

アドホックネットワークにおけるデータの複製を考慮した k 最近傍データ検索手法

駒井 友香¹ 佐々木 勇和¹ 原 隆浩¹ 西尾 章治郎¹

概要: k 最近傍検索は, 指定された位置 (クエリポイント) に対して, その位置から距離の近い上位 k 個の位置依存データ (k 最近傍データ) を取得する検索方法である. 本稿ではアドホックネットワークにおいて, データの複製を用いて, トラフィックおよび検索時間の低減を目的とする k 最近傍データ検索手法を提案する. 提案手法では, クエリポイントをクエリを発行する端末の位置と想定するため, 移動時に周辺端末とメッセージの交換を行い, 複製を近傍データに更新する. 検索時には, 検索範囲内の各端末が, 自身の保持するデータの複製に関連する地理的範囲と検索範囲の重複が大きいほど早く返信を行い, 返信を傍受した端末は, 返信済みの範囲のデータを返信しない. これにより, できるだけ少ない端末による重複のないデータの返信を実現する. シミュレーション実験より, 提案手法は取得精度を維持しつつ, トラフィックおよび検索時間を低減していることを確認した.

Processing k Nearest Neighbor Queries with Cached Location-dependent Data in Ad Hoc Networks

YUKA KOMAI¹ YUYA SASAKI¹ TAKAHIRO HARA¹ SHOJIRO NISHIO¹

1. はじめに

近年, 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は既存の通信基盤を必要とせず, 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, 災害時の救助活動やイベント会場での情報共有, およびセンサネットワークでの情報収集などへの応用が期待されている. 広い領域において多数の端末がアドホックネットワークを構成している状況では, 周辺情報の収集等, 位置に依存したデータを検索するような位置依存サービスが求められている. 位置依存情報の検索方法として, ユーザが指定した特定の位置に関連したデータを取得する, k 最近傍 (k -Nearest Neighbor, k NN) 検索がある. k 最近傍検索では, 検索要求 (クエリ) 発行端末が指定した位置 (クエリポイント) に対して, その位置からデータの発生位置までの距離が近い上位 k 個の位置依存データ (k 最近傍データ) を検索することにより, 必要なデータを効

率的に取得することができる.

筆者らの研究グループはこれまでに, アドホックネットワークにおいて小さいトラフィックで k 最近傍データを検索する手法を提案した [2]. この手法では, 端末が自身の保持するデータの発生位置から一定の距離 (保証距離) 以上離れた場合, そのデータをデータの発生位置に最も近い端末へ転送する. これにより, データを発生位置付近に留めておくことができる. 検索の際には, 保証距離を考慮して k 最近傍データが存在する範囲 (検索範囲) を推定し, 範囲内の端末の持つデータを検索することで k 最近傍データを取得する. この手法により, 一箇所 (一つの端末) に全データを集めることや, ネットワーク全体へクエリを転送する必要なく, 位置依存データに対して k 最近傍検索を行うことができる.

ここで, アドホックネットワークにおけるデータ検索では, データの複製を利用することが有用である. 各端末がデータの複製を効果的に保持することにより, 狭い範囲の検索においても多くのデータを得ることができるため, 検索時の検索範囲を小さくすることができる. これによりト

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
Yamadaoka 1-5, Suita, Osaka 565-0871, Japan

ラヒックおよび検索時間をより小さくすることができると考えられる。特に、 k 最近傍検索では自身の現在位置をクエリポイントとして指定する場合が多いと考えられるため、各端末が自身の周辺に存在するデータの複製を保持することにより、効率的な検索を行うことができる。

ここで、アドホックネットワークでは端末が移動するため、自身の保持しているデータの複製が、自身の位置から最も近い k 個のデータである保証がない。また、必要なデータ数が自身の保持するデータ数を上回る場合には、周辺の端末へデータの間合わせを行い、必要なデータを取得する必要がある。このとき自身の周辺の端末は、それぞれの位置をクエリポイントとする k 最近傍データの複製を保持しているため、重複したデータを保持している可能性が高い。重複したデータを収集してしまうと無駄なトラヒックが大きく発生する。また、各端末が保持するデータを一元的に管理しておらず、必要なデータを全て取得できているかを保証することが困難であるため、必要なデータを全て取得できているかをクエリ発行端末が自律的に判断する仕組みが必要である。

