

アドホックネットワークにおける 2 フェーズ Top-k 検索手法

佐々木勇和[†] 原 隆浩[†] 西尾章治郎[†]

[†] 大阪大学大学院情報科学研究科マルチメディア工学専攻 〒565-0871 大阪府吹田市山田丘1番5号

E-mail: †{sasaka.yuya,hara,nishio}@ist.osaka-u.ac.jp

あらまし 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。筆者らはこれまでに、アドホックネットワークにおける効率的な Top-k 検索手法を提案している。しかし、この手法では、ネットワーク全体の k 番目のスコアを一部のデータのスコアから推定するため、推定誤差が発生する可能性があり、返信データの絞り込みが十分ではなかった。そこで、本稿では、2 フェーズの通信を用いた Top-k 検索手法を提案する。提案手法では、まず第 1 フェーズでデータのスコアの情報を収集し、その後、収集した k 番目のスコアを閾値として、第 2 フェーズでデータを収集する。これにより、取得精度を維持しつつ、さらなるトラヒックの削減を可能とする。

キーワード アドホックネットワーク, Top-k 検索

A Two-Phase Top-k Query Processing Method in Mobile Ad Hoc Networks

Yuya SASAKI[†], Takahiro HARA[†], and Shojiro NISHIO[†]

[†] Department of Multimedia Engineering, Graduate School of Information Science and Technology,
Osaka University, 1-5 Yamadaoka, Suita, Osaka, 565-0871 Japan

E-mail: †{sasaka.yuya,hara,nishio}@ist.osaka-u.ac.jp

1. はじめに

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 [8], [11]。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 [4], [12]。

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い場合、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、不要なトラヒックが発生する。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。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。

そこで、これまでに筆者の研究グループは、文献 [4], [12] において、アドホックネットワークにおけるトラヒックの削減と検索結果の取得精度の低下の抑止を実現する Top-k 検索手法を提案した。これらの手法では、クエリ経路上の端末のもつデータからネットワーク全体の k 番目のスコアを推定することにより、固定数返信の手法と比較して、不要な返信データの削減と取得精度の維持の両方を実現できる。また、これらの手法では、クエリ実行中のリンク切断の影響を小さくすることを考慮して、クエリ処理を 1 往復の通信のみで実行するアプローチを採用し

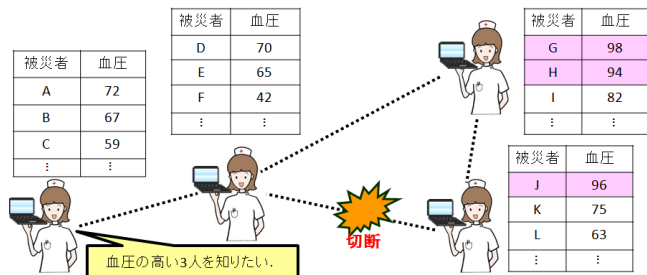


図 1 アドホックネットワークにおける Top-k 検索例

た．しかし，一部の端末のもつデータのスコアのみから k 番目のスコアを推定するため，推定精度は十分ではなく，返信データの絞り込みが十分ではないという問題があった．

そこで，本稿では，2 フェーズの通信を用いた Top-k 検索手法を提案する．提案手法では，まず第 1 フェーズでデータのスコアの情報を収集し，その後，収集した k 番目のスコアを閾値として，第 2 フェーズでデータを収集する．これにより，取得精度を維持しつつ，さらなるトラヒックの削減を目指す．さらに，クエリ応答時にリンク切断を検出した場合，他の隣接端末に返信データを送信することで，取得精度を維持する．提案手法の有効性をシミュレーション実験により評価する．

以下では，2. で関連研究について述べ，3. で提案手法について説明する．その後，4. でシミュレーション実験の結果を示し，最後に 5. 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

本章では，Top-k 検索に関する従来研究について紹介する．Top-k 検索は，さまざま分野において盛んに研究が行われているが，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 [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ヒ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，検索クエリの転送にユニキャストを想定している点，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．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．具体的には，最初の検索クエリによって，各端末から最上位 (1 番目) のデータのスコアを返送してもらい，その中で最もスコアの高いデータ (ネットワーク内で 1 番目のスコアをもつデータ) を取得する．次に，ネットワーク内で 1 番目のスコア

