

モバイルアドホックネットワークにおける Top-k 検索のための位置情報を用いた複製配置手法 Location-based Replication Method for Top-k Query in MANETs

佐々木 勇和[†] 原 隆浩[†] 西尾 章治郎[†]
Yuya SASAKI[†], Takahiro HARA[†], and Shojiro NISHIO[†]

1. まえがき

近年、無線通信が可能な移動端末のみで構成される、モバイルアドホックネットワークへの関心が高まっている。モバイルアドホックネットワークでは、通信インフラが必要ないため、災害地での救助活動などへの活用が期待されている。また、効率的な情報共有の技術は非常に重要である。必要な情報のみを取得する検索方法として、最も関連のある上位 k 個のデータを取得する Top-k 検索が提案されている。本稿では、モバイルアドホックネットワークにおいて、効率的な Top-k 検索処理の実現を目標とする。

モバイルアドホックネットワーク環境において、Top-k 検索を単純に実行する場合、ネットワーク全体に検索クエリをフラッディングする必要があるため、検索範囲が非常に広がる。このとき、他の端末がもつデータを複製として保持する、“複製配置”を用いることにより、検索範囲を限定的にすることが可能である。複製配置では、基本的に、各端末はアクセス頻度が高いデータを複製として保持することで、データ検索にかかるオーバーヘッドを削減することができる。その一方で、Top-k 検索では、例えば、スコアが最も高いデータは常に検索されるなど、アクセス頻度に大きな偏りがあるため、複製の多様性が低下してしまう可能性がある。データの多様性を損なわずに、各端末が保持する複製を決定する必要がある。さらに、モバイルアドホックネットワークのような、無線通信を用いるネットワークでは、地理的に近い端末からデータを検索することにより、オーバーヘッドの削減が可能である。そのため、位置情報を用いた複製配置手法が数多く提案されている。しかし、モバイルアドホックネットワークでは、端末が自由に移動するため、複製の再配置が重要な課題となる。

そこで、本稿では、効率的に Top-k 検索を実行するために、位置情報を用いた複製配置手法を提案する。この手法では、ネットワークを六角形のグリッドに区切り、グリッド毎に配置する複製を決定する。配置する複製は、データのアクセス頻度、無線通信距離、および端末数によって決定する。端末が移動した場合や、データが削除および生成された場合、各端末は自律的に複製を再配置する。また、Top-k 検索を実行する際は、周辺のグリッドのみを検索するだけで、上位 k 個のデータを取得することが可能である。シミュレーション実験により、提案手法は狭い範囲のみの検索で必要な情報を取得できることを確認する。

2. 関連研究

複製配置手法、および Top-k 検索手法における、代表的な既存手法について説明する。

2.1 複製配置手法

モバイルアドホックネットワークにおける複製配置手法について説明する。文献 [3] では、データの可用性の向上を目的として、三つのアルゴリズムが提案されている。それぞれのアルゴリズムは、データのアクセス率、隣接端末が保持する複製、およびネットワークポロジを考慮して、配置する複製を決定する。文献 [5] では、LACMA と呼ばれる位置依存の複製配置手法が提案されている。LACMA では、データは特定の区切られた領域（グリッド）に配置され、グリッド内を検索することによって要求データを取得できることを保証する。これらのブ

ロトコルは、基本的に一回の検索で一つのデータを検索することを想定している。しかし、Top-k 検索のような、ランクベースのクエリでは、一度の検索で複数のデータを取得するため、複製データの位置を決定することがより難しい。

文献 [4] において、モバイル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が提案されている。この手法では、返信データを受信および中継した端末がそのデータを複製として配置する。文献 [4] では、複製配置手法のみを提案しており、複製が配置されている状況における、Top-k 検索手法については検討していない。さらに、この複製配置手法は、スコアの低いデータを必要以上に多く配置してしまう可能性があり、効率的とはいえない。