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作成する環境を想定し、トラヒックおよび検索時間の低減を目的とする k 最近傍データ検索手法である FA (Filling Area) 法を提案する。FA 法では、オリジナルデータはこれまでに提案した手法 [2] 同様、発生位置付近に留め、さらに各端末は、自身の位置をクエリポイントとした近傍データの複製を保持する。端末は一定の距離以上移動した場合、周辺端末とメッセージの交換を行い、自身の現在地点に関連するデータに複製を更新する。 k 最近傍データの検索時には、 k 最近傍データの複製を保持している端末が存在すると考えられる範囲を検索範囲として、その範囲内の端末へクエリをフラッディングする。検索範囲内の各端末は自身が保持するデータの複製をクエリ発行端末へ返信する。このとき、自身の保持するデータの複製が関連する地理的範囲 (担当領域) と検索範囲の重複が大きいほど早く返信を行い、返信を傍受した端末は、自身の担当領域のうち傍受した返信により包含された領域を自身の担当領域から除外する。これにより、できるだけ少ない端末による重複のないデータの返信を実現する。さらに、検索範囲を網羅的に検索できていることを保証するため、他の端末からの返信により包含されていない担当領域がある場合には、その範囲にデータがない場合にも、クエリ発行端末へ通知を行う。これにより、クエリ発行端末は、検索範囲を網羅的に検索できていることを保証することができる。シミュレーション実験により、提案手法はデータの複製を用いることにより、取得精度を維持しつつ、トラヒックおよび検索時間を低減できていることを確認した。

以下では、2. で関連研究について述べる。次に、3. で提案手法について述べ、4. でシミュレーション実験の結果を示す。

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

2.1 位置依存サービスにおけるデータの取り扱い

文献 [6] では、アドホックネットワークにおいて位置依存データの共有を行うための複製配置手法を提案している。この方法では、効果的に複製を配置することで、データへの高いアクセス成功率を達成している。この研究では、特定の位置に関する一つのデータへのアクセスを想定しているが、本研究では特定の位置に近い複数のデータを取得することを目的としている。文献 [7] では、センサネットワークにおいてインフラを用いることなく位置依存サービスを提供するシステムを提案している。このシステムでは、エージェントが端末間を移動し、特定のエリア内にエージェントを留めることを目的としている。本稿では、データを発生位置に留める手法として、この方法を参考にしている。

2.2 k 最近傍検索

文献 [5] では、クエリ発行端末が移動する環境において、クエリ発行端末周辺の k 最近傍物体の検索手法を提案している。この手法では、サーバに全てのデータが収集できていることを前提としており、本研究とは想定が異なる。文献 [4] では、モバイル P2P 環境における k 最近傍検索手法を提案している。この手法では、端末は協調してクエリを処理し、分散処理による通信コストの低減を実現して k 最近傍検索を行っている。この文献により扱っている問題は本稿と似ているが、この文献では k が端末の保持できる複製のデータ数より小さい場合のみを扱っており、本稿とは想定が異なる。文献 [8] では、モバイル P2P 環境において、近隣端末と協調して範囲検索および k 最近傍検索を処理する手法を提案している。この手法では、R 木を基とした木構造である SAT を提案し、隣接端末の保持するデータの情報を管理することにより、必要な端末のみにクエリを送信することで、小さなトラヒックによる検索を実現している。しかし、この研究では、各端末が自身の隣接端末を完全に把握している状況を想定しており、提案手法ではこれを必要としない点で想定が異なる。

3. 提案手法 : FA 法

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、位置依存データに対して k 最近傍検索を行う環境を想定する。位置依存データは、位置情報を保持しており、データはその位置に関連しているものとする。具体的には、データを生成した端末のその時点での位置をデータの関連位置とする。 k 最近傍検索を行う端末は、自身の位置をクエリポイントとして指定した検索クエリを発行し、ネットワー

ク内の位置依存データの中から、データの関連位置がクエリポイントから近い上位 k 個のデータを取得する。ネットワーク内には m 個のデータが存在し、一意の識別子が付与されているものとする。各データに更新はないものとする。データの発生位置に偏りはしないものとし、簡単化のため、すべてのデータのサイズは等しいものとする。ネットワーク内には、同等の性能を持つ n 台の端末（識別子： M_1, M_2, \dots, M_n ）が存在し、各々が自由に移動する。ネットワークに参加する端末は既知であり、ネットワークを構築する前に各端末に一意の識別子が付与されているものとする。各端末は、GPS などにより自身の位置情報を把握しているものとする。各端末には、データ C 個分の複製を配置可能な記憶領域があるものとする。

3.2 データの管理

ユーザが自身の位置周辺の k 最近傍データを繰り返し検索する場合は多いと考えられる。その場合、ユーザの持つ各端末にそれぞれ自身の周辺の位置依存データの複製を保持することが有効である。データの複製を保持した場合、検索を行う端末は、保持できるデータの複製数 C より k が小さい場合には、クエリを発行することなく、自身の保持するデータの複製より k 最近傍データを取得することができる。一方、保持できるデータの複製数 C より k が大きい場合には、周辺の端末もそれぞれの周辺のデータの複製を保持しているため、限られた端末のみを検索することにより k 最近傍データを取得することができる。

アドホックネットワークでは端末は移動するため、端末の位置が変化し、それに伴い自身の周辺のデータも変化する。現在の自身の周辺のデータの複製を保持し続けることは大きな課題である。端末が保持しているデータの複製を現在の位置に関連するデータに更新するため、保持しているデータもしくはその情報を頻繁に交換すると非常に大きなトラフィックが発生し、パケットロスの原因となる。一方、自身の保持するデータの複製を更新しなければ、検索時に必要なデータを取得できない可能性がある。