を返送した端末に対して，2 番目のデータのスコアを返送してもらい，他の端末が返送した 1 番目のデータのスコアと比較することで，ネットワーク内の 2 番目のスコアをもつデータを取得する．この処理を k 番目のスコアをもつデータを取得するまで繰り返す．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．

文献 [10] 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．この研究では，検索結果に入らない不要データの返信によるトラヒックが多く，1 章で述べた問題点を解決できない．また文献 [9] では，P2P コンテンツ検索システムにおける Top-k 検索手法として，検索クエリの受信時に各端末が上位 k 個のスコアをもつデータのみを返信する Fixed-k query 手法を提案している．この手法では，各端末が他端末からのクエリ応答を中継する際，上位 k 個に入る可能性のあるデータのみを返信することで，クエリ応答によるトラヒックをある程度削減できるが，基本的に各端末が k 個のデータを返信するため，その効果は十分ではない．また Fixed-k query 手法を拡張し，各端末が，クエリ発行端末からのホップ数と自身の隣接端末数を考慮して，クエリ応答で返信するデータ数を決定する手法，および全ての隣接端末からのクエリ応答を受信した後，返信すべき順位のスコアをもつデータのみを返信する手法を提案している．これらの手法は，クエリ応答によるトラヒックを削減できるが，各端末のもつデータやそのスコアに偏りがある環境では，上位 k 個のデータが返信されず，取得精度が低下する．

これまでに筆者の研究グループは，文献 [4], [12] において，アドホックネットワークにおけるトラヒ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．文献 [4] の手法では，ネットワーク全体のデータのスコア分布を推測するのではなく，クエリが伝搬した経路上の端末がもつデータのスコアから， k 番目のスコアを，そのスコアをもつデータが確実に k 個以上存在することを保証できるスコアとして推定する．具体的にこの手法では，クエリ経路上の端末がもついくつかのデータのスコアを基準値としてクエリに添付し，中継端末においてこれらの基準値が更新されていく．クエリの返送時には，この基準値のうち， k 番目のスコアに該当する値を閾値として，その値以上のスコアをもつデータのみをクエリ発行端末に返信していく．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ができる．文献 [12] では，さらにトラヒックを削減する Top-k 検索手法を提案した．この手法では，各端末が自身の保持するデータのスコアからヒストグラムを作成し，そのヒストグラムからネットワーク全体のスコア分布および k 番目のスコアを推定する．また，各端末はヒストグラムを検索クエリおよび返信に添付することにより，より精度よくスコア分布と k 番目のスコアを推定することができる．さらに，推定した k 番目のスコアと実際の k 番目のスコアの誤差を考慮して，過去の k 番目のスコアやセーフティマージンを用

いて、推定した k 番目のスコアを補正する．返信時には、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により、トラフィックを削減する．しかし、これらの手法では、正確な k 番目のスコアを推定することができないため、検索結果の絞り込みの効果は十分ではない．さらに、文献 [12] で提案した手法では、推定した k 番目のスコアに誤差がある場合、必要なデータが返信されない可能性がある．本稿で提案する手法では、第 1 フェーズにおいて、クエリ発行端末が正確な k 番目のスコアを推定できるため、検索結果のさらなる絞り込みが可能である．

3. 提案手法

本章では、まず想定環境を述べ、その後、提案手法の概要を述べる．最後に、提案手法の検索手順を詳しく説明する．

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．また、データは発生、消滅するため、ネットワーク全体のデータ数 n は動的に変化する．簡単化のため、全てのデータのサイズは等しく、各端末は複製を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．

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．

3.2 提案手法の概要

提案手法では、2つのフェーズ（検索フェーズ、収集フェーズ）によって Top- k 検索を行う．検索フェーズで、クエリ発行端末はネットワーク全体のすべての端末がもつデータのスコアの情報を収集し、収集したスコアの k 番目のスコアを閾値とする．その後、収集フェーズで、クエリ発行端末はクエリに閾値を添付し、閾値以上のスコアをもつデータを収集する．これにより、クエリ発行端末は、ネットワーク全体の k 番目のスコアを把握でき、 k 番目のスコア以上のスコアをもつデータのみを集めることにより、返信データ数を削減することができる．

