

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

萩原 亮^{†1} 篠原 昌子^{‡2}
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。本論文では、トラヒックの削減と検索結果（上位 k 個のデータ）の取得精度の維持を実現する Top-k 検索のためのメッセージ処理手法を提案する。提案手法では、検索結果に入らないデータの返信を抑制するため、各端末が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。また、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必ず取得できるようにするため、各端末がリンク切断の検出時に、他の隣接端末を経由する返信経路を検出し、クエリ応答をクエリを発行した端末まで返信する。

A Message Processing Method for Top-k Query in Ad Hoc Networks

RYO HAGIHARA,^{†1} MASAKO SHINOHARA,^{‡2}
TAKAHIRO HARA^{†1} and SHOJIRO NISHIO^{†1}

In 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 message processing method for a top-k query for reducing traffic and keeping the accuracy of the query result, i.e., guaranteeing that data items with the k highest scores in the entire network are acquired. In this method, when a mobile node transmits query and reply messages, it can reduce candidates of data items included in the top-k result to prevent unnecessary data transmissions. Moreover, if a mobile node detects a disconnection of a radio link, it searches for an alternative path to transmit the reply message to the query-issuing node.

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ともない、ルータ機能を持つ無線通信端末のみで一時的な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では、既存の通信基盤を必要とせずに、端末のみで自律分散的にネットワークを構築できるため、災害時の救助活動における情報共有やセンサネットワークでの情報収集への応用が期待されている。このようなアプリケーションでは、限られた通信帯域を効率的に利用するため、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には、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上位 k 個のスコアを持つデータを返信した場合、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）を必ず取得できる。しかしこの場合、実際には上位 k 個以内でない（検索結果に入らない）データまで返信されてしまうため、不要なトラヒックが発生してしまう。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必要なデータを取得できず、検索結果の取得精度が低下する可能性がある。

たとえば図 1 は、災害時の救助活動において、救助隊員が被災者の負傷度の情報を管理し、隊員の持つ端末どうしてネットワークを構築している様子を示す。このような環境において、左下の隊員が、限られた数の救急車で搬送する負傷度の高い 2 人の被災者を検索する場合、各隊員が自身の管理情報から負傷度の高い 2 人の被災者の情報を返信すると、救急車で搬送できる人数を超える被災者の情報が返信されてしまう。また、右下と左上の隊員の持つ端末間の無線リンクが切断された場合、右下の隊員の持つ被災者の情報が返信されず、左下の隊員は負傷度の高い 2 人の被災者の情報を取得できない。

^{†1} 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
Department of Multimedia Engineering, Graduate School of Information Science and Technology,
Osaka University
^{‡2} 株式会社富士通研究所
Fujitsu Laboratories Ltd.

ラヒックを削減する手法を提案している．この研究は，過去のクエリ応答が再利用される環境を想定している点で本研究とは異なる．

3. Top-k 検索のためのメッセージ処理手法

本章では，まず提案手法の設計方針，および想定環境について述べ，その後，各端末が設定する基準値について説明する．次に，基準値を用いた検索クエリの転送，およびクエリ応答の返信について説明する．さらに，リンク切断時の返信経路の検出について説明する．

3.1 設計方針

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．Top-k 検索の実行中に端末間のリンクが切断された場合，検索結果の取得精度が低下する可能性があるため，提案手法では，検索クエリを発行した端末が，検索結果を短時間で取得できるよう，1 回の検索クエリの送信とクエリ応答の返信で Top-k 検索を実行する．

また，アドホックネットワークでは通信帯域が限られているため，各端末が返信するデータ数が多い場合，パケット衝突による転送遅延の増大やスループットの低下が発生してしまう．そこで提案手法では，各端末が返信するデータ数をできる限り削減できるように，検索結果に入るスコアを推測し，推測したスコアに基づいて返信データを決定する．

さらに，提案手法では，Top-k 検索の実行中にネットワークポロジが変化した場合においても，検索結果の取得精度を維持するため，リンク切断を検出した端末は他の経路を探索し，クエリ応答を確実に検索クエリを発行した端末まで返信する．

3.2 想定環境

本論文では，アドホックネットワークを構成する端末が，他の端末が持つデータに対して，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，ネットワーク内の端末の持つデータの中から，上位 k 個のスコアを持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．本論文で提案するメッセージ処理手法は，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 ）が存在する．各端末は，自身の隣接端末を把握しているものとする．取得可能なデータ集合を統一するため，端末の移動によってネットワークポロジが変化しても，ネットワークの分断は発生しないものとする．また，検索クエリやクエリ応答を転送している間に，切断したリンクが再び接続することはしないものとする．これらの想定は，本論文で提案するメッセージ処理手法に必須では

ないが，簡単化のために行う．

各端末は， c 個のデータを保持している．簡単化のため，すべてのデータのサイズは等しく，データの更新は発生しないものとする．さらに，データの複製は作成されないものとする．

3.3 基準値の設定

提案手法では，各端末が検索結果に入るスコアを推測し，そのスコアの一部を検索クエリに添付することで，検索結果の候補を絞り込む．ここで，検索クエリに添付するスコアの数をもくした場合，検索クエリを受信した端末は検索結果に入るデータを高い精度で推測できる．たとえば，上位 k 個のスコアのすべてを検索クエリに添付した場合，検索クエリを受信した端末は，検索クエリを中継したすべての端末と自身が持つデータに対して，上位 k 個のスコアを正確に把握できる．しかし， k 個のスコアを検索クエリに添付するため，要求データ数 k が大きくなるにつれて，検索クエリによるトラヒックも増大してしまう．一方，検索クエリに添付するスコアの数が少ない（たとえば， k 番目のスコアの 1 つのみを検索クエリに添付した）場合，検索クエリによるトラヒックは k に依存しないが，各端末が検索結果に入るデータを推測する精度が低くなり，絞り込みの効果が小さくなる．

