

モバイル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

Research on Top-k Query Processing in Mobile Ad Hoc Networks

佐々木 勇和[◆]
Yuya SASAKI

本研究は、Top-k 検索をモバイルアドホックネットワーク上で効率的に処理することを目的とする。モバイルアドホックネットワークは、移動端末のみで構築されているため、基地局を介さずに通信できる。災害時の救助活動など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。一方、Top-k 検索では、指定した検索条件に関連する上位 k 個のデータを検索する。モバイルアドホックネットワーク上で効率的に Top-k 検索を処理することにより、通信量削減によるバッテリー消費の軽減などが可能となる。これにより、更なるモバイルアドホックネットワークの普及が考えられる。

In this paper, our goal is to efficiently process top-k queries in mobile ad hoc networks. Mobile ad hoc networks are composed of solely mobile nodes, and thus nodes can communicate each other without any infrastructure. Thus, mobile ad hoc networks are expected to develop application for rescue effort in disaster sites and data gathering in sensor networks. On the other hand, top-k queries retrieve the most relevant k data items. Efficient top-k query processing in mobile ad hoc networks can reduce communication overhead, and thus enable to reduce energy consumption. Therefore, we consider that our research facilitates the spread of mobile ad hoc networks.

1. はじめに

近年、IEEE802.11 や Bluetooth, Zigbee などをはじめとする無線通信技術の発展と、携帯電話などの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、端末が自身の無線通信範囲内に存在する端末と通信する場合、基地局を介さずに通信できる。一方、自身の無線通信範囲内に存在しない端末と通信する場合、これらの端末の間に存在する端末がパケットを中継してマルチホップ通信を行う。このように、アドホックネットワークは既存の通信インフラを必要とせず、端末同士で自律分散的にネットワークを構築できるため、災害時の救助活動および山岳での遺跡発掘調査など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。

一方、情報量の増加と多様化に伴い、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の検索技術が求められている。その一つとして、指定した検索条件に関連

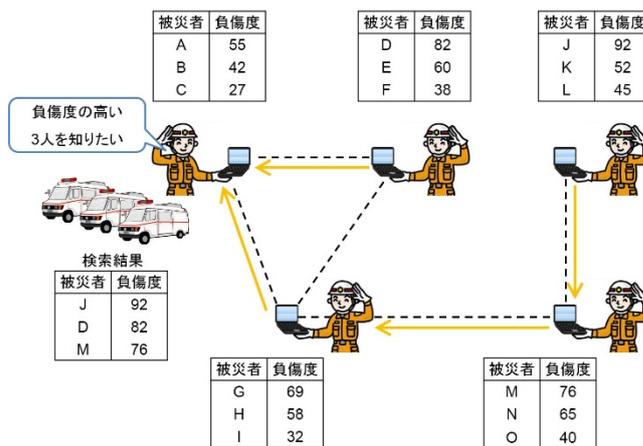


図1 モバイルアドホックネットワークにおける Top-k 検索の例

Fig.1 An example of top-k query in mobile ad hoc networks

する上位 k 個のデータを検索する Top-k 検索が幅広く研究されている。Top-k 検索では、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。もともとは、集中管理システムにおいて、計算量を削減するための研究が一般的であったが、近年では、P2P ネットワークやセンサネットワークなどの分散環境において研究が盛んに行われている。

本論文では、Top-k 検索をモバイルアドホックネットワークに適用することを想定とする。例えば、各端末 (ユーザ) 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など、様々な状況において Top-k 検索は有効と考える。図1は、緊急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、救助隊員の持つ端末同士でネットワークを構築している様子を示している。このような環境において、左上の救助隊員が限られた数の救急車で、負傷度の高い被災者を優先的に搬送する場合、左上の救助隊員が負傷度の高い数人 (この例では3人) の被災者を検索するために Top-k 検索を行うことが有効と考えられる。

2. 研究動機

2.1 通信量の削減および検索精度の維持

モバイルアドホックネットワークにおいて、Top-k 検索を実行する端末 (クエリ発行端末とよぶ) は、まず検索クエリメッセージをネットワーク内に存在する端末に送信する。クエリメッセージを受信した端末は、自身が保持するデータを返信する。モバイルアドホックネットワークでは、無線通信帯域に制限があるため、自身が保持する全てのデータを返信すると、ネットワーク内の通信量 (トラフィック) が増加し、パケットロスが発生してしまう。パケットロスが発生した場合、上位 k 個に入る必要なデータをクエリ発行端末が取得できず、検索精度が低下してしまう。トラフィックを減少させるために、返信データ数を削減することが重要である。

しかし、不用意に返信データ数を削減すると、必要なデータの返信もしない可能性がある。そのため、上位 k 個に入らない無駄なデータの返信によるトラフィックを削減しつつ、上位 k 個のデータに入る必要なデータのみを返信することが求