2.2 Top-k 検索処理手法

文献 [1, 2] において、モバイルアドホックネットワークにおける Top-k 検索処理手法が提案されている。文献 [2] では、クエリ発行端末が検索メッセージをネットワーク全体に送信し、受信した端末はスコアが高いデータのみを返信する。これにより、無駄な返信データの最小化を目的としている。文献 [1] では、必要な端末にのみメッセージを送信する Top-k 検索のためのルーティング手法を提案している。これらの手法では、複製を考慮していない点、およびネットワーク全体へメッセージを送信する点が問題となる。

3. システムモデル

移動端末のみで構成されるモバイルアドホックネットワークを想定し、クエリ発行端末は、スコアが高いデータ k 個のデータを取得することを目的とする。ネットワーク内に M 台の端末が存在し、各端末はそれぞれ識別子 m_i ($i = 1, \dots, M$) が付与されており、自由に移動する。また、それぞれの端末は B 個のデータを複製として配置できるものとする。各端末は、GPS などの位置測位装置を保持しており、自身の位置情報を既知であるとする。さらに、データのアクセス頻度および端末が存在している領域を既知であるものとする。各端末は、通信半径 R の通信機器を保持しており、通信半径内の端末と通信することが可能である。

ネットワーク内に D 個のデータが存在し、それぞれ識別子 d_i ($i = 1, \dots, D$) が付与されている。本稿では、添え字をデータの順位（つまり、 d_1 は最もスコアが大きいデータ、および d_k は k 番目にスコアが大きいデータ）とする。それぞれのデータは特定の端末に保持されており、各端末は自由にデータを削除および生成する。簡単化のため、全てのデータのサイズは同じとする。

4. 複製配置手法

本章では、提案する Top-k 検索のための位置情報を用いた複製配置手法について説明する。

4.1 設計方針

Top-k 検索では、低オーバーヘッドおよび低遅延でスコアが高い上位 k 個のデータを取得することが重要である。スコアが高いデータを自身周辺の端末が複製として保持すれば、効率的に必要なデータを検索することができる。そこで、本稿では、領域を六角形のグリッドに区切り、グリッド内にスコアが高いデータを留まらせることにより、検索範囲を限定する。しかし、各端末が保持することができる複製数には制限があるため、グリッドのサイズやグリッドに留まらせる複製データが重要となる。提案する複製配置手法では、データのアクセス頻度 (k が

[†]大阪大学大学院情報科学研究科
Grad. Sch. of Information Science and Tech., Osaka Univ.

指定される確率), 無線通信距離, およびグリッド内の端末数を考慮して, グリッドサイズと留まらせる複製データを決定することにより, 効率的な配置を実現する.

4.2 複製配置の方針

各端末は, 上位 k 個のデータを小さいホップ数で取得することができれば, 低オーバーヘッドおよび低遅延を実現することができる. そこで, 各グリッドにスコアが高いデータを配置した際に, 平均何ホップで取得できるかを計算する. ここで, 六角形のグリッドの対角線は, 通信半径 R とする (つまり, 同じグリッド内の端末のデータは 1 ホップで取得可能である).

$$H_{avg} = \sum_{i=1}^N \alpha_i \cdot (2i - 1) \quad (1)$$

$$c/k_{max} \geq \sum_{i=1}^N \frac{\alpha_i}{6 \cdot i - 5} \quad (2)$$

$$\sum_{i=1}^N \alpha_i = 1 \quad (3)$$

これらの式において, c はグリッド内に配置できる複製数 (グリッド内の端末台数 $\times B$), k_{max} は指定される最大の k , および N は最大探索範囲を示す. $\alpha_i \cdot k_{max}$ 個のデータの探索範囲を, $(2 \cdot i - 1)$ ホップとして, 平均ホップ数 H_{avg} を計算している. i が大きくなるほど, 各グリッドで分担して保持する複製数が増加する (式 (2) の右項) が, 探索ホップ数が大きくなる (式 (1) の右項).