そのため，提案手法では各端末が推測した上位 k 個のスコアの中から， N ($\leq k$) 個のスコアを検索クエリに添付し，その値を基準値に設定する．本論文では，簡単化のため，各端末が推測した上位 k 個のスコアの中から， N ($\leq k$) 個のスコアを均一に抽出し， $\frac{k}{N}$ 個ごとの値を基準値に設定する．すなわち，端末 M_a における N 個の基準値 $B_a(i)$ ($1 \leq i \leq N$) はそれぞれ，検索クエリを発行した端末が $\frac{k}{N}i$ ($1 \leq i \leq N$) 個のデータを取得できることを保証するスコアを示す．ここで， M_a の基準値の全体集合を $B_a = \{B_a(1), B_a(2), \dots, B_a(N)\}$ とする．また，検索クエリを受信した端末は，検索クエリに含まれる基準値と自身の持つデータのスコアから，基準値をより精度の高いものへと更新していく．

たとえば， $B_a(1)$ つまり $\frac{k}{N}$ 位のスコアと，クエリ受信端末が持つデータのうちの $\frac{k}{N}$ 番目のスコア（ここでは s_1 とする）に対して， $B_a(2) \leq s_1 \leq B_a(1)$ となるとき，ネットワーク内にスコアが s_1 以上のデータが少なくとも $\frac{2k}{N}$ 個は存在するため， $B_a(2)$ を現在の値から s_1 に変更できる．このように，多数の基準値を設定し，それらを細かく変更していくことで，検索結果に入るスコアを高い精度で推測できる．

3.4 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，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各自の基準値を更新

することで、絞り込みの効率を向上させる．以下では、Top-k 検索を発行した端末 M_p と検索クエリを受信した端末の動作について説明する．

- (1) M_p は検索条件、および検索での要求データ数 k を指定する．また、検索条件から自身の持つデータのスコアを計算し、自身の基準値を以下のように設定する．

$$B_p(i) = S_p \left(\frac{k}{N} i \right) \quad (1 \leq i \leq N)$$

ここで $S_a(b)$ は、端末 M_a が算出したスコアのうちの b 番目のスコアを示す．つまり、 M_p の基準値は、自身の持つデータの $\frac{k}{N}$ 個ごとのスコアを示す．

- (2) 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、 M_p の基準値リスト $B_p = \{B_p(1), B_p(2), \dots, B_p(N)\}$ 、およびホップ数が含まれる．基準値リストにはメッセージを送信した端末の持つ基準値が格納され、ホップ数は検索クエリ発行端末までのホップ数を示し、 M_p 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．ここでは、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受信した端末 M_q は、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．受信した検索クエリメッセージが初めて受信したものであれば、 M_q はメッセージ送信端末を自身の親とし、手順 (1) と同様に検索条件から自身のスコアを算出する．その後、手順 (4) へ進む．一方、受信した検索クエリメッセージがすでに前に受信したものであれば、手順 (6) へ進む．
- (4) 検索クエリメッセージを初めて受信した M_q は、自身の基準値を式 (2) のように設定する．

$$B_q(i) = RANK(i, S_{q-base} \cup L) \quad (1 \leq i \leq N)$$

ここで、集合 $S_{q-base} = \{S_q(\frac{k}{N} \cdot 1), S_q(\frac{k}{N} \cdot 2), \dots, S_q(\frac{k}{N} \cdot N)\}$ は M_q の持つデータをスコアの降順に並べた際の $\frac{k}{N}$ 個ごとのスコアの全体集合を示し、集合 $L = \{L(1), L(2), \dots, L(N)\}$ はメッセージに含まれる基準値リストのスコアの全体集合を示す．また、 $RANK(j, A)$ は集合 A の中で j 番目のスコアを示す．つまり、 M_q の基準値は集合 S_{q-base} と集合 L の中で上位 N 個のスコアとなる．これにより、 M_q は自身も含めてメッセージが転送された経路上の端末の持つデータの中に、 $B_q(i) (1 \leq i \leq N)$ 以上のスコアを持つデータがそれぞれ $\frac{k}{N}i$ 個以上存在することを保証できる．手順 (5) へ進む．

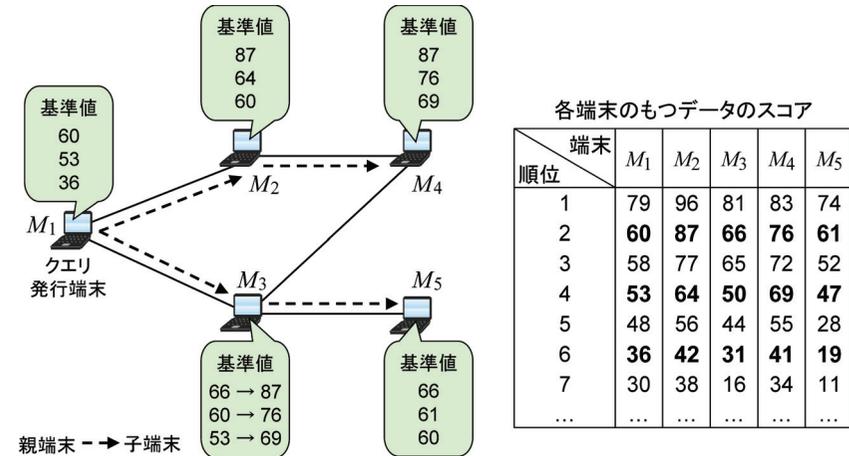


図 2 検索クエリの転送
Fig. 2 Query transmission.