図 2 を用いて、端末 M_1 が上位 3 個のスコアをもつデータを要求するクエリ ($k = 3$) を発行した場合の検索フェーズの例を示す．表 1 は各端末が保持するデータのスコア（降順）を表す．図 2 において、図中の吹き出しは自身および子端末のもつデータの上位 3 個のスコアを表す．例えば、端末 M_3 が検索クエリを受信した場合、 M_3 には子端末がないため、自身のもつデータの上位 3 個のスコア（81、66、および 65）を返信する． M_1 は、ネットワーク全体の端末がもつデータの上位 3 個のスコアを知ることができるため、検索結果を取得するための閾値（83）を決定することができる．

次に、図 3 を用いて、収集フェーズの例を示す．図 3 において、図中の吹き出しは、自身の返信データのスコアを表す．各

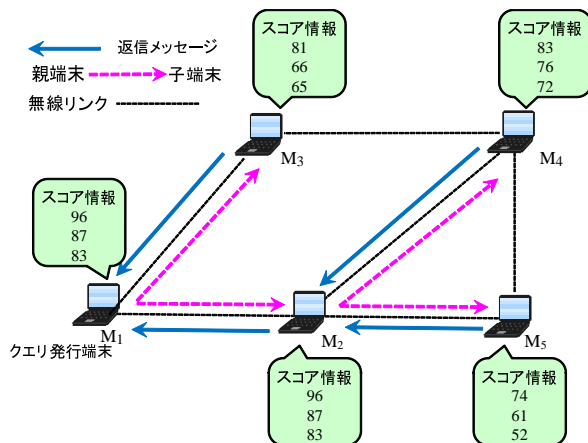


図 2 検索フェーズの例

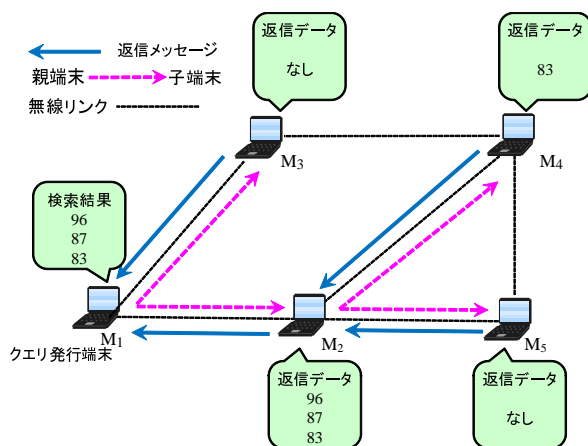


図 3 収集フェーズの例

表 1 各端末が保持するデータのスコア

順位	M_1	M_2	M_3	M_4	M_5
1	79	96	81	83	74
2	60	87	66	76	61
3	58	77	65	72	52
4	53	64	50	69	47
⋮	⋮	⋮	⋮	⋮	⋮

端末は、閾値以上のスコアをもつデータのみを返信することにより、可能な限り返信データ数を減らすことができる．

3.3 検索手順

本節では、メッセージの処理手順を説明する．検索フェーズと収集フェーズでは、メッセージに含まれる要素のみが異なり、基本的に同様のメッセージの処理手順をとる．両方のフェーズとも、まずネットワーク全体の端末に検索クエリをフラッディングし、その後、受信した端末はクエリ応答を返信する．

3.3.1 検索クエリの転送

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．
- (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、および経路端末リストが含まれる。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。収集フェーズの場合は、検索フェーズの際に決定した閾値も検索クエリに含まれる。

(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、自身の識別子 M_q を経路端末リストの末尾に追加し、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、自身が、経路端末リストの末尾から 2 番目の端末の場合、経路端末リストの末尾の端末を自身の子とする。

検索フェーズにおける検索クエリの転送では、各端末は経路端末リストにより、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。収集フェーズの場合、各端末は返信すべきデータのスコアの閾値を知ることができる。

3.3.2 クエリ応答の返信

検索フェーズでは、各端末は自身のもつデータのスコア情報を返信し、収集フェーズでは、閾値以上のスコアをもつデータを返信する。これにより、検索フェーズでは、ネットワーク全体の端末のもつスコアを収集でき、収集フェーズでは、必要なデータ (閾値よりもスコアの高いデータ) のみを収集することができる。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) 自身の子がいない端末 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、返信リストが含まれる。ここで、返信リストには、検索フェーズの場合、 M_r のもつデータの中で上位 k 個のスコアが含まれる。収集フェーズの場合、 M_r のもつデータの中で閾値以上のスコアをもつデータとそのスコアが含まれる。