FA 法では、初期動作として、各端末が保持するデータの複製を配置するため、データ生成端末が周辺の端末にデータを周知する必要がある。そのため、データ生成端末がデータの複製を保持すべき端末の存在する範囲にデータを含む周知メッセージをフラッディングする。この範囲は、データの複製を保持すべき端末が存在する範囲を包含し、かつ無駄な範囲へメッセージの送信が行われないよう適切に設定される必要がある。

3.2.1 オリジナルデータの管理

データを保持している端末がデータの発生位置を離れると、効率的に k 最近傍データを検索することができない。データをその発生位置付近に留めることで、クエリポイント周辺を検索するだけで、効率的に k 最近傍データを取得

できる。そのため、提案手法では、文献 [2] 同様、データ保持端末がデータの発生位置から閾値 d （保証距離）以上離れた場合、ジオルーティングにより、発生位置に最も近い端末までデータを伝搬する。

具体的には、まず、送信元端末（データ保持端末）が隣接端末検索メッセージをブロードキャストし、これを受信した端末が送信元端末より発生位置に近い場合、返信待ち時間を設定する。この返信待ち時間は、発生位置に近ければ近いほど短く設定されるため、発生位置に最も近い端末が最も早く送信元端末へ返信を行う。返信を受け取った送信元端末は、返信元端末へデータを送信する。これを繰り返し、発生位置に最も近い端末までデータを伝搬する。

このジオルーティング手法により、1本の経路で転送が行えるため、少ないトラフィックで発生位置に最も近い端末までメッセージを伝搬することが可能である。

3.2.2 複製データの管理

各端末は、 C 個の複製を保持できるため、自身の現在の位置に近い C 個のデータの複製を保持する。そのために、各端末は周辺端末とメッセージを交換して自身の保持するデータの複製を更新する。各端末は、自身が以前にデータの複製を更新した地点（更新点）をクエリポイントとする C 最近傍データを保持している。ここで、本稿において、更新点を中心とし、更新点から最も遠いデータ（ C 番目のデータ）の発生位置までの距離を半径とする円範囲のことを担当領域と定義する。

端末が更新点から d 以上離れた場合、複製の更新を開始する。具体的には、隣接端末へ更新要求メッセージを送信する。このメッセージにはすでに複製として保持しているデータの識別子および現在の端末の位置が含まれる。更新要求メッセージを受信した端末は、メッセージに含まれている端末の位置と自身の更新点の距離を基に以下の式により返信のための待ち時間を決定する。

$$waiting\ time = \alpha \cdot dist. \quad (1)$$

$dist$ は、更新要求メッセージに含まれる端末の位置と受信端末の更新点の距離であり、 α はパラメータである。この式により、更新要求メッセージを送信した端末の位置に更新点に近い端末ほど早く返信を行う。これは、更新点に近い端末ほど更新要求メッセージを送信した端末が保持すべきデータの複製を多く保持している可能性が高いためである。そのため、最も小さな待ち時間を設定した端末は、待ち時間経過後、自身の保持している複製より、更新要求メッセージを送信した端末の現在位置から発生位置に近い C 番目のデータに含まれるデータのうち保持されていないデータを返信する。この時、他の端末は返信を傍受し、すでに返信されているデータを重複して返信しない。これにより、無駄なトラフィックを抑制することができる。

3.3 k 最近傍データの検索手順

FA 法では、クエリ発行端末が自身を中心とした特定の円範囲（検索範囲）にクエリを送信し、必要なデータを取得する。検索範囲内の端末は、 k 最近傍データに含まれるデータを返信する。ここで、各端末は他の端末からの返信を傍受してすでに返信されているデータを重複して返信しないようにする。アルゴリズム 1 に、クエリ処理の疑似コードを示す。ここで、 $Assignment(node)$ とは、括弧内に指定した端末の担当領域を表す。

FA 法では、まず、検索を行う端末が k を指定し、自身の保持する複製により必要なデータを取得できるかを確認する（アルゴリズム 1, 2 行目）。 k が保持できるデータの複製数 C より小さい場合、自身の保持しているデータのうち端末の現在位置（クエリポイント）に近い k 個のデータが「保証範囲」以内の場合、これを k 最近傍データとしてクエリを発行することなく検索を終えることができる。ここで、保証範囲とは、中心を端末自身の位置とし、半径を（クエリ発行端末の担当領域の半径）－（クエリ発行端末の現在位置と更新点の距離）とする円範囲である（図 1 参照）。発生位置が保証範囲内にある場合、自身の担当領域の外には発生位置がより近いデータはないことを保証できる。そのため、保持しているデータのうち、この範囲内に k 個以上のデータが存在する場合、最も近い k 個のデータが k 最近傍データであることを保証できる。一方、 k が C より大きい場合、もしくは、自身のもつクエリポイントに近い k 個のデータが保証範囲内でない場合、周辺端末への問い合わせにより k 最近傍データを取得する。