- (5) M_q は、メッセージ送信端末とその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 $B_q(i) (1 \leq i \leq N)$ に更新する．また、ホップ数を 1 増やし、自身の隣接端末に送信する．手順 (3) へ戻る．
- (6) 検索クエリメッセージを再受信した M_q は、メッセージ送信端末の親端末が自身である場合、メッセージ送信端末を自身の子とする．そうでない場合、手順 (4) と同様に、自身の基準値 $B_q(i) (1 \leq i \leq N)$ を更新する．ただし、トラフィックの増加を抑制するため、検索クエリメッセージは再送信しない．

検索クエリの転送では、基準値を用いることで検索結果の候補となるデータを絞り込む．また、各端末が他の端末から受信した基準値を用いて、自身の基準値を逐次更新することで、検索結果の候補となるデータをさらに絞り込む．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を把握できる．また、ホップ数とメッセージ送信端末の組から、親子関係のない隣接端末と自身の位置関係を把握できる．

図 2 は、端末 M_1 が、 $k = 6, N = 3$ 、つまり 3 個の基準値を用いて上位 6 個のスコアを持つデータを検索する動作を示す．吹出しは、各端末が最初に設定した基準値を、右側は後から更新した基準値を示す．また図中の表は、各端末が検索条件から算出したデータのスコア

コアを示し、太字は各端末における $\frac{k}{N}$ 個ごとのスコアを示す。たとえば、 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 60, 53, 36 と自身の $\frac{k}{N}$ 個ごとのスコア 87, 64, 42 の中から上位 3 個のスコア 87, 64, 60 を M_2 の基準値とし、検索クエリの転送を続ける。一方、 M_3 は、 M_1 から受信したメッセージにより、基準値を 66, 60, 53 に設定した後、自身と親子関係にない M_4 からメッセージを再び受信し、さらに基準値を 87, 76, 69 に更新する。

3.5 クエリ応答の返信

3.4 節より、各端末は検索結果の候補となるデータを絞り込んでいるため、クエリ応答の返信では、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端末が、受信した情報から返信の必要がないデータを検出した場合、そのデータを返信しない。以下では、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 M_r 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 の識別子、閾値、および返信リストが含まれる。閾値はメッセージを返信する端末が推測した k 番目のスコアで、返信リストには自身の持つ閾値以上のスコアとそのスコアを持つデータが格納される。ここで 3.4 節より、 M_r は自身の基準値 $B_r(i)$ ($1 \leq i \leq N$) 以上のスコアを持つデータが、ネットワーク内に $\frac{k}{N}i$ 個以上あることを保証できるため、自身の最小の基準値である $B_r(N)$ を閾値とする。
- (2) クエリ応答メッセージを受信した M_s は、自身の閾値を、受信したメッセージに含まれる閾値と自身の最小の基準値 $B_s(N)$ の大きい方に設定する。 M_s はメッセージに含まれる返信リストから、閾値未満のスコアを持つデータとそのスコアを削除する。また、自身の持つデータのうち、閾値以上のスコアを持つデータと、そのスコアを返信リストに追加する。ここで、返信リストに含まれるデータ数が k より大きくなる場合、閾値を返信リストの k 番目のスコアに更新し、閾値未満のスコアを持つデータとそのスコアを返信リストから削除する。 M_s は、すべての子からクエリ応答を受信すると、自身の親に更新したクエリ応答メッセージを送信する。
この処理を、検索クエリ発行端末がすべての子からクエリ応答メッセージを受信するまで繰り返す。

クエリ応答の返信では、検索結果の候補となるデータのみを返信することで、トラフィックを削減する。また、閾値を用いることで、中継端末で検索結果に入らないと推測されたデー

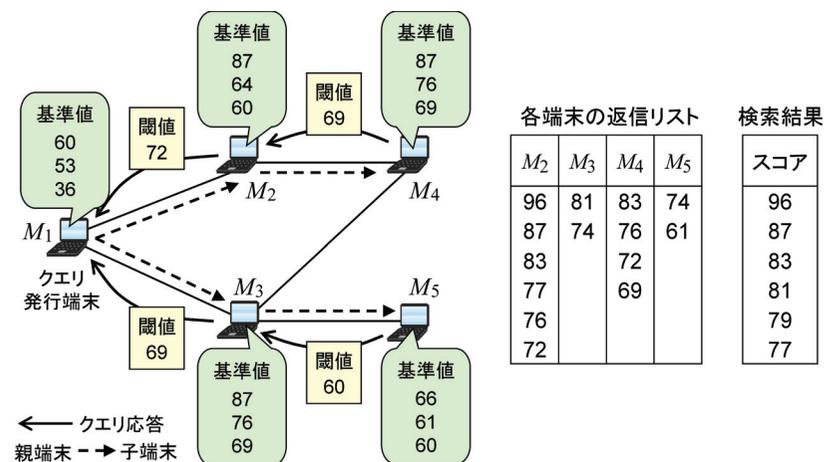


図 3 クエリ応答の返信例
Fig. 3 Transmission of query-reply.