(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_r は、クエリ応答を作成し、自身の親に送信する。検索フェーズの場合、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_r のもつデータのスコア中で、上位 k 個のスコアが含まれる。収集フェーズの場合、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_r のもつデータの中で、

閾値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。

(3) クエリ応答の返信後、別の子からクエリ応答メッセージを受信した端末 M_r は、クエリ応答メッセージを作成し、自身の親に送信するが、フェーズによって処理が異なる。検索フェーズの場合、受信したクエリ応答メッセージをそのまま転送する。収集フェーズの場合、手順 (2) と同様の処理を行うが、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。

クエリ応答の返信では、各端末は、スコア情報、もしくは、閾値以上のスコアをもつデータのみを返信する。データサイズはスコア情報に比べサイズが大きいので、不必要なデータ (スコアの高い上位 k 個以外のデータ) を返信しないことにより、トラフィックを大幅に削減できる。

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、クエリ応答メッセージ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、リンク切断を検出した端末は、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。以下では、親とのリンク切断を検出した端末 M_t が、別経路でクエリ応答を返信する動作について説明する。

(1) リンク切断時に親端末と切断した端末 M_t は、親以外の隣接端末にクエリ応答メッセージを転送する。ここで、自身の隣接端末リストから、転送先の隣接端末を決定する。複数の端末と隣接している場合、トラフィックや遅延の増分をできる限り小さくするため、隣接端末のうちクエリ発行端末からのホップ数が最小のものに転送する。

どの端末とも隣接していない場合、クエリ応答メッセージを破棄する。

(2) 自身の子以外の端末からクエリ応答メッセージを受信した端末 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、受信したメッセージを子からのメッセージと同様に扱い、3.3.2 項の手順に従って親に返信する。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 が親に返信したメッセージは端末 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メッセージ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、クエリ応答メッセージを破棄する。

リンク切断の検出では、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、検索結果をできる限りクエリ発行端末まで返信し、取得精度の低下を抑止できる。ただし、切断を検出した端末が孤立している場合や親からメッセージを受信した端末が他の端末と隣接していない場合、クエリ応答がクエリ発行端末まで返信されないため、取得精度が低下してしまう。

表 2 パラメータ設定

パラメータ	意味	値
k	要求データ数	100 および 10 (1~100)
d	データサイズ	1[KB] (0.01~10)
R	領域サイズ	600[m] (600~2,000)

4. シミュレーション評価

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 [13] を用いた。

4.1 シミュレーション環境

$R[m] \times R[m]$ 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 [3] に従い、0.5 [m/秒] の速度で移動する。停止時間は 60[秒] 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] のサイズのデータ存在し、各端末が 30 から 200 個のデータを保持するものとする。具体的には、シミュレーション開始時に各端末は 80 から 120 個のデータを保持しており、シミュレーション開始後、2400 秒までは、各端末のデータ数が 300 秒ごとに 0 から 5 個の範囲でランダムに増加し、2400 秒から 4800 秒まで 300 秒毎に 0 から 5 個の範囲でランダムにデータが消滅するものとした。さらに、4800 秒経過後は、300 秒毎に 0 から 5 個の範囲でランダムにデータが発生するものとした。また、ネットワーク全体のスコア分布は正規分布に従うものとし、スコアの幅は [80, 320) とした。

各端末は、3000 [秒] の間隔で Top-k 検索クエリを発行する。Top-k 検索手法には、提案手法、文献 [12] で提案した手法、および単純手法を用いた。文献 [12] で提案した手法では、各端末がクエリおよび応答にデータのスコアから構築されたヒストグラムを添付し、それをもとに k 番目のスコアを推定する。文献 [12] では、推定方法として、2 つの手法を提案しており、本実験では、確率密度関数を用いて k 番目のスコアを推定する方法を用いた。単純手法は、各端末が自身のもつデータの中からスコアの高い k 個のデータを返信する手法である。