5 つの c/k_{max} の場合において, H_{avg} がほぼ最小となる α_i を以下に示す.

- $c/k_{max} \geq 1$ (つまり, 上位 k_{max} 個のデータを 1 つのグリッドに配置可能): $\alpha_1 = 1.0$ のとき, $H_{avg} = 1$.
- $c/k_{max} = 0.75$: $\alpha_1 = 0.7$ および $\alpha_2 = 0.3$ のとき, $H_{avg} = 1.6$.
- $c/k_{max} = 0.5$: $\alpha_1 = 0.4$ および $\alpha_2 = 0.6$ のとき, $H_{avg} = 2.2$.
- $c/k_{max} = 0.25$: $\alpha_1 = 0.12$ および $\alpha_2 = 0.88$ のとき, $H_{avg} = 2.76$.
- $c/k_{max} = 0.1$: $\alpha_1 = 0$, $\alpha_2 = 0.5$ および $\alpha_3 = 0.5$ のとき, $H_{avg} = 4$.

c/k_{max} が非常に小さい場合を除いて, α の添え字は, 2 までとなっている. これは, 自身および隣接の端末を検索すれば k_{max} 個のデータを検索できることを示す. 非常に小さい c/k_{max} は, 例外的であると考えられるため, 本稿では扱わない. さらに, 上記の計算では, 六角形の対角線を R としたが, 対角線を $2R$ とした場合, グリッド内に配置できる複製数は 4 倍となる. そのため, ホップ数の平均 H_{avg} が 2 を超える場合, 六角形のサイズを大きくするべきである.

上記の議論をもとに, それぞれの端末は自身が所属するグリッドと隣接するグリッドから必要なデータを取得できるように配置する複製を決定する. まず, ネットワークを六角形のグリッド (グリッドのサイズについては後述) に分け, それぞれのグリッドに識別子 $g(i, j)$ ($i, j = 0, \dots, N$) を割り当てる (図 1). 配置する複製は, 以下の式で決定する.

$$k_1 = k_{max}/2 \quad (4)$$

$$c_k = k_1 + \lceil (k_{max} - k_1)/7 \rceil + \text{margin} \quad (5)$$

これらの式において, k_1 は自身のグリッドのみで取得するデータの順位, c_k は各グリッドが配置すべき複製数, および margin は再配置を防ぐための余分に保持する複製数を示す. それぞれのグリッド $g(i, j)$ は, 1 から k_1 位までのデータと, $(k_1 + \lceil \frac{k_{max}-k_1}{7} \rceil \cdot \lfloor \frac{i+5 \cdot (j \bmod 2)}{2} \rfloor - 2 \cdot (j \bmod 2)) \bmod 7 + 1$ から $(k_1 + \lceil \frac{k_{max}-k_1}{7} \rceil \cdot \lfloor \frac{i+5 \cdot (j \bmod 2)}{2} \rfloor - 2 \cdot (j \bmod 2)) \bmod 7 +$

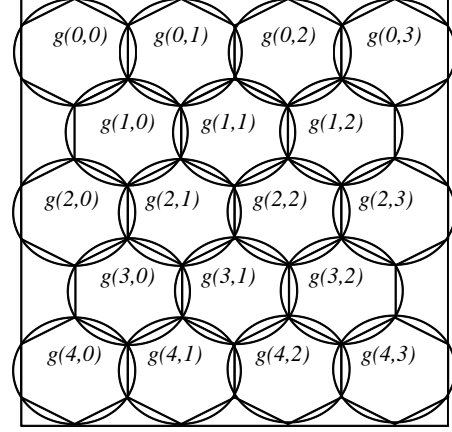


図 1: グリッドの配置

$\lceil \frac{k_{max}-k_1}{7} \rceil$ までのデータを保持する. ここで, margin は, それぞれの範囲を拡張して保持する. 具体的には, それぞれの範囲を $\frac{\text{margin}}{3}$ の幅で拡張する. 一方, グリッドに所属していない端末は, 上位 B 個のデータを保持する.