クエリ発行端末は、検索範囲の半径 R を自身の保持しているデータの複製および k を基に以下の式を用いて決定する。

$$R = \beta \cdot l \cdot \sqrt{\frac{k}{C}} - \gamma \cdot l + d. \quad (2)$$

ここで、 l はクエリ発行端末の担当領域の半径、 β はデータの密度の見積りの誤りを防ぐためのパラメータ、および γ は検索範囲を縮小するためのパラメータである。まず、第一項により、 R はクエリ発行端末の保持するデータの複製より求めたデータの密度を基に決定される（ $\pi l^2 \cdot C = \pi x^2 \cdot k$, x は必要なデータの発生位置が存在する範囲の円半径）。さらに、各端末はそれぞれの更新点から l 程度発生位置が離れたデータの複製を保持しているため、 l だけ検索範囲の半径を小さく設定しても十分必要なデータを取得することができる（第二項）。このとき、データを保持している円範囲の半径がすべての端末において l であるならば γ は 1 とすればよいが、各端末ごとに違うため、精度を保つために基本的に 1 より小さな値に設定する。また、各端末は最大で d だけ更新点から離れている場合があるため、 R は d だけ大きく設定する必要がある（第三項）。

クエリ発行端末は、 R を設定した後、検索範囲内の端末

Algorithm 1 提案手法におけるクエリ処理の疑似コード

```

1: // 検索開始
2: if Assignment(query-issuing node) guarantees  $k$ NNs then
3:   complete search
4: else
5:   calculate search area
6:   send Query
7:   set WT
8: end if
9: // 初めてクエリ受信時
10: if within search area then
11:   store ID of parent
12:   store Assignment(query-issuing node)
13:   send Query
14:   set RD
15: end if
16: // 返信受信時
17: store Assignment(replyNode)
18: update RD
19: if parent of source node then
20:   if query-issuing node then
21:     store data items
22:   else
23:     send Reply to parent
24:   end if
25: end if
26: // RD 経過後
27: if Assignment(Node) is not covered then
28:   if no data within Assignment(Node) then
29:     send EmptyReply to parent
30:   else
31:     send Reply to parent
32:   end if
33: end if
34: // WT 経過後
35: if  $k$  data items are acquired and area of  $k$ NNs is covered then
36:   complete search
37: else
38:   calculate new search area
39:   send Query
40:   set WT
41: end if

```

へクエリをフラグディングする。クエリには、クエリ発行端末の識別子、残り要求データ数 k' (k - 保証範囲内にあるデータ数)、 R 、クエリ発行端末の担当領域（更新点と半径）、クエリポイント（クエリ発行端末の位置）が含まれる。この時、返信を持つための待ち時間 WT を設定する。 WT は、 $\delta \cdot [R/l]$ で決定し、 δ は事前に設定する正の整数である（アルゴリズム 1, 7 行目）。

クエリを初めて受信した検索範囲内の端末は、自身の担当領域に応じて以下の式により返信遅延時間 RD を設定する。

$$RD = \delta \cdot lap \cdot \frac{maxArea - replyArea}{maxArea}. \quad (3)$$

$$lap = \begin{cases} 1 & (distance < l). \\ [distance/l] & (otherwise). \end{cases} \quad (4)$$

ここで、 $replyArea$ は、「受信端末の担当領域 \cap クエリ発行端末の担当領域 $\cap \forall$ 返信端末の担当領域」で表される領域であり、返信を傍受することができた他の端末からの返信により包含されておらず、自身のみが包含している返信範囲である。 $maxArea$ は事前に設定される正の

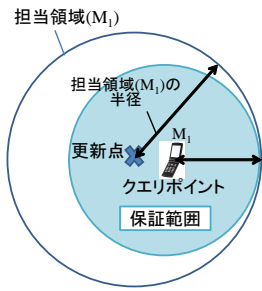


図 1 保証範囲

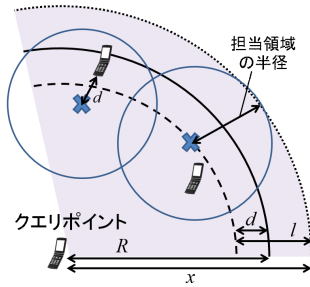


図 2 検索範囲

整数, $distance$ はクエリポイントから受信端末までの距離である. この式により, 受信端末の返信範囲が大きく, 受信端末の位置がクエリポイントに近いほど, RD が小さく設定され早く返信を行う. 返信範囲が大きいと, 返信するべきデータをより多く持っている可能性が高く, これらをまとめて返信することができるため, 返信範囲ができるだけ大きな端末から返信を行うためである. これにより, 返信を行う端末数を抑制することができる. また, lap を用いることでクエリポイントまでの距離が近い端末から返信を行うため, 同じデータを保持している端末のうち, より少ないホップ数でクエリ発行端末まで返信を行うことができる.