Top-k 検索クエリには制限時間を設けて、クエリ発行端末は、 α 秒経過後、全ての端末から返信がない場合、そこで検索を終了する。予備実験の結果より、各手法（および各パラメータ）において、最も取得精度が高くなる α を用いた。提案手法のみ 2 種類の α を用い、1 つは最も取得精度が高くなる値、もう 1 つは文献 [12] で提案した手法において最適な α 以下の値である。

本実験のパラメータを表 2 に示す。左側の値を基本的には用い、その影響を調べる際に、括弧内の値で変化させる。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 α [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。

- トラヒック: 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（クエリ 1 回分）をトラヒックとする。

- 取得時間: 全クエリに対する、クエリ発行端末が検索クエリを発行してから検索結果を取得するまでの平均時間（最大 α [秒]）を取得時間とする。

4.2 データサイズの影響

データサイズ d を変化させた時の各手法の性能を図 4 に示す。図中の各グラフにおいて、「Proposed」は、提案手法において制限時間が適切な場合、「Proposed(short)」は、提案手法において制限時間を短く設定した場合、「Method in [12]」は、文献 [12] で提案されている手法、「Fixed k 」は、各端末が自身のもつデータの中からスコアの高い k 個のデータを返信する単純手法を示す。これらのグラフにおいて、横軸はデータサイズ d を表している。縦軸は、図 4(a) では取得精度、図 4(b) ではトラヒック、図 4(c) では取得時間を表している。

図 4(a) の結果から、 d が大きくなるほど、取得精度が低下することがわかる。これは、データサイズが大きくなると、パケットロスが頻繁に発生するためである。データサイズが小さい時、全ての手法で取得精度が高い。しかし、取得精度が 100% を達成できないのは、ネットワーク分断が発生しているためである。データサイズが大きい時、提案手法は他の手法より取得精度が高い。これは、図 4(b) の結果が示すように、提案手法が不必要な返信データを効果的に削減するため、トラヒックが小さく、パケットロスの発生頻度が他の手法より少ないからである。

図 4(c) の結果から、提案手法（“Proposed”）は、文献 [12] で提案されている手法より取得時間が長い。これは、スコア情報の収集とデータの収集（検索フェーズと収集フェーズ）のために 2 往復の通信を用いているためである。単純手法では、 d が大きくなるにつれて、初め取得時間が増加し、その後、減少する。これは、 d が大きい場合、パケットロスが頻繁に発生するため、待ち時間 (α) が短い方が不必要な再送を避けることができ、取得精度が上がるためである。提案手法における 2 つの制限時間を比較すると、理想的な制限時間の方が、スコアの情報を多く集めることができるため、より取得精度が若干高くなる。

4.3 要求データ数の影響

次に、要求データ数 k を変化させた時の各手法の性能を図 5 に示す。これらのグラフにおいて、横軸は要求データ数 k を表している。縦軸は、図 5(a) では取得精度、図 5(b) ではトラヒック、図 5(c) では取得時間を表している。

この結果から、 k が大きくなると、提案手法の制限時間が短い場合（“Proposed (short)”）と文献 [12] で提案されている手法では、初め取得精度が下がり、その後上がって、再び下がる。また、他の手法では、 k が大きくなるにつれて、取得精度は低下する。 k が非常に小さい場合、理想的な制限時間の提案手法と単純手法の取得精度が他の手法より高い。制限時間が短い場合の提案手法では、クエリ制限時間が短いため、すべての端末からの返信を待つことができず、取得精度が低下してしまう。特に、 k が非常に小さいときには、検索結果に入るべきデータを取りこぼした際の影響が大きくでている。また、文献 [12] で