全てのグリッドで配置できる複製数 $c_{g(i, j)}$ は, c_k を超える必要がある. 全ての $c_{g(i, j)}$ で, c_k を超える最小の $x \cdot R$ (x は整数) をグリッドの対角線の長さとする.

この配置により, 多くのデータを自身が所属するグリッドのみで取得でき, k_{max} 位までのデータを隣接するグリッドを検索することにより, 取得することが可能である.

4.3 複製の配置

本稿で提案する複製配置では, まず 1 台の端末 (代表端末とよぶ) がネットワーク全体の $(k_{max} + \text{margin})$ 位までのデータおよび全ての端末の位置情報を収集する. その後, 代表端末が各端末の位置からグリッドのサイズを計算する. このとき, グリッドの識別子 $g(0, 0)$ の位置を基準とする. 次に, 代表端末は, 上位 $(k_{max} + \text{margin})$ 位までのデータ, 全ての端末の位置情報, およびグリッドのサイズが添付されたメッセージをネットワーク上の全ての端末にフラッディングにより送信する. メッセージを受信した端末は, データの複製を配置する. このとき, 各端末が配置する複製は, グリッドの中心に近い端末ほど, 以下に示す優先度が高いデータを配置する.

1. 第一優先データ: $(k_1 + \lceil \frac{k_{max}-k_1}{7} \rceil \cdot \lfloor \frac{i+5 \cdot (j \bmod 2)}{2} \rfloor - 2 \cdot (j \bmod 2)) \bmod 7 + 1$ から $(k_1 + \lceil \frac{k_{max}-k_1}{7} \rceil \cdot \lfloor \frac{i+5 \cdot (j \bmod 2)}{2} \rfloor - 2 \cdot (j \bmod 2)) \bmod 7 + \lceil \frac{k_{max}-k_1}{7} \rceil$ 位のデータ.
2. 第二優先データ: 1 から k_1 位までのデータ.
3. マージンデータ: マージンのための margin .
4. その他: 上記に含まれない $(k_{max} + \text{margin})$ 位までのデータ.

この優先度により, 必要なデータを複製としてグリッド内に配置することができる. $(k_{max} + \text{margin})$ 個以上のデータを配置できる場合, $\lfloor \frac{(k_{max} + \text{margin}) + 1}{B} \rfloor$ 番目に中心から遠い端末から, 第一優先データから順に重複して複製として配置する. 1 から k_1 位までの最もスコアが高いデータ群が第二優先度である理由は, 全てのグリッドでこれらのデータを配置するためである.

メッセージを受信した端末は, $(k_{max} + \text{margin})$ 位までのデータのスコアおよび識別子, および自身のグリッド内に存在する端末台数も把握することができる. そのため, グリッド内に配置されている複製を把握でき, 検索の際に必要なグリッドにのみメッセージを送信することができる.

4.4 複製の再配置

この節では、4.3 節で複製データを各端末に配置した後の、複製データの再配置について説明する。具体的に、データの削除および生成、および端末の移動における再配置について説明する。

4.4.1 データ削除

($k_{max} + margin$) 位までのデータが削除された場合、全ての端末に“削除メッセージ”を送信する。この削除メッセージには、削除されたデータの識別子およびデータを削除した時間が添付されている。削除メッセージを受信した端末は、スコアの順位を変更する。また、受信端末が削除されたデータを保持している場合、そのデータの複製を削除し、データを削除した時間を記録する。このとき、データを削除した端末が、自身のグリッド内に存在する端末は、グリッド内に存在する複製を更新し、空き容量がこのとき何個できたかを把握する。データが削除された場合でも、マージンにより余分に複製を保持しているため、データの取得を保証することができる。