RD 経過後, 検索範囲内の端末は返信を行う. 返信メッセージには, 返信データ, および自身の担当領域 (更新点と半径) が含まれている. ここで, 返信範囲が他の端末の返信により含まれていない ($replyArea \neq 0$) 場合のみ, 返信を行う (アルゴリズム 1, 27 行目). 返信データは, 傍受した他の端末により返信されておらず, 発生位置が図 2 の色つき部分 ($Filling Area$) に存在する k' 個以内のデータである. 自身の返信範囲が包含されていないが, 自身の返信範囲内に返信すべきデータが存在しない場合, 返信メッセージではなく, 包含通知メッセージを送信する. 包含通知メッセージには, 自身の担当領域のみが含まれる. 包含通知メッセージを受信したクエリ発行端末は, 包含通知メッセージに含まれる領域内には, データが存在しないことを把握することができる. このメッセージにより, $Filling Area$ を網羅的に検索できていることを保証することができる. 返信メッセージ, もしくは包含通知メッセージを受信 (もしくは傍受) した端末は, 重複してデータの返信を行わないようにするため, 返信により包含された範囲 (返信を行った端末の担当領域) を記録し, 自身の返信範囲と重なる場合, 重なった部分を除くように返信範囲を更新し, それに伴い, RD の更新も行う.

返信メッセージを受信したクエリ発行端末は, k 最近傍データを更新し, メッセージ内の担当領域を記録する. WT 経過後, k 個以上のデータを取得し, かつ $Filling Area$ をすべて検索できている場合, 検索を終了し, 保持している複製を k 最近傍データにより更新する. これに当てはまら

ない場合, 再度 R を設定し (β をより小さく設定する等), 再度検索を行う.

図 3 を用いて M_1 がクエリを発行した場合のクエリ処理例について説明する. クエリを受信した M_2, M_3, M_4 , および M_5 は, クエリ発行端末 M_1 がデータを保持している範囲を除いた自身のデータを保持する範囲およびクエリポイントまでの距離からそれぞれ応答返信待ち時間を設ける. まず, クエリポイントから近く自身のデータを保持する範囲の広い M_2 が親端末である M_1 へクエリ応答を送信し (図 3(a)), 灰色部分のデータを返信データとしてクエリ応答に添付する. このメッセージを受信した M_1 は取得データを保存する. また, このメッセージを傍受した端末 M_3, M_4 , および M_5 は, M_2 の保持するデータの範囲を記録し, それぞれの応答返信待ち時間を更新する. 次に, M_3 は自身の応答返信待ち時間の経過後, クエリ応答を開始する. このメッセージを傍受した M_4, M_5 は, M_3 の保持するデータの範囲を記録し, それぞれの応答返信待ち時間を更新する. 同様に, M_5 は自身の応答返信待ち時間の経過後, クエリ応答を開始し, メッセージを傍受した M_4 は, M_5 の保持するデータの範囲を記録し, 応答返信待ち時間を更新する (図 3(b)). 応答返信待ち時間が経過した M_4 は, 自身の保持するデータの範囲が返信を傍受した端末によって包含されているため, 返信は中止する (図 3(c)).

FA 法では, 検索範囲を設定し, 各端末が他の端末からの返信を傍受することにより, 無駄な範囲へのクエリ送信や, データの重複した返信を抑制することができる.

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. 本実験では, ネットワークシミュレータ Qualnet5.2 を用いた*1.

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 500 台の端末 (M_1, \dots, M_{500}) が存在する. 各端末はランダムウェイポイント [1] に従い, 0.1~0.5[m/秒] の速度で移動し, 停止時間は 200[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. また, 各端末は自身の位置を常に把握しているものとした.

領域内にデータは 500 個存在するものとし, データのサイズは 128[B] とする. 各端末は C 個のデータを複製として配置することができるとする. クエリポイントは, クエリ発行端末の現在位置とした. d , 式 (2) 中の α, β, γ および式 (3) 中の δ は, 事前実験を基にそれぞれ 50, 1, 1.1, および 0.9 と設定した.

*1 Scalable Network Technologies: "Qualnet," <http://www.scalable-networks.com/>.

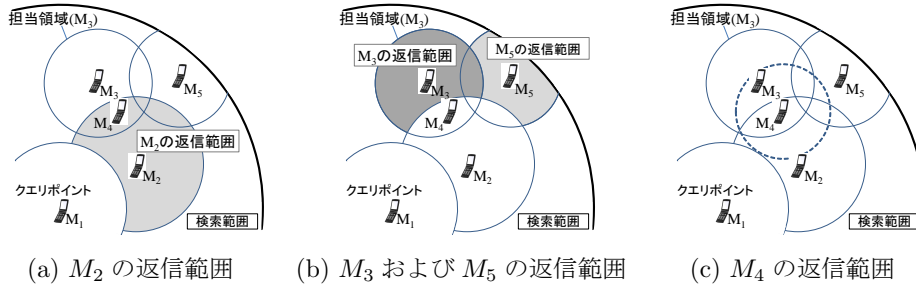


図 3 FA 法の例

表 1 パラメータ設定

パラメータ	意味	値
C	保持複製数	20 (10~50)
k	要求データ数	40 (1~60)

