合したものととなる。 $M_s$  は、このヒストグラム  $H'$  を用いて、 $k$  番目のスコアの推定 (3.3.4 項) を行う。
- (3)  $M_s$  は、クエリ応答を作成し、自身の親に送信する。ここで、返信リストは、受信した全てのクエリ応答に含まれる返信リストおよび  $M_s$  のもつデータの中で、推定した  $k$  番目のスコア以上のスコアをもつデータとそのスコアが含まれる。ただし、該当するデータの数が要求データ数  $k$  より大きい場合、上位  $k$  個のスコアとそのスコアをもつデータのみが含まれる。また、応答ヒストグラムは、ヒストグラム  $H_s$  と受信した全ての応答ヒストグラムを統合したものととなる。

クエリ応答の返信では、推定した  $k$  番目のスコア以上のスコアをもつデータのみを返信する。また、中継端末で検索結果に入らないと判断できたデータを返信しない。さらに、自身および自身の子孫端末のヒストグラムを統合した応答ヒストグラムをクエリ応答に添付することで、クエリ発行端末に近い端末ほど、ヒストグラムに含まれるスコア数が増加し、 $k$  番目のスコアをより正確に推定できる。その結果、検索結果に入らないデータの返信を抑制できるため、取得精度を低下させずにトラフィックを削減できる。

図 3 および図 4 を用いて、クエリ応答の返信例を説明する。図 4 において、吹出しは各端末がクエリ応答に添付した応答ヒストグラムを示し、矢印はクエリ応答の流れを示す。例えば、 $M_4$  は、図 3 に示すクエリヒストグラム ( $M_1$ [黄色],  $M_2$ [緑色],  $M_4$ [青色] のヒストグラムを統合したもの) を用いて、 $k$  番目のスコアを推定し、応答ヒストグラム ( $M_4$ [青色] のヒストグラム) と推定した  $k$  番目のスコア以上のスコアをもつデータを、親である  $M_2$  に返信する。

### 3.3.3 リンク切断の検出時の処理

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する。ここで、親とのリンクが切断された端末は、クエリ応答を返信できないため、検索結果の取得精度が低下する。そこで、親とのリンクが切断した端末は、他の隣接端末にクエリ応答を送信する。以下では、親とのリンクが切断した端末  $M_t$  が、他の隣接端末にクエリ応答を返信する動作について説明する。

- (1) 端末  $M_t$  は、親以外の隣接端末にクエリ応答を転送する。ここで、複数の端末と隣接している場合、隣接端末のうちクエリ発行端末までのホップ数が最小のものに転送す

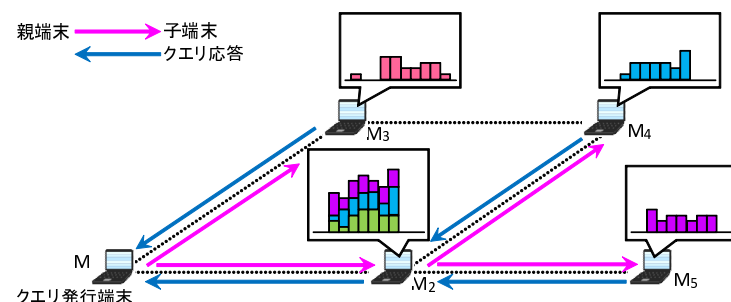


図 4 クエリ応答の返信例

る。どの端末とも隣接していない場合、クエリ応答を破棄する。

- (2) 自身の子以外の端末からクエリ応答を受信した端末  $M_u$  は、端末  $M_t$  がクエリ発行端末から自身までの経路上に存在しない場合、受信したクエリ応答を子からのクエリ応答と同様に扱い、3.3.2 項の手順に従って親に送信する。ただし、データの重複を防ぐため、既に返信したデータとそのスコアは返信リストから削除する。一方、端末  $M_t$  がクエリ発行端末から自身までの経路上に存在する場合 (端末  $M_t$  が端末  $M_u$  の親端末である場合も含む)、 $M_u$  が親に送信したクエリ応答は端末  $M_t$  まで返信されてしまう。そこでこの場合、端末  $M_u$  は受信したクエリ応答を親以外のいずれかの隣接端末に送信する。親以外の端末と隣接していない場合、クエリ応答を破棄する。