4.4.2 データ生成

($k_{max} + margin$) 位のデータよりスコアが大きいデータが生成された場合、全ての端末に“生成メッセージ”を送信する。この生成メッセージには、新しく生成されたデータが添付されている。生成メッセージを受信した端末は、スコアの順位を変更する。複製の再配置には 4 つのパターンがある。1 つ目は、ランク外 ($(k_{max} + margin + 1)$ 位) になったデータを保持している場合、そのデータと新しいデータを入れ替える。2 つ目は、複製のための空き容量がある端末が存在する場合、最も古い時刻に空き容量ができた端末がデータを複製として新しいデータを保持する。3 つ目は、グリッド内で最も優先度が低いデータの複製と新しいデータを入れ替える。4 つ目は、新しく生成されたデータがグリッド内で最も優先度が低い場合、そのデータは複製として保持しない。この方法では、各端末は自身もつグリッド内の複製の情報だけで、データの入れ替えを行うため、グリッド内でのメッセージの交換が必要ない。さらに、グリッド内で保持されている複製の情報も自律的に更新することが可能である。

4.4.3 端末の移動

端末がグリッド内から移動したとき、移動を知らせるメッセージを移動元と移動後のグリッドに知らせる必要がある。このとき、グリッド間の移動は、現在所属する六角形のグリッドの内接円から出たときを移動とする。これにより、グリッドの境界線付近の移動による頻繁なグリッド間移動を抑制することができる。

端末がグリッドを移動するとき、“移動メッセージ”を隣接端末に送信する。この移動メッセージには、保持している第一優先データ、保持している第一優先データ以外の複製の識別子、空き容量の情報（複製を削除した時間）、および移動元と移動後のグリッドの識別子が含まれている。移動メッセージを受信した移動元のグリッド内の端末は、受信した第一優先データをグリッド内の端末が保持していない場合、空き容量に複製として配置、もしくはグリッド内で最も優先度が低い複製と入れ替える。移動メッセージを受信した移動後のグリッド後の端末は、受信したメッセージをもとにグリッド内の複製情報を更新し、これをメッセージに添付して送信する。これにより、移動した端末は移動後のグリッド内の複製情報を知ることができる。それぞれの移動元および移動後のグリッド内の端末は、グリッドのサイズが $2 \cdot R$ 以上の場合、メッセージをそれぞれのグリッドにフラッディングする。これらの処理により、グリッド内の複製の情報の整合性を保つことができる。

5. Top-k 検索処理手法

提案する Top-k 検索処理手法では、限られたグリッドにのみ、検索クエリをフラッディングする。自身が所属するグリッ

ド内に存在するデータは把握しているため、上位 k 個のデータをこのグリッドだけで取得できるかを確認する。取得できる場合、このグリッドにのみ検索クエリメッセージをフラッディングする。一方、できない場合、隣接するグリッドが第一優先データを保持しているものとし、必要なデータを取得するためのグリッドも検索範囲とする。検索クエリには、クエリ発行端末の識別子、 k 、検索条件、グリッド識別子、送信端末の識別子、親端末の識別子、およびデータ識別子リストが添付されている。データ識別子リストは、端末が保持する上位 k 個のデータの識別子の集合であり、それぞれの端末が追加していくことにより、重複したデータの返信を防ぐ。また、添付されたグリッド識別子に該当するグリッド内に所属する端末のみが、メッセージを再転送する。

メッセージを受信した端末は、送信端末を親端末とする。また、メッセージを再受信した端末は、メッセージ内の親端末の識別子が自身の場合、送信端末を自身の子端末とする。データ返信の際は、子端末がいない端末から、上位 k 個のデータを親端末に返信する。このとき、データ識別子リストに添付されているデータは、返信しない。全ての子端末から受信した端末は、データを集約して、自身の親に送信する。これを繰り返すことにより、クエリ発行端末は、上位 k 個のデータを限られた範囲から効率的に取得することができる。