以下の二手法を比較手法として用いた。これら二手法では、オリジナルデータのみ FA 法と同じ方法によりそれぞれのデータの発生位置付近の端末に留めておく。まず、1つ目は、複製を用いない手法である EXP 法 [2] であり、各端末がネットワーク内のオリジナルデータ数 m 、および端末が存在する領域の大きさを把握しているものとし、検索範囲の半径 R を次式のように設定する。

$$R = \sqrt{\frac{k \cdot \text{area}}{\pi \cdot m}} + d. \quad (5)$$

ここで、 area は領域の大きさ ($\text{area} = 1,000 \times 1,000$) である。クエリ発行端末は、検索範囲内の端末へクエリをフラディングし、これを受信した端末は、自身の保持しているデータのうち、クエリポイントを中心とし、 $(R - d)$ を半径とする円範囲内にデータの発生位置が存在するデータのみを葉端末から集約しつつクエリ発行端末へ返信する。グラフ中において、この手法を「EXP」と表記する。2つ目の方法は、FA 法において、保持データの更新を行わない手法であり、グラフ中では「FA⁻」と記述する。この方法では、3.2 節で示した複製データの更新を除き、クエリ処理およびオリジナルデータの管理は FA 法と同様に行う。この方法と比較を行うことにより、FA 法における保持データの更新の効果を示す。FA 法を含めた全ての手法で検索範囲の設定および検索範囲内の検索は一度のみとし、 k 個のデータを取得できない場合も再検索は行わない。

本実験のパラメータを表 1 に示す。各パラメータは、その影響を調べる際に、括弧内の値で変化させた。

FA 法および FA⁻ 法では、まず各端末が保持しているデータを周辺端末へ配布し、これを受信した端末が自身の位置から近い C 個のデータを保持データとして記録する。その後、1,000 秒経過後から、0.5 秒ごとにランダムに選択した端末が自身の位置をクエリポイントとした k 最近傍検索クエリを発行した。そして、1,000 回のクエリが発行された時点で以下の評価値を調べた。

- トラヒック

k 最近傍検索クエリを処理するために送信されたメッ

表 2 メッセージの種類とサイズ

メッセージの種類 (手法)	サイズ [B]
隣接端末検索 (データ)	48
返信 (データ)	16
データ送信 (データ)	40+128
データ通知 (データ)	32+128
データ要求 (データ)	32
データ返答 (データ)	24+128· q
クエリ (FA および FA ⁻)	104
クエリ (EXP)	64
クエリ応答 (FA および FA ⁻)	64+128· q
クエリ応答 (EXP)	32+128· q
包含通知 (FA および FA ⁻)	16
返信受信 Ack	16

ッセージの平均バイト数 (1 クエリ当たりの平均)。表 2 は提案手法、および比較手法における全てのメッセージのサイズを示す。この表において、 q はそれぞれのメッセージに含まれるデータ数である。ここで、(データ) は、3.2 節で述べたデータ管理のためのメッセージである。

- 検索時間

クエリ発行端末が検索を開始してから検索結果を取得するまでの平均時間 (1 クエリ当たりの平均)。

- 取得精度

順位付き検索結果の性能を測る MAP (Mean Average Precision) の値を取得精度とする [3]。MAP は、各クエリの平均精度 AP (Average Precision) を平均化したものである。AP は以下の式で求める。

$$AP_i = \frac{1}{k} \sum_{j=1}^k \left(\frac{g}{j} \cdot e \right), \quad (6)$$

AP_i は i 番目のクエリの平均精度であり、 g は、取得した解の上位 j 個のうち正解集合の j 位以内である解の個数である。また、 e は j 番目の解が正解集合に含まれる場合に 1、含まれない場合に 0 とする。MAP は、 AP_i ($i = 1, \dots, 1,000$) の平均である。MAP は、クエリポイントに近い上位 k 個のデータのうち、より上位のデータの情報を取得できているほど、高い値となる。