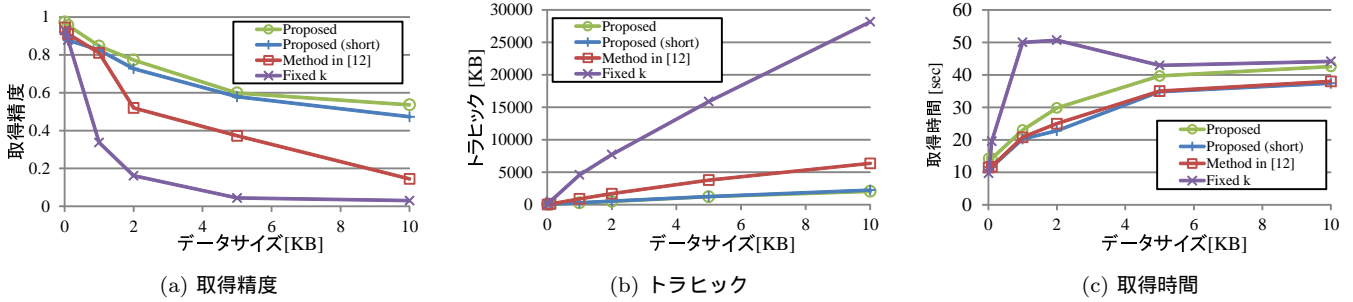


図 4 データサイズの影響

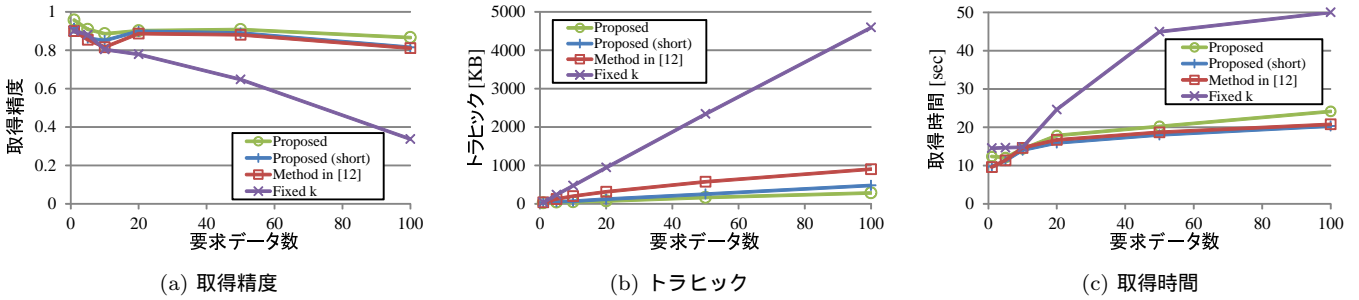


図 5 要求データ数 k の影響

提案された手法では、 k が非常に小さい場合、推定誤差の影響が大きく、必要なデータが返信されない際に、取得精度が低下する。

k が大きくなると、パケットロスが発生し、すべての手法で取得精度が低下するが、図 5(b) の結果より、トラヒックに大きな差がでているのがわかる。理想的な制限時間の提案手法では、トラヒックを可能な限り削減できるため、他の手法より取得精度が高い。図 5(b)、および図 5(c) の結果より、 k が大きくなると、返信メッセージのサイズが大きくなるため、トラヒックおよび取得時間が大きくなる。提案手法では、第 1 フェーズでスコア情報を収集するため、文献 [12] で提案する手法より、返信データ数を削減できる。しかし、2 往復の通信を用いるため、取得時間が若干長くなる。提案手法の制限時間が短い場合、文献 [12] とほぼ同じ取得時間だが、トラヒックが小さくなっている。単純手法は、返信データ数が多いため、パケットロス、および再送の影響が大きく、他の手法に比べてトラヒック、および取得時間ともに大きい。

4.4 領域サイズの影響

最後に、領域サイズ R の影響を調べるために、異なる領域サイズにおいて、データサイズを変化させた時の各手法の性能を図 6、図 7 および図 8 に示す。領域サイズは、図 6 では 1000 [m]、図 7 では 1500 [m]、図 8 では 2000 [m] とした。全ての場合において、要求データ数 k は 10 とした。領域を広くすると、一般的に接続できる端末数が減少し、それに伴って返信データ数も少なくなり、必要なデータが返信されない可能性が高くなる。そこで、必要なデータを取りこぼした際の影響を調べるため、 $k = 10$ と小さい値を用いた。これらのグラフにおいて、横軸はデータサイズ d を表している。縦軸は、図 6(a)、図 7(a) および図 8(a) では取得精度、図 6(b)、図 7(b) および図 8(b) ではトラヒック、図 6(c)、図 7(c) および図 8(c) では取得時間