タを返信しないため、さらにトラフィックを削減できる。

図 3 を用いて、クエリ応答の返信例を説明する。図において、吹き出しはクエリの転送が終了した時点での各端末の基準値、端末間のリンク上の四角は、クエリ応答メッセージに含まれる閾値を示す。また図中の各表は、各端末が返信したクエリ応答メッセージに含まれる返信リスト、および M_1 が取得した検索結果を示す。たとえば、子がない M_5 は、3 番目の基準値 60 を閾値とし、60 以上のスコアとそのデータを M_3 に返信する。 M_3 は自身の閾値を、 M_5 から受信したメッセージに含まれる閾値 60 と自身の基準値 69 の大きい方に更新する。 M_3 は、返信リストから閾値 69 未満のデータとそのスコア (61) を削除し、自身の持つデータのうち、スコアが閾値 69 以上のデータとそのスコア (81) を追加する。 M_2 も同様に閾値を 69 とするが、 M_2 の持つスコア 69 以上のデータを追加すると、返信リストに含まれるデータ数が $7 (> k)$ となるため、閾値を 6 番目のスコア 72 に更新し、不要な 7 番目のスコアを持つデータとそのスコア (69) を削除してから M_1 に返信する。

3.6 リンク切断時の処理

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ト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、他の隣接端末を經由

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送信する動作について説明する。

- (1) M_t は、自身の子以外の隣接端末にクエリ応答メッセージを転送する。複数の端末と隣接している場合、トラヒックおよび遅延をできる限り抑制するため、自身の隣接端末のうち、クエリ発行端末までのホップ数が最小のものに転送する。一方、自身の子とのリンク切断を検出した端末は、その子の情報を削除する。切断した子を除いたすべての子からクエリ応答メッセージを受信すると、親にクエリ応答メッセージを送信する。
- (2) 自身の子以外からクエリ応答メッセージを受信した M_u は、自身がクエリ応答メッセージを親に送信していない（別の子端末からの返信を待っている）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、3.5 節の手順に従って親に返信する。一方、 M_u がクエリ応答メッセージをすでに親に送信している場合、3.5 節の手順 (2) と同様に、 M_u の閾値を更新して、クエリ応答メッセージを自身の親に返信する。ただし、データの重複を防ぐため、 M_u がすで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。

ここで、 M_t が自身の子のみと隣接している場合、隣接端末に送信したクエリ応答メッセージは M_t に返信されてしまう。この場合 M_t は以下の手順により、経路探索メッセージを用いることで、クエリ応答を送信する別経路を構築する。

- (1) M_t は、すべての自身の子に経路探索メッセージを送信する。このメッセージには、メッセージ送信端末 M_t および経路情報リストが含まれる。経路情報リストには、リンク切断を検出した端末からメッセージ送信端末までの経路上の端末の識別子が格納される。ここでは、 M_t のみとなる。
- (2) 経路探索メッセージを受信した端末 M_v は、受信したメッセージに含まれるメッセージ送信端末を親としない隣接端末が存在するか調べる。存在する場合、 M_v から検索クエリ発行端末 M_p までの経路は切断した無線リンクを含まないため、この経路を経路情報リストに格納し、メッセージを M_t に返信する。これを受信した M_t は検出した経路に沿ってメッセージを送信し、 M_v も 3.5 節の手順と同様にクエリ応答メッセージを親に返信する。一方、上記のような隣接端末が存在しない場合、 M_v は、経路情報リストに自身を追加し、自身の子に経路探索メッセージを送信する。

親とのリンク切断を検出した端末は、自身の子以外の隣接端末にクエリ応答メッセージを送信する。また、自身の子とのみ隣接している端末は、自分自身にクエリ応答メッセージを返信

していない端末までの経路を構築し、構築した経路に沿ってクエリ応答メッセージを返信する。これにより、検索クエリ発行端末は、ネットワーク内のすべての端末からクエリ応答メッセージを受信できる。

図 3 を用いて、リンク切断時の処理例を説明する。たとえば、 M_1-M_3 間のリンクが切断した場合、 M_3 は自身の子でない M_4 にクエリ応答メッセージを送信する。一方、 M_1-M_2 間のリンクが切断した場合、 M_2 は経路探索メッセージを自身の子である M_4 に送信する。 M_4 は、 M_2 を親としない隣接端末 M_3 を発見し、経路情報リスト (M_2 M_4 M_3) を M_3 に返信し、 M_3 は検出した経路に沿ってクエリ応答メッセージを送信する。

4. 性能評価

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。

4.1 シミュレーション環境

2次元平面上に、100 台の端末（識別子： M_1, M_2, \dots, M_{100} ）を各端末の隣接端末の数が 1 から 5 台になるように配置した。各端末は、ネットワークの分断が発生しないように、0 から 10 [m/秒] の乱数で設定された速度で、0 から 2π [rad] の乱数で設定された方向に移動するものとした。各端末は、IEEE 802.11g や Bluetooth などの無線通信方式を用いていることを想定し、無線通信半径は 70 [m] の円とした。また、各端末はデータサイズが C であるデータを 200 個持つものとし、各端末の持つデータのスコアは一樣乱数に従うものとした。