親とのリンクが切断した場合、端末が親以外の隣接端末にクエリ応答を返信することで、検索結果の取得精度の低下をできる限り抑制できる。図 4 を用いて、リンク切断時の処理例を説明する。例えば、 $M_1 - M_2$  間のリンクが切断した場合、端末  $M_2$  は  $M_4$  にクエリ応答を送信する。これを受信した  $M_4$  は、 $M_2$  がクエリ発行端末から自身までの経路上に存在するため、 $M_3$  にこのクエリ応答を送信する。

### 3.3.4 $k$ 番目のスコアの推定

提案手法では、各端末はクエリ応答を返信する際、クエリ発行端末から自身までの経路上の端末、および自身の子孫端末のヒストグラムを統合した推定用ヒストグラム  $H'$  を用いて、ネットワーク内の  $k$  番目のスコアを推定する。このとき、 $H'$  がネットワーク全体のデータのスコア分布と等しいと考える。具体的には、以下の式を満たす最大の自然数  $a$  ( $\leq C$ ) を求める。

$$\sum_{l=a}^C \frac{n}{n_{H'}} \cdot e_l \geq k, \quad (1)$$



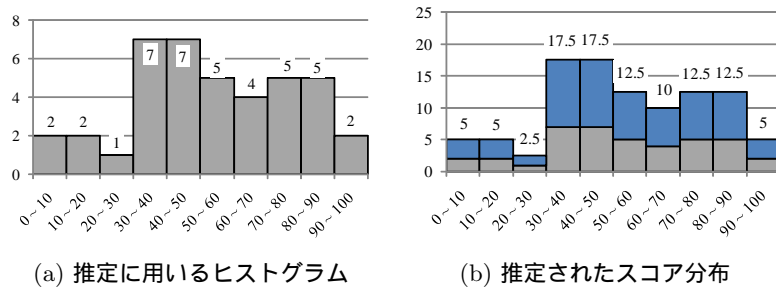


図5 スコア分布の推定

$$n_{H'} = \sum_{l=1}^C e_l. \quad (2)$$

ここで、 $n$  はネットワーク全体のデータ数、 $k$  は要求データ数、 $C$  は推定用ヒストグラム  $H'$  の階級数、および  $e_j$  ( $1 \leq j \leq C$ ) は階級  $c_j$  に含まれるスコアの数を示す。つまり、 $n_{H'}$  は推定用ヒストグラムに含まれるスコア数を示し、式 (1) は推定用ヒストグラム内のスコア数とネットワーク全体のデータ数の比に基づいて、推定用ヒストグラムの各階級に含まれるスコア数を増加したヒストグラムを、ネットワーク全体のデータのスコア分布として推定している。このとき  $a$  は、階級  $c_a$  以上の階級に含まれるスコア数の合計が  $k$  以上となる最大の整数を示す。次に端末は、階級  $c_a$  の下限値を  $k$  番目のスコア  $S_k$  と推定する。

図 5 (a) のヒストグラムを用いて、推定を行う例を説明する。この図は、図 4 に示すクエリ応答の返信時に  $M_3$  が作成した推定用ヒストグラムを示す。式 (2) より  $n_{H'} = 40$  となるため、 $n = 100$  とすると、 $M_3$  はヒストグラムの各階級に含まれるデータ数を 2.5 倍したものを、ネットワーク全体のデータのスコア分布と推定する (図 5 (b))。式 (1) において、 $k = 10$ 、 $c = 10$  とすると、 $a = 9$  の場合に  $k$  を超えるため、 $M_3$  は  $k$  番目のスコア  $S_k$  を階級  $c_9$  の下限値 80 と推定する。