4.2 保持複製数 C の影響

FA 法では、端末が保持できるデータの複製数によって

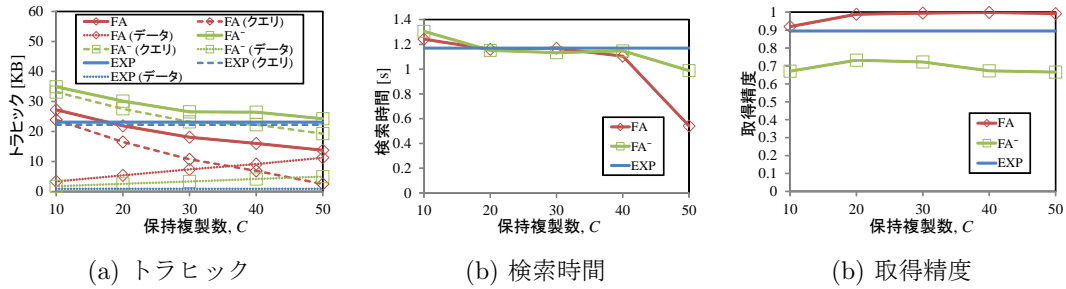


図 4 保持複製数 C の影響

性能が左右されるため、保持複製数 C の影響について調べる。端末の保持複製数 C を変化させた場合の結果を図 4 に示す。これらのグラフにおいて、横軸は保持複製数 C を表し、縦軸は、それぞれ図 4(a) はトラフィック、図 4(b) は検索時間、および図 4(c) は取得精度を表す。図中の判例で、「手法名 (クエリ)」および「手法名 (データ)」は、それぞれ手法ごとのクエリ処理のために発生したトラフィック、および 3.2 節で記述したオリジナルデータおよび保持データの管理のために発生したトラフィックを示す。全ての「手法名 (データ)」には、オリジナルデータの管理が含まれており、これに加えて「FA⁺ (データ)」には、データ発生後のデータの各端末への配布のためのトラフィック、さらに、「FA (データ)」には、データ発生後のデータの各端末への配布のためのトラフィックおよび保持データの更新のためのトラフィックが含まれる。

図 4(a) より、FA 法では、 C が大きくなるほどデータ管理のためのトラフィックが大きくなっている一方、クエリ処理のためのトラフィックが小さくなっている。これは、保持複製数が増加し、更新のために送信されるデータが増加する一方、 k 最近傍データのうちクエリ発行端末の保持するデータが多くなり、クエリ処理中に他の端末から返信されるデータ数が少なくなるためである。データの更新を行わない FA 法では、クエリ処理のためのトラフィックが、FA 法よりも大きくなっている。データの更新を行わない場合、各端末の保持するデータは、自身の現在の位置と関係がなくなっているため、隣接端末ではない遠くの端末が同様のデータを保持している場合が多い。そのため、他の端末の返信を傍受して、すでに他の端末から返信されたデータの返信を行わない仕組みがうまく機能せず、同じデータが重複して返信されてしまう。

図 4(b) より、FA 法では、保持複製数が 50 の場合、検索時間が小さくなっている。これは、クエリ発行端末の保持している複製により全ての必要なデータを取得できる場合が多いためである。一方、保持複製数が 40 以下の場合、EXP 法と同等の検索時間となっている。これは、FA 法では、検索時間が δ および lap に依存するが、 δ を小さく設定すると、返信の待ち時間に差がつかず、データを重複し

て返信するケースが多く発生するため、 δ をある程度大きな値にする必要があることに起因する。ただし、図 4(c) からわかるとおり、EXP 法は、検索時間は FA 法と同等であるが、取得精度は大幅に低い。また、データの更新を行わない FA 法では、保持複製数が 50 の場合、FA 法と比べて検索時間が大きい。これは、クエリ発行端末が保持している複製により結果を取得できない場合が多いためである。

図 4(c) より、FA 法では、取得精度が非常に高いことがわかる。これは、データの更新が適切に機能しており、クエリ発行端末へ返信されるデータが少ないためである。ここで、 C が 10 の場合に取得精度が低下しているが、これは、式 (2) における l が非常に小さくなり、その結果として R が小さく設定されるため、クエリ発行端末で全ての k 最近傍データを取得できない場合があるからである。この問題は、適切なパラメータの設定により、改善できると考えられる。データの更新を行わない FA 法では、FA 法に比べて取得精度が低下している。これは、更新点から移動したにもかかわらず複製配置の更新を行わないため、各端末の保持している複製が現在の端末の位置に対応しておらず、必要なデータを保持する端末が検索範囲内に存在しない場合が多いからである。一方、EXP 法では、FA 法に比べて取得精度が低い。これは、検索範囲内の全ての端末が返信を行うことにより、パケットの衝突が頻繁に起こり、必要なデータを取得できないためである。

4.3 要求端末数 k の影響

要求端末数 k を変化させた場合の結果を図 5 に示す。これらのグラフにおいて、横軸は要求端末数 k を表し、縦軸は、それぞれ図 5(a) はトラフィック、図 5(b) は検索時間、および図 5(c) は取得精度を表す。

図 5(a) より、 k が大きくなるほど、すべての手法でトラフィックが大きくなるのがわかる。これは、クエリ処理における検索範囲が大きくなり、返信されるデータ数が多くなるためである。FA 法では k が C より小さい場合に、クエリ処理のためのトラフィックがほとんど 0 となっていることがわかる。これは、クエリ発行端末が保持しているデータが k 最近傍データを全て含んでいるためである。 k が小さい場合、FA 法ではクエリ処理のためのトラフィックに比べてデータ管理のためのトラフィックが比較的大きくなって