ここで、最大のホップ数が 10 ホップほどのネットワークであれば、11 [Mbps] 程度の通信帯域を用いて、100 [KB] 程度のメッセージを送信する場合、検索クエリの転送およびクエリ応答の返信にかかる時間は 1 [秒] 以下程度となるため、検索クエリやクエリ応答の転送などの片道の通信に要する時間は、端末の移動にともなうネットワークトポロジの変化の間隔に比べて十分に小さいものと考えられる。そのため、本実験では、検索クエリの転送中およびクエリ応答の返信中のそれぞれにおいて、端末の移動によってネットワークトポロジが変化しないものとした。一方、検索クエリの転送からクエリ応答の返信までの間は、数秒のタイムアウトを設定して動作を制御する必要があるため、その間に端末の移動の影響を受ける可能性がある。そのため、実験では、ある初期トポロジにおいてクエリ発行端末が検索クエリを転送し、各端末が 1 タイムスロット経過した際の移動後のトポロジにおいてクエリ応答を返信するものとした。また、簡単化のため、クエリ発行端末の検索クエリの作成や各端末のスコアの算出など、メッセージの転送以外に要する遅延も十分に小さいとして無視

表 1 パラメータ設定
Table 1 Parameter configuration.

パラメータ	意味	値	値の範囲
N	基準値の数	10	(0~100)
C	データサイズ	10	(1~1,000)
k	要求データ数	100	(1~300)

した。さらに、MAC 層でのパケット衝突や再送は考慮せず、メッセージを転送する端末は無線通信帯域のすべてを使用できるものとした。同様の理由により、トランスポート層も、プロトコルを特定せず、再送も考慮しないものとした。

表 1 に、本実験で用いたパラメータを示す。各パラメータは基本的には定数値をとるが、いくつかのパラメータは括弧内の範囲で変化させた。

以上のシミュレーション環境において、100 回のシミュレーション実験を行ったときの評価値を調べた。なお、 i 回目の実験では端末 M_i が Top-k 検索を行うものとした。また、各試行において、端末の移動による影響を評価するため、ある初期トポロジ T においてクエリ発行端末が検索クエリを転送し、各端末が移動後のトポロジ T' においてクエリ応答を返信するものとした。ここで、全試行において、初期トポロジ T と移動後のトポロジ T' はそれぞれ共通のものを用いた。

- 検索結果の取得精度
ネットワーク全体の上位 k 個のスコアを持つデータのうちクエリ発行端末が獲得したデータ数の割合。
- トラヒック
ネットワーク全体で送信されたメッセージ（検索クエリ、クエリ応答および経路探索メッセージ）のデータの合計サイズの平均。ただし、各メッセージに含まれるデータ以外の要素のサイズを 1、データサイズ C は 4.4 節以外では 10 とした。
- トラヒックの削減率
基準値を用いない場合に比べて、最もトラヒックを削減できていた場合における削減したトラヒックの割合の平均。
- 最適基準値数
基準値を用いない場合に比べて、トラヒックの削減率が最大の場合における基準値の数の平均。

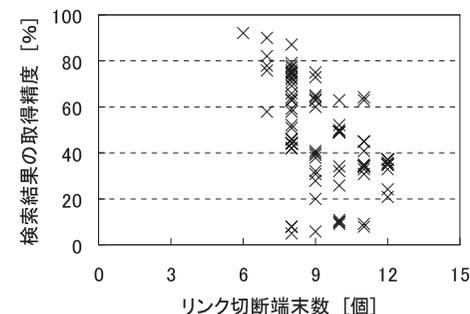


図 4 検索結果の取得精度

Fig. 4 Accuracy of query result.

4.2 リンク切断の処理による性能への影響

リンク切断の処理による性能への影響を調べるため、リンク切断時の処理を行わない場合の検索結果の取得精度を調べた。その結果を図 4 に示す。この図において、プロットは各実験における結果を表し、横軸はリンク切断を検出した端末数、縦軸は検索結果の取得精度を表す。

図 4 の結果より、リンク切断端末数が増加すると、検索結果に入るデータが返信されない可能性が高くなるため、検索結果の取得精度は全体的に減少することが分かる。また、リンク切断端末数が同数であっても、検索結果の取得精度は大きく異なる場合があることが分かる。これは、たとえば、検索クエリ発行端末までのホップ数の小さい端末がクエリ応答を返信しない場合、その端末を根とする部分木に属するすべての端末の持つデータが返信されないなど、リンク切断を検出する端末によって返信されるデータ数が異なるからである。

リンク切断の検出時の処理を用いなければ、平均すると、9.3 個の端末が親とのリンク切断を検出し、検索結果の取得精度が 53% 低下する。この結果から、提案手法におけるリンク切断時の処理が有効であることが分かる。なお、本論文の評価では、ネットワークの分断、端末の離脱およびパケット衝突が発生しない環境を想定しているため、提案手法において、リンク切断の検出時の処理を行うことにより、クエリ発行端末がネットワーク内のすべての端末からクエリ応答を受信でき、検索結果の取得精度はつねに 100% となる。

以降の節では、すべての試行においてリンク切断の検出時の処理を行うことで、検索結果の取得精度はつねに 100% となっている。

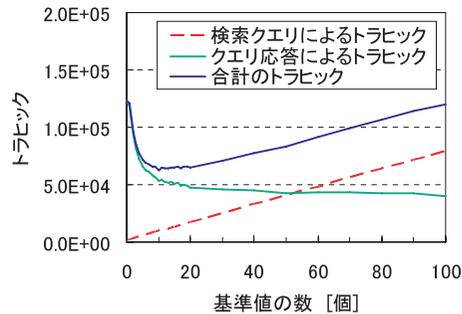


図 5 基準値の数の影響
Fig. 5 Effect of number of standard scores.