#### 4. 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 Qualnet4.0<sup>(6)</sup> を用いた。

##### 4.1 シミュレーション環境

1,000[m]×1,000[m] の 2 次元平面上の領域に 50 台の端末 ( $M_1, \dots, M_{50}$ ) が存在する。各端末はランダムウェイポイント<sup>(3)</sup> に従い、0.5 [m/秒] の速度で移動する。停止時間は 60[秒]

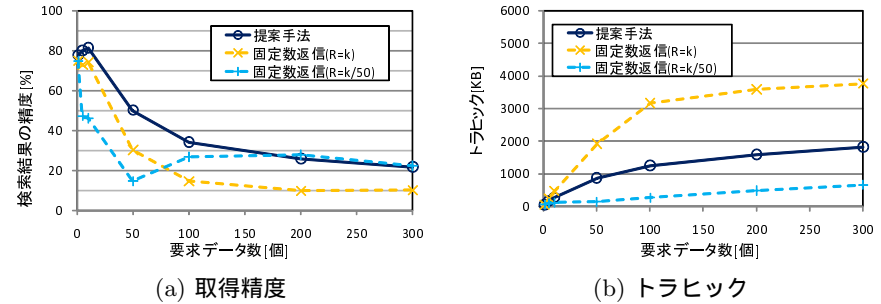


図6 要求データ数  $k$  の影響

とした。各端末は、IEEE802.11b を使用し、伝送速度 11 [Mbps]、通信伝搬距離が 100 [m] 程度となる送信電力でデータを送信する。ネットワーク内には、 $d$  [KB] のサイズのデータが 5,000 個存在し、各端末はそれぞれ 100 個のデータを保持するものとした。また、ネットワーク全体のスコア分布は正規分布に従うとし、スコアの幅は [80, 320] とした。

各端末は、1000 から 2000 [秒] の間隔で Top-k 検索クエリを発行する。ここで、比較手法として、各端末が自身のもつデータの中からスコアの高い  $R = k$ 、 $k/50$  個のデータを固定数返信する手法を用いた。ここで、 $R = k/50$  の場合は、全端末で  $k$  個のデータを返信する。本実験では、要求データ数  $k$  は基本的に 100 とし、4.2 節では 1~300 の間で変化させた。またデータサイズ  $d$  は基本的に 1 [KB] とし、4.3 節では 0.01~10 [KB] 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、各端末の初期位置をランダムに決定し、7,200[秒] を経過させたときの以下の評価値を調べた。

- 平均取得精度  
上位  $k$  個のデータの中で、検索クエリの発行後 60 [秒] の間に取得できたデータの数の割合を取得精度とする。平均取得精度は、シミュレーション時間内に発行された全クエリに対する取得精度の平均である。
- トラフィック  
シミュレーション時間内に発行された全クエリに対する、送信された検索クエリおよびクエリ応答の平均データ量 (1 回分)。

##### 4.2 データ要求数 $k$ の影響

要求データ数  $k$  を変化させた時の提案手法の性能を調べた。その結果を図 6 に示す。これらの図において、横軸は要求データ数  $k$  を表している。縦軸は、図 6 (a) では検索結果の取得精度、図 6 (b) ではトラフィックをそれぞれ表す。

この結果から、 $k$  が大きくなると返信されるデータが多くなるため、全ての手法でトラ

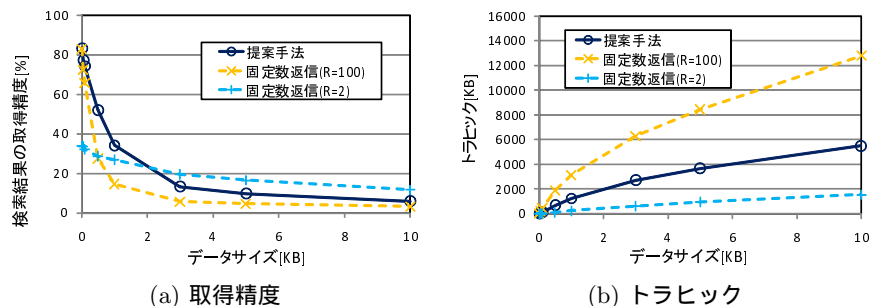


図 7 データサイズの影響

ヒックが増加する。また、トラフィックが増加すると、メッセージが損失する機会が増えるため、検索結果の取得精度が低下する。提案手法は、 $R = k/50$  の場合よりもトラフィックは大きいですが、比較手法よりも  $k$  の増加に伴う検索結果の取得精度の低下を抑止できている。