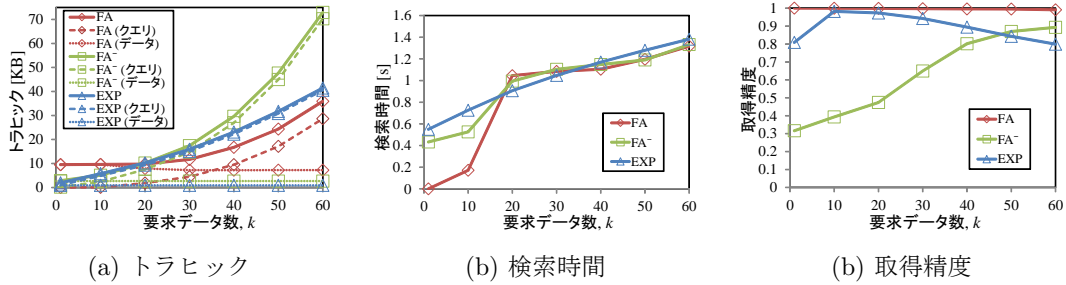


図 5 要求端末数 k の影響

いる。そのため、 k が大きい等、検索範囲がより広くなる場合に、データの複製を保持し、管理するための効果が大いといえる。一方、データの更新を行わない FA 法では、 k が大きくなるにつれてトラフィックが非常に大きくなっている。これは、検索範囲が大きくなるにつれて、同じデータを重複して返信する機会がより頻繁に発生するためである。

図 5(b) より、FA 法では、 k が C より小さい場合に、クエリ発行端末の保持している複製のみで結果を取得できる場合が多いため、検索時間が小さくなっている。一方、 k が C 以上の場合、検索時間は lap に依存し、 k が 20 ($k=C$) のとき、返信データ数 k' は小さいが、検索時間を低減できていないことがわかる。このことから、クエリ発行端末の保持するデータ (複製) により、検索範囲がほぼ被覆されているような状況では、FA 法はあまり適していないといえる。

図 5(c) より、FA 法では、図 4(c) 同様、高い取得精度を達成できていることがわかる。データの更新を行わない FA 法では、 k が大きくなるほど取得精度が高くなっている。これは、 k が大きくなるほど検索範囲が大きくなり、検索範囲内の端末が k 最近傍データを偶然保持している場合が増加するためである。一方、EXP 法の取得精度は k が大きくなるほど低下するが、これは、多くの返信により衝突によるパケットロスが頻繁に発生するためである。 k が 1 の場合、EXP 法の取得精度が低くなっているが、これは、検索範囲が非常に小さく設定され、検索範囲内で発生したデータが存在しない場合が発生するためである。

5. 結論

本稿では、アドホックネットワークにおいてデータの複製を作成する環境を想定し、トラフィックおよび検索時間の低減を目的とする k 最近傍データ検索手法である FA 法を提案した。FA 法では、オリジナルデータはこれまでに提案した手法同様、発生位置付近に留め、さらに各端末は、自身の位置をクエリポイントとした近傍データの複製を保持する。検索時には、検索範囲内へクエリをフラッシングし、検索範囲内の各端末が自身の保持するデータの複製をクエリ発行端末へ返信する。このとき、自身の保持する

データの複製に関連する地理的範囲 (担当領域) が大きいほど早く返信を行い、返信を傍受した端末が、返信済みのデータを返信しないことにより、できるだけ少ない端末による重複のないデータの返信を実現する。

シミュレーション実験により、提案手法はデータの複製を適切に管理することにより、取得精度を維持しつつ、トラフィックおよび検索時間を低減できていることを確認した。今後は、検索時間をさらに短縮するために、提案手法を拡張する予定である。

謝辞

本研究の一部は、文部科学省科学研究費補助金・基盤研究 S(21220002) および基盤研究 B(24300037) 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] T. Camp, J. Boleng, and V. Davies, "A Survey of Mobility Models for Ad Hoc Networks Research," *Wireless Communications and Mobile Computing*, Vol.2, No.5, pp. 483-502, 2002.
- [2] 駒井友香, 佐々木勇和, 原 隆浩, 西尾章治郎, "アドホックネットワークにおける k 最近傍データ検索手法," マルチメディア通信と分散処理ワークショップ (DPSWS 2012), pp.17-25, 2012.
- [3] C.D. Manning, P. Raghavan, and H. Schütze, "Introduction to Information Retrieval," *Cambridge University Press*, 2008.
- [4] T.P. Nghiem, A.B. Waluyo, and D. Taniar, "A Pure Peer-to-Peer Approach for k NN Query Processing in Mobile Ad Hoc Networks," *Personal and Ubiquitous Computing*, 2012.
- [5] Z. Song and N. Roussopoulos, " K -Nearest Neighbor Search for Moving Query Point," *Proc. Int. Symposium on Spatial and Temporal Databases*, pp.79-96, 2001.
- [6] G. Tsuchida and S. Ishihara, "Replica Arrangement for Location Dependent Data in Consideration of Network Partition in Ad Hoc Networks," *Int. Journal of Communication Networks and Distributed Systems*, Vol.2, No.4, pp. 401-423, 2009.
- [7] 屋代智之, T.F.L. Porta, "Nomadic Agent System: インフラに依存しない位置情報サービス提供システム," 情報処理学会論文誌, Vol.46, No.12, pp. 2952-2962.
- [8] Q. Zhu, D.L. Lee, and W.-C. Lee, "Collaborative Caching for Spatial Queries in Mobile P2P Networks," *Proc. Int. Conf. on Data Engineering*, pp. 279-290, 2011.