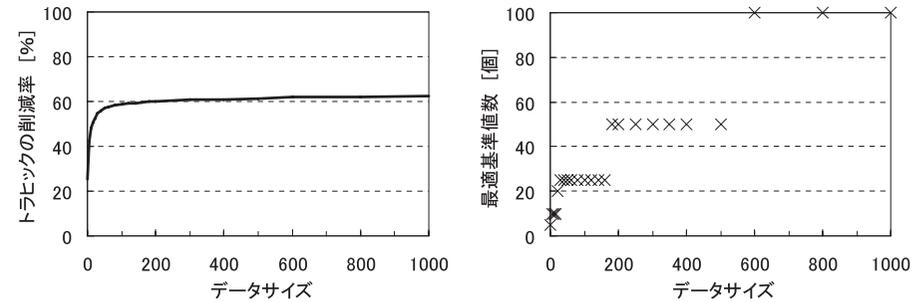
4.3 基準値の数 N の影響

次に、基準値の数 N 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 5 に示す。この図において、横軸は N 、縦軸はトラヒックを表す。ここで、 $N = 0$ の場合は、提案手法において基準値を検索クエリメッセージに添付せず、各端末は受信したデータと自身のデータの中から上位 k 個のスコアを持つデータを返信する場合である。

図 5 の結果から、 N が増加すると、検索クエリメッセージのサイズが大きくなるため、検索クエリによるトラヒックは増加することが分かる。一方、 N が増加すると、検索結果の候補となるデータを効率的に絞り込めるため、クエリ応答によるトラヒックは減少する。また、 N が非常に大きい場合、検索クエリの転送経路上の端末が持つデータの中で、上位 k 個のスコアを持つデータのみが返信されるため、クエリ応答のトラヒックはほぼ一定になる。合計のトラヒックは、 N が増加すると、1 度低くなった後、高くなることが分かる。これは、 N が小さい範囲では、データの絞り込みによるクエリ応答によるトラヒックの減少の影響が大きく、一方、 N が大きい範囲では、検索クエリによるトラヒックの増加の影響が大きいためである。合計のトラヒックは、 $N = 10$ の場合に最小となり、基準値を用いない場合と比べて、49%程度に減少することが分かる。

4.4 データサイズ C の影響

4.3 節の結果から、トラヒックを最も削減できる基準値の数は、端末の持つデータのサイズに依存すると考えられるため、データサイズ C 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 6 に示す。この図において、横軸は C を表す。縦軸は、図 6(a) ではトラヒックの削減率、図 6(b) では最適基準値数を表す。



(a) トラヒックの削減率 (b) 最適基準値数

図 6 データサイズの影響
Fig. 6 Effects of data size.

図 6 の結果から、 C が非常に小さい場合、検索クエリによるトラヒックの影響が大きいため、トラヒックの削減率、最適基準値数ともに小さい ($C = 1$ の場合、トラヒック削減率 25%、最適基準値数 1 個)。 C が大きくなると、データの絞り込みによってクエリ応答によるトラヒックが減少する影響が大きくなるため、トラヒックの削減率、最適基準値数ともに増加する。 C がさらに大きくなると、検索クエリによるトラヒックによる影響が無視できるほど小さくなるため、トラヒックの削減率はほぼ一定 (約 62%) になり、最適基準値数はクエリ応答によるトラヒックを最大限削減できる k になる。

4.5 要求データ数 k の影響

トラヒックを最も削減できる N は、要求データ数 k にも依存すると考えられるため、 k 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。この図において、横軸は k を表す。縦軸は、図 7(a) ではトラヒックの削減率、図 7(b) では最適基準値数を示す。

図 7 の結果から、 k が非常に小さい場合、各端末が返信しなければならないデータ数が少ないため、トラヒックの削減率、最適基準値数ともに小さい ($k = 10$ の場合、トラヒック削減率 30%、最適基準値数 2 個) ことが分かる。 k が大きくなると、データの絞り込みによってクエリ応答によるトラヒックが減少する影響が大きくなるため、トラヒックの削減率、最適基準値数ともに増加する。 k がさらに大きくなると、上位 k 個のスコアを持つデータのみが返信されるため、トラヒックの削減率はほぼ一定 (約 54%) になる。このとき、データの絞り込みに多くの基準値が必要なため、最適基準値数はやや増加する。

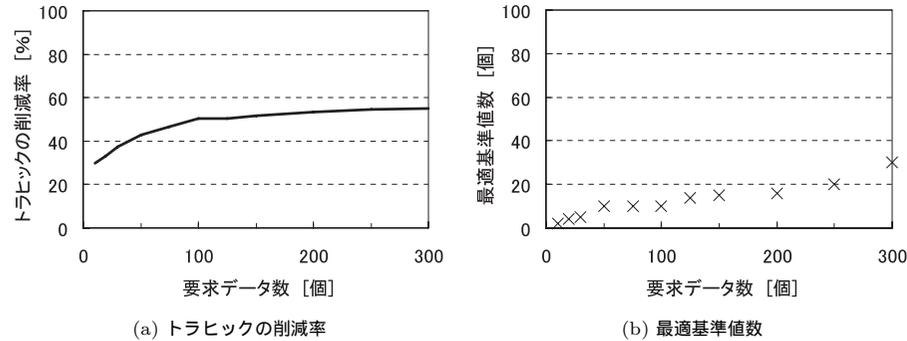


図 7 要求データ数の影響
Fig. 7 Effect of number of requested data.