なお、 $R = k$  の場合において、要求データ数が小さい (50 未満の) 範囲では、各端末の返信データ数が少なくなるため、全端末から要求データ数以上のデータを取得でき、取得精度が非常に高くなる。 $R = k/50$  の場合において、要求データ数  $k = 50$  のときに、取得精度が非常に低くなっている。これは、上位 50 個に入るデータを複数もつ端末において、1 つのデータしか返信しないためである。

#### 4.3 データサイズの影響

データサイズ  $d$  を変化させたときの提案手法の性能を調べた。その結果を図 7 に示す。これらの図において、横軸はデータサイズ  $d$  を表す。縦軸は、図 7 (a) では検索結果の取得精度、図 7 (b) ではトラフィックをそれぞれ表す。

この結果から、 $d$  が大きくなると各端末が返信するクエリ応答のサイズが増加するため、全ての手法でトラフィックが増加する。また、4.2 節と同様にメッセージが損失する機会が増えるため、検索結果の取得精度が減少する。 $d$  が小さい範囲では、提案手法は  $R = 100$  の場合よりも大幅に小さいトラフィックで同等の検索結果の取得精度を実現できている。一方、 $d$  が大きい範囲では、 $R = 2$  の場合よりも検索結果の取得精度が低くなってしまふ。これは、提案手法の方が返信データ数が多いためにクエリ応答のサイズが増加し、パケット損失が多く発生してしまうためである。

### 5. 結 論

本稿では、アドホックネットワークにおいて、トラフィックの削減と検索結果の取得精度の低下の抑止を目的とする Top-k 検索のためのメッセージ処理手法を提案した。提案手法で

は、各端末が検索クエリやクエリ応答にヒストグラムを添付して送信することで、ネットワーク全体のデータのスコア分布と  $k$  番目のスコアを推定する。各端末は推定した  $k$  番目のスコア以上のスコアをもつデータを返信することで、できる限り検索結果に入るデータのみを返信する。また、返信先の端末とのリンクが切断した端末は、他の隣接端末にクエリ応答を転送することで、取得精度の低下を抑止する。

シミュレーション実験の結果から、提案手法は  $k$  番目のスコアを推定して、検索結果に入るデータのみを返信することで、トラフィックを削減しつつ、取得精度の低下を抑止できることを確認した。

本稿では、正規分布や Zipf 分布などの特定のスコア分布は想定しなかったが、データによってはこのような特定のスコア分布に従うものも多い。特定のスコア分布に従う場合、分布の確率密度関数などを用いて、より高精度に分布を推定できる可能性がある。また、提案手法はデータのスコア分布に地理的な局所性があることを想定していないが、実環境ではセンシングデータなど地理的な局所性が強いものが多く存在する。地理的な局所性が存在する場合、それを考慮しスコア分布を予測することで、さらに性能が向上する可能性がある。今後は、このような手法について検討する予定である。

### 参 考 文 献

- 1) R. Akbarinia, E. Pacitti, and P. Valduriez, "Reducing network traffic in unstructured P2P systems using top-k queries," Distributed and Parallel Databases, vol.19, no.2-3, pp.67-86, May 2006.
- 2) 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.
- 3) T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," Wireless Communications and Mobile Computing (WCMC), 2(5): 483-502, 2002.
- 4) R. Hagihara, M. Shinohara, T. Hara, and S. Nishio, "A message processing method for top-k query for traffic reduction in ad hoc networks," Proc. Int. Conf. on Mobile Data Management, pp.11-20, May 2009.
- 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) Scalable Networks: makers of QualNet and EXata, the only multi-core enabled network simulation and emulation software.[Online]. Available: <http://www.scalable-networks.com/>.