6. 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*。

6.1 評価環境

1,000[m]×1,000[m] の 2 次元平面上の領域に 500 台の端末が存在し、初期位置をランダムに決定した。各端末はランダムウォークモデルに従い、0.5 から 1[m/秒] の速度で移動する。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。各端末は、初期状態として、それぞれデータを 10 個保持し、30 秒から 3600 秒間のランダムな時間でデータを削除、または生成する。データサイズは、32[B] として、データのスコアは、0 から 100 の間で一様乱数に従い、ランダムに決定する。また、各端末は、 $B(5 \sim 100)$ 個のデータを複製として配置することができる。クエリ発行端末は、 k として、25, 50, 75, および 100 を等しい確率で指定し、本稿で提案した Top-k 検索手法によってデータを取得する。

以上のシミュレーション環境において、まず、1 台の端末が端末の位置と 109 位 ($k_{max} = 100, margin = 9$) までのデータを収集し、全ての端末に通知する。その後、30 秒毎にクエリを発行し、10,800 秒経過させた場合の以下の評価値を調べた。

- **検索精度**: 上位 k 個のデータうち、取得できたデータ数の平均割合。
- **遅延 [秒]**: 検索クエリ発行後、データ取得までに経過した平均時間。
- **検索オーバーヘッド [KB]**: 検索クエリおよび返信メッセージの総バイト数の 1 クエリ当たりの平均値。
- **再配置オーバーヘッド [KB]**: 再配置のためのメッセージの総バイト数の 1 クエリ当たり (30 秒間) の平均値。

6.2 評価結果

配置可能複製数を変化させて、提案手法の性能を調べた。その結果を図 2 に示す。それぞれ図 2 の横軸は、配置可能複製数として、縦軸は、(a) では検索オーバーヘッド、(b) では検索

* Scalable Networks: makers of QualNet and EXata, the only multi-core enabled network simulation and emulation software. [Online]. Available: <http://www.scalable-networks.com/>.

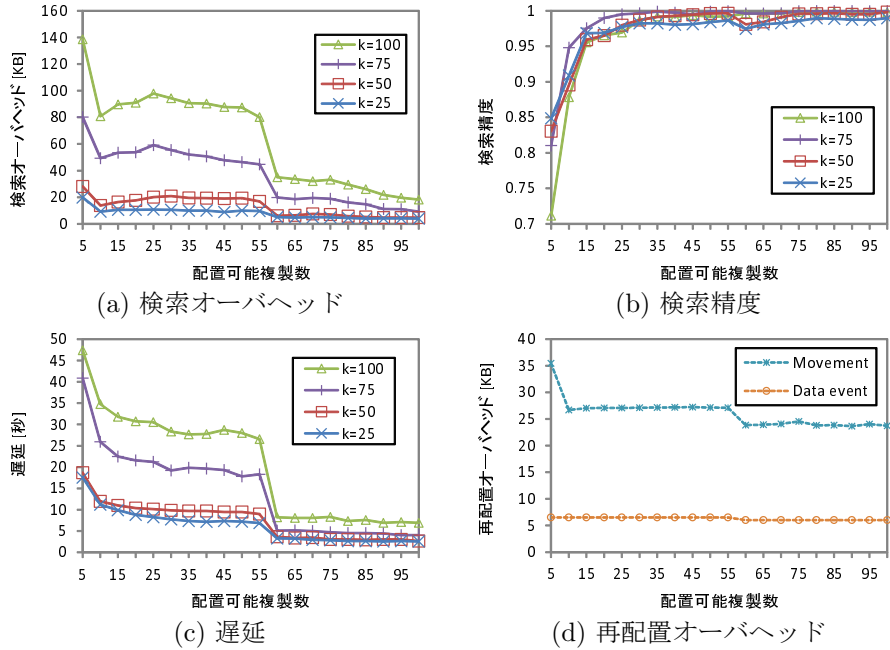


図 2: 配置可能複製数の影響

精度, (c) では遅延, および (d) では再配置オーバーヘッドである。グラフの凡例において, $k = x$ はクエリ発行端末が k として x を指定した場合の結果, “Movement” は端末移動に伴う再配置オーバーヘッド, および “Data event” はデータの削除および生成に伴う再配置オーバーヘッドを示す。