5. 考 察

本論文では、簡単化のため、ネットワークの分断が発生せず、端末が離脱しない環境を想定した。また、性能評価において、各端末が無線を用いて通信を行う際、パケット衝突が発生しないものとした。しかし、実環境におけるアドホックネットワークでは、これらが頻繁に発生し、提案手法の性能に影響を与えるものと考えられる。そこで、本章では、ネットワークの分断、端末の離脱、およびパケット衝突が提案手法の性能に与える影響について考察する。

5.1 ネットワークの分断による影響

ネットワークの分断が発生する環境では、クエリ発行端末が存在するネットワークから一部の端末が孤立し、それらの端末からのクエリ応答を受信できない。また、孤立した端末が上位 k 個に入るスコアを持つデータを保持している場合、そのデータを検索結果に反映させることができず、検索結果の取得精度は低下してしまう。たとえば、ネットワークが半分に分断した場合、提案手法および比較手法の検索結果の取得精度はともに半分程度まで低下してしまうものと考えられる。

5.2 端末の離脱による影響

5.1 節と同様に、端末がネットワークから離脱する環境では、クエリ発行端末が離脱した端末からのクエリ応答を受信できない。また、離脱した端末が上位 k 個に入るスコアを持つデータを保持している場合、そのデータを検索結果に反映させることができず、検索結果

の取得精度は低下してしまう。

5.3 パケット衝突による影響

パケット衝突が発生する環境では、端末が送信しているメッセージが損失してしまい、検索結果の取得精度が低下する可能性がある。そのため、各端末がパケット衝突によるメッセージ損失を検知した場合には、損失したメッセージを再送する必要がある。ここで、パケット衝突は送信しているメッセージのサイズが大きいくほど発生しやすい。提案手法は、基準値を用いない場合に比べてトラフィックを低く抑えられるため、パケット衝突の発生をできる限り抑えることができると考えられる。したがって、パケット衝突の影響は、比較手法よりも小さくなるものと予測される。

6. おわりに

本論文では、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維持を目的として、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、各端末が検索結果に入るスコアを推測し、検索クエリを発行した端末が $\frac{k}{N}i$ ($1 \leq i \leq N$) 個のデータを取得できることを保証する N 個のスコアを基準値に設定する。各端末は、検索クエリの転送時、およびクエリ応答の返信時に基準値を用いることで、検索結果の候補となるデータを絞り込む。また、端末がクエリ応答の返信時にリンク切断を検出した場合、他の隣接する端末を経由する返信経路を検出する。

シミュレーション実験の結果より、提案手法は基準値を用いることで、検索結果の候補となるデータを絞り込み、クエリ応答の返信によるトラフィックを大幅に削減できることを確認した。また、リンク切断時に返信経路を検出することで、検索結果の取得精度を維持できることを確認した。

本論文では、簡単化のため、端末の移動によるネットワークの分断や端末の離脱が発生しない環境を想定した。ネットワークの分断や端末の離脱が発生する環境では、提案手法において、すべての子からクエリ応答メッセージを受信できず、その子からのクエリ応答メッセージを待ち続ける端末が発生する。そのため、一部の端末がクエリ応答メッセージを検索クエリを発行した端末まで返信できず、検索結果の取得精度が低下する可能性が高くなる。そこで今後は、各端末に一定のタイムアウト時間を設け、その時間になったらすべての子からクエリ応答を受信していなくても、自身のクエリ応答を返信するように処理手順を変更する必要がある。さらに、このような環境における検索結果の取得精度の低下を抑制するために、文献 3)、4) の提案手法を参考に、各端末がデータの複製を配置する手法について検討

する予定である。

また、本論文では、提案手法が実装されるアプリケーション層のみを考慮したシミュレーション実験により、性能評価を行った。しかし、実環境では、その下位層に、トランスポート層や MAC 層のプロトコルが存在し、シミュレーション実験で考慮していない事項が提案手法の性能に大きな影響を与える可能性がある。たとえば、端末が密集する環境などでは、パケット衝突と再送が頻繁に発生することで、データの転送に多大な時間やトラヒックを要してしまうことが考えられる。また、天候や周囲の状況によって十分な転送速度が得られない環境では、検索クエリの転送中およびクエリ応答の返信中に端末の移動によってネットワークポロジが頻繁に変化する可能性も考えられる。これらは提案手法以外の手法にも同等に影響するため、本論文のシミュレーション実験で示した手法間の性能差の相対性にはほとんど影響がないものと考えられるが、各手法の性能自体には大きな影響を与える可能性がある。今後は、これら下位層のプロトコルも考慮し、提案手法を総合的に評価する予定である。

さらに、本論文では、基準値を推定した上位 k 個のスコアの中から均一に抽出した。今後は、スコアの分布などを考慮して基準値を適切に選択する方法について検討する予定である。

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) および平成 20 年度総務省委託研究「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。