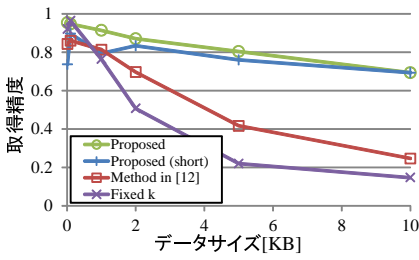
を表している。

図 6(a)、図 7(a)、および図 8(a) の結果から、領域が広がると、すべての手法で取得精度が低下する。これは、領域が広がったことにより、ネットワーク分断がより頻繁に発生するためである。提案手法の制限時間が短い場合 (“Proposed (short)”) において、 d が非常に小さい場合、他の手法より精度が低い。これは、クエリの制限時間が短いために、データの返信を十分に待たず、必要なデータを取得できないためである。提案手法 (“Proposed”) は、特に、ホップ数が長く、 d が小さい (データ返信のための通信時間が短い) 場合、2 往復の通信をすることによる遅延が制限時間に大きく影響し、比較手法より長い取得時間が必要となる。

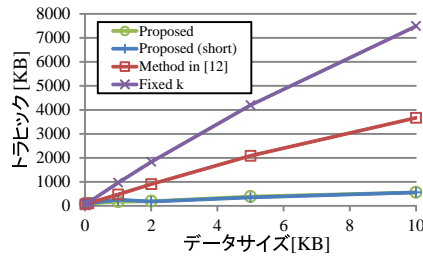
図 6(a)、図 7(a)、および図 8(a) の結果から、領域が 1500 [m] ($R = 1500$ [m]) のときが最もトラヒックが大きく、領域が 2000 [m] ($R = 2000$ [m]) のときが最もトラヒックが小さい。領域が 1500 [m] のとき、領域が 1000 [m] のときに比べて、相互接続している端末間のホップ数が増加するため、トラヒックが大きくなる。一方、領域が 2000 [m] のとき、ネットワーク分断が発生するために、返信データが少なく、トラヒックが小さくなる。取得時間は、領域が 1500 [m] のとき、ホップ数が長い場合、最も取得時間が長い。

5. おわりに

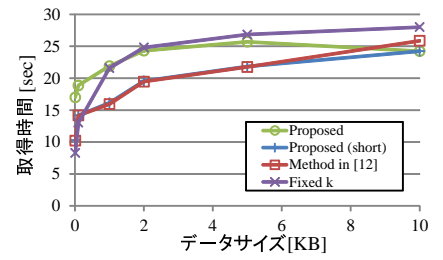
本稿では、2 フェーズの通信を用いた Top- k 検索手法を提案した。提案手法では、まず第 1 フェーズでデータのスコア情報を収集し、その後、収集した k 番目のスコアを閾値として、第 2 フェーズでデータを収集する。これにより、従来手法と比較して、取得精度を維持しつつ、さらなるトラヒックの削減できる。さらに、クエリ応答時にリンク切断を検出した場合、他の隣接端末に送信することで、取得精度を維持する。



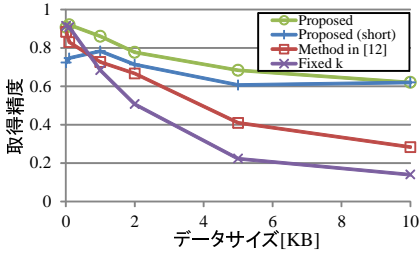
(a) 取得精度



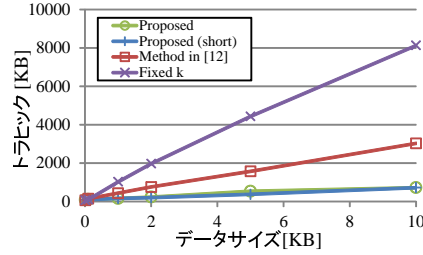
(b) トラフィック



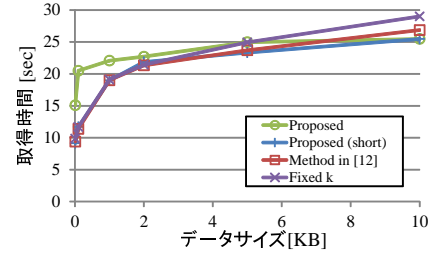
(c) 取得時間

図 6 データサイズの影響 ($R=1000[m]$)