図 2(a) より, B が小さく, k が大きい場合, オーバーヘッドが非常に大きい。これは, グリッドサイズが大きく, 周辺のグリッドにクエリメッセージを送信する必要があるためである。 k が小さい場合, 自身が所属するグリッドのみを検索するため, オーバーヘッドが小さい。 $B = 10$ になると, グリッドサイズが小さくなり, オーバーヘッドが小さくなる。その後, B が大きくなると, 返信データ数が増加するため, オーバーヘッドが若干大きくなり, さらに B が大きくなると, 検索範囲が小さくなるため, オーバーヘッドが若干小さくなる。 $B = 60$ になると, さらにグリッドサイズが小さくなるため, オーバーヘッドが減少する。

図 2(b) より, B が小さいとき, パケットロスにより, 検索精度が低い。これは, 検索範囲が広く, 検索オーバーヘッドが大きい場合, パケットロスが発生しやすいためである。 B が大きくなると, 検索精度が高くなる。これには, 2つの理由がある。まず, クエリ発行端末がより多くの複製を保持できるためである。次に, 返信データ数が減少し, パケットロスが減少するためである。 $B = 60$ のとき, k が 25 および 50 のときに精度が低くなっている。これは, グリッドサイズが R で, グリッド内の端末が同時にデータを返信する場合が多く, パケットロスが生じやすいためである。 B が 60 より大きくなると, 自身の複製から検索結果を取得できる可能性が高くなるため, 精度が高くなる。

図 2(c) より, k が小さいとき, 遅延が小さい。これは, 探索範囲が小さいためである。また, $B = 10$ および $B = 60$ のときに, グリッドが小さくなるため, 大きく遅延が小さくなる。

図 2(d) より, まず, B が大きくなると, 移動に伴う再配置オーバーヘッドが減少する。これは, 端末が移動した場合, メッセージを移動元および移動後のグリッドにメッセージをフラッシングするときのオーバーヘッドが小さくなるためである。データの削除および生成に伴うオーバーヘッドは, 配置可能複製数に依存しないため, 一定である。

7. むすび

本稿では, Top- k 検索のための位置情報を用いた複製配置手法について提案した。提案手法では, ネットワークを六角形のグリッドに分割し, グリッド毎に複製に配置する。Top- k 検索を実行する際は, 自身が所属するグリッドおよび隣接グリッドのみを検索することによって, 必要な上位 k 個のデータを取得することができる。シミュレーション実験より, 提案手法は, 低オーバーヘッド, 低遅延, および高取得精度を達成することを確認した。今後, 端末の粗密がある環境における効率的な複製配置について検討する予定である。

【謝辞】

本研究の一部は, 文部科学省研究費補助金・基盤研究 S (21220002), 基盤研究 B (24300037), および特別研究員奨励費 (24-293) 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] D. Amagata, Y. Sasaki, T. Hara, and S. Nishio. A robust routing method for top- k queries in mobile ad hoc network. In *Mobile Data Management (MDM)*, pages 250–256, 2013.
- [2] R. Hagihara, M. Shinohara, T. Hara, and S. Nishio. A message processing method for top- k query for traffic reduction in ad hoc networks. In *Mobile Data Management (MDM)*, pages 11–20. IEEE, 2009.
- [3] T. Hara. Effective replica allocation in ad hoc networks for improving data accessibility. In *INFOCOM*, volume 3, pages 1568–1576. IEEE, 2001.
- [4] T. Hara, R. Hagihara, and S. Nishio. Data replication for top- k query processing in mobile wireless sensor networks. In *Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC), 2010 IEEE International Conference on*, pages 115–122. IEEE, 2010.
- [5] S. Lee, S. Wong, K. Lee, and S. Lu. Content management in a mobile ad hoc network: Beyond opportunistic strategy. In *INFOCOM*, pages 266–270. IEEE, 2011.