参 考 文 献

- 1) Akbarinia, R., Pacitti, E. and Valduriez, P.: 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) Balke, W.T., Nejdil, W., Siberski, W. and Thaden, U.: Progressive Distributed Top-k Retrieval in Peer-to-peer Networks, *Proc. IEEE ICDE'05*, pp.174-185 (2005).
- 3) Hara, T.: Effective Replica Allocation in Ad Hoc Networks for Improving Data Accessibility, *Proc. IEEE INFOCOM'01*, pp.1568-1576 (2001).
- 4) Hara, T. and Madria, S.K.: Data Replication for Improving Data Accessibility in Ad Hoc Networks, *IEEE Trans. Mobile Computing*, Vol.5, No.11, pp.1515-1532 (2006).
- 5) Kalnis, P., Ng, W.S., Ooi, B.C. and Tan, K.L.: Answering similarity queries in

- peer-to-peer networks, *Information Systems*, Vol.31, No.1, pp.57-72 (2006).
- 6) Klemm, A., Lindemann, C. and Waldhorst, O.P.: A Special-Purpose Peer-to-Peer File Sharing System for Mobile Ad Hoc Networks, *IEEE VTC'03*, pp.2758-2763 (2003).
- 7) 松波秀和, 寺田 努, 西尾章治郎: P2P 型コンテンツ検索システムのための効率的な Top-k 検索処理手法, *情報処理学会論文誌*, Vol.47, No.9, pp.2850-2859 (2006).
- 8) Mase, K., Sato, T., Nakano, K., Sengoku, M. and Shinoda, S.: Efficient Flooding Schemes in Mobile Ad Hoc Networks, *Proc. IEEE MDMC'01*, pp.152-157 (2001).
- 9) Pagani, E. and Rossi, G.P.: Providing reliable and fault tolerant broadcast delivery in mobile ad-hoc networks, *Mobile Network and Applications*, Vol.4, No.3, pp.175-192 (1999).
- 10) Silberstein, A., Braynard, R., Ellis, C., Munagala, K. and Yang, J.: A Sampling-Based Approach to Optimizing Top-k Queries in Sensor Networks, *Proc. IEEE ICDE'06*, p.68 (2006).
- 11) Silberstein, A., Munagala, K. and Yang, J.: Energy-Efficient Monitoring of Extreme Values in Sensor Networks, *Proc. ACM SIGMOD*, pp.169-180 (2006).
- 12) 寺田 努, 松波秀和, 西尾章治郎: P2P 型コンテンツ検索システムにおけるコンテンツ分布を考慮した Top-k 検索処理手法, *情報処理学会論文誌: データベース*, Vol.48, No.SIG14(TOD35), pp.1-10 (2007).
- 13) Wu, M., Xu, J., Tang, X. and Lee, W.-C.: Monitoring Top-k Query in Wireless Sensor Networks, *Proc. IEEE ICDE'06*, p.143 (2006).
- 14) Wu, M., Xu, J., Tang, X. and Lee, W.-C.: Top-k Monitoring in Wireless Sensor Networks, *IEEE Trans. Knowledge and Data Engineering*, Vol.19, No.7, pp.962-976 (2006).

(平成 21 年 3 月 30 日受付)

(平成 21 年 10 月 2 日採録)



萩原 亮

2008 年大阪大学工学部電子情報エネルギー工学科卒業。現在、同大学大学院博士前期課程在学中。モバイル環境におけるデータ検索技術に興味を持つ。日本データベース学会の学生会員。



篠原 昌子 (正会員)

2004年大阪大学工学部電子情報エネルギー工学科卒業。2006年同大学大学院情報科学研究科博士前期課程修了。2009年同大学院情報科学研究科博士課程修了。博士(情報科学)。現在、株式会社富士通研究所所属。2009年本学会論文賞受賞。センシングを用いたコンテキスト管理に興味を持つ。



原 隆浩 (正会員)

1995年大阪大学工学部情報システム工学科卒業。1997年同大学大学院工学研究科博士前期課程修了。同年同大学院工学研究科博士後期課程中退後、同大学院工学研究科情報システム工学専攻助手、2002年同大学院情報科学研究科マルチメディア工学専攻助手、2004年より同大学院情報科学研究科マルチメディア工学専攻准教授となり、現在に至る。工学博士。1996年本学会山下記念研究賞受賞。2000年電気通信普及財団テレコムシステム技術賞受賞。2003年本学会研究開発奨励賞受賞。2008年、2009年本学会論文賞受賞。データベースシステム、分散処理に興味を持つ。IEEE、ACM、電子情報通信学会、日本データベース学会の各会員。



西尾章治郎 (フェロー)

1975年京都大学工学部数理工学科卒業。1980年同大学大学院工学研究科博士後期課程修了。工学博士。京都大学工学部助手、大阪大学基礎工学部および情報処理教育センター助教授、大阪大学大学院工学研究科情報システム工学専攻教授を経て、2002年より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教授となり、現在に至る。2000年より大阪大学サイバーメディアセンター長、2003年より大阪大学大学院情報科学研究科長、その後2007年より大阪大学理事・副学長に就任。この間、カナダ・ウォータールー大学、ビクトリア大学客員。データベース、マルチメディアシステムの研究に従事。現在、Data & Knowledge Engineering等の論文誌編集委員。本会理事を歴任。本会論文賞を受賞。電子情報通信学会フェローを含め、ACM、IEEE等8学会の各会員。