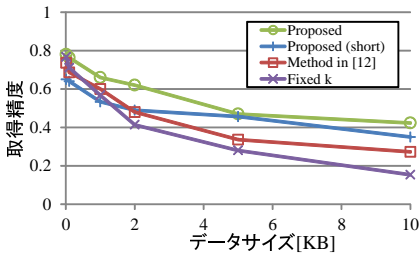
(a) 取得精度



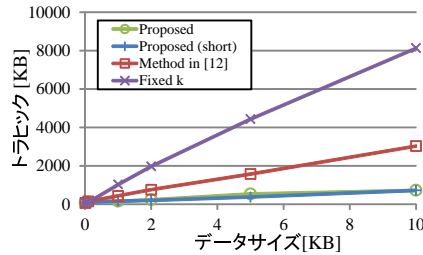
(b) トラフィック



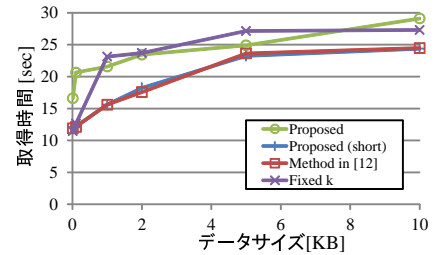
(c) 取得時間

図 7 データサイズの影響 ($R=1500[m]$)

(a) 取得精度



(b) トラフィック



(c) 取得時間

図 8 データサイズの影響 ($R=2000[m]$)

シミュレーション実験により、提案手法は比較手法よりもトラフィックを削減できることがわかった。トラフィックを削減することにより、パケットロスも抑えることができ、取得精度も向上した。しかし、2回の通信を行うことにより、取得時間が文献[12]で提案した手法よりも長くなってしまふ。検索クエリの制限時間を短くすることも可能であるが、必要なデータの取りこぼしにより、取得精度の低下が発生する可能性がある。

本稿で提案した手法では、第1フェーズでネットワーク全体の端末からスコア情報を収集する。しかし、端末数が多い場合やホップ数が長い場合などでは、スコア情報の収集に長い時間がかかってしまふ。そこで今後は、一部の端末のみからスコア情報を収集し、それを用いて k 番目のスコアを推定する手法の検討を予定している。

謝 辞

本研究の一部は、(財)近畿移動無線センター・モバイルワイヤレス助成金、文部科学省科学研究費補助金・基盤研究S(21220002)、特定領域研究(18049050)、および総務省委託研究「ユビキタスサービスプラットフォーム技術の研究開発」の研究助成によるものである。ここに記して謝意を表す。

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, 2006.
- [2] W.-T. Balke, W. Nejdl, W. Siberski, and U. Thaden, "Progressive distributed top-k retrieval in peer-to-peer networks," in *Proc. Int. Conf. on Data Engineering*, pp. 174-185, 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, 2009.
- [5] T. Hara, and S.K. Madria, "Data replication for improving data accessibility in ad hoc network," *IEEE Trans. Mob. Comput.*, vol. 5, no. 11, pp. 1515-1532, 2006.
- [6] T. Hara, and S.K. Madria, "Consistency management strategies for data replication in mobile ad hoc networks," *IEEE Trans. Mob. Comput.*, vol. 8, no. 7, pp. 950-967, 2009.
- [7] T. Hara, N. Murakami, and S. Nishio, "Replica allocation for correlated data items in ad hoc sensor networks," *SIGMOD Record*, vol. 33, no. 1, pp. 38-43, 2004.

- [8] D. B. Johnson, "Routing in ad hoc networks of Mobile Hosts," Proc. IEEE WMCSA'94, pp. 158–163, 1994.
- [9] H. Matsunami, T. Terada, and S. Nishio, "A query processing mechanism for top-k query in P2P networks," in *Proc. Int. Conf. in SWOD*, pp. 84–87, 2005.
- [10] 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, 2006.
- [11] Perkins. C. E, and Ooyer. E .M, "Ad-hoc On-Demand Distance Vector Routing," Proc. IEEE WMCSA'99, pp.90–100, 1999.
- [12] 佐々木勇和, 原 隆浩, 西尾章治郎, "アドホックネットワーク上のスコア分布推定による Top-k 検索手法," 情報処理学会論文誌, Vol. 52, No. 1 (2011 年 1 月).
- [13] Scalable Networks: makers of QualNet and EXata, the only multi-core enabled network simulation and emulation software.: <http://www.scalable-networks.com/>.