◆ 学生会員 大阪大学大学院情報科学研究科
sasaki.yuva@ist.osaka-u.ac.jp

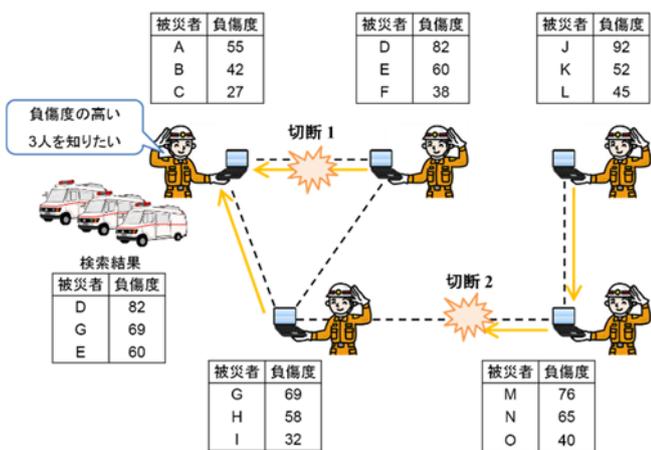


図2 ネットワーク分断の例
Fig.2 An example of network partition

められる。各端末は、他の端末が保持するデータに関する情報をもたないため、検索クエリにどのような情報を添付して、メッセージを送信するかが重要となる。

図1では、クエリ発行端末は、負傷度が大きい上位3人のデータを要求している。この例では、データJ, D, およびMのみが必要なデータとなるため、その他のデータの返信は、無駄なトラヒックとなる。

2.2 ネットワークトポロジの変化

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、ネットワークトポロジが動的に変化する。隣接する端末とのリンクが切断すると、構築した転送経路によるデータ転送ができなくなる可能性がある。そのため、できる限りリンク切断が発生する前に検索処理を終わらせること、および、それでもリンク切断が発生した場合は新たな経路を探す処理が必要となる。さらに、ネットワークが複数のグループに分断してしまうことも考えられる。この場合、他のグループに所属する端末へのアクセスができず、必要なデータを取得できない可能性がある。

図1の例において、リンク切断が発生した場合の影響を図2を用いて説明する。リンク切断1が発生した場合は、中央上の端末がデータを左上のクエリ発行端末に直接送信することができないため、左下の端末を経由して、送信する必要がある。一方、リンク切断2が発生した場合、ネットワークが分断してしまうため、クエリ発行端末は右側のグループに所属する端末のデータを取得することができず、検索精度が低下する。

2.3 複製配置の利用

他の端末が保持するデータを自身の記憶領域に保持する複製配置により、より効率的なTop-k検索を実現することができる。例えば、頻繁にアクセスされるデータを多くの端末が保持することにより、近く of 端末から必要なデータを取得できる。これにより、ネットワーク分断が発生した場合でも、正確な検索結果の取得を可能にする。しかし、複数の端末が同じデータを保持することにより、同じデータが複数返信されることが考えられる。これは、トラヒック増加の原因となり、パケットロスを引き起こす可能性がある。

例えば、図2において、上位3個に入るデータJおよびMを左側のグループの端末が複製として保持することにより、リンク切断2によりネットワークが分断しても、検索結果の

精度を保つことができる。しかし、ネットワーク分断が発生していない場合では、JおよびMを保持する複数の端末がこれらのデータを重複して返信し、無駄なトラヒックが発生する。

3. 提案手法

本章では、本論文において提案している三つの手法について簡単に説明する。

3.1 データ分布推定によるTop-k検索手法

データ分布を推定することにより、返信データ数を削減するTop-k検索手法を提案した。返信データ数を削減するには、正確なk番目のスコアを求める必要があるが、検索クエリメッセージ転送の際は、一部の端末しか経由しないため、全ての端末が保持するデータの情報を取得することは困難である。そのため、一部の端末のデータからネットワーク全体の端末が保持するデータのスコアを推定することが求められる。本研究では、データのスコアがあるデータ分布に従っているものと想定し、このデータ分布の推定により、k番目のスコアを求めることを試みた。具体的には、検索クエリメッセージにデータスコアから作成されるヒストグラムを添付し、各端末はこのヒストグラムからネットワーク全体のデータが従う分布を推定する。分布の推定では、ヒストグラム内のスコア数を用いる方法、および確率密度関数を用いる方法を用いた。ヒストグラム内のスコア数を用いる方法では、ヒストグラムのそれぞれの値を、メッセージを経由した端末数とネットワーク全体の端末数から求める係数倍する。また、確率密度関数を用いる方法では、分布の概形を決定する要素（正規分布では、平均と分散）を計算する。推定した分布からk番目のスコアを求め、そ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、推定の誤差により、検索精度の低下を防ぐために、推定したk番目のスコアを補正するセーフティマージンも提案した。さらに、ネットワークトポロジの変化に伴う端末間のリンクが切断した場合、データを迂回路を用いて返信する。この方法により、従来方法より、返信データ数を削減しつつ、検索精度を維持することが可能となった。

3.2 複製配置を考慮したTop-k検索手法

データの複製を考慮して、Top-k検索を実行する手法を提案した。検索実行時に、各端末が返信データを複製として、配置する手法が提案されている。この手法を含めて、一般的に、複製配置では、頻繁にアクセスされるデータの複製を配置する。Top-k検索では、スコアが大きいデータが頻繁にアクセスされるため、各端末はスコアが大きいデータを複製として保持することになる。Top-k検索において、データのスコアを考慮し、配置する複製を決定する手法が提案されている。しかし、この手法では、多くの端末がスコアが大きいデータを配置するため、同じデータが重複して複数返信されてしまう可能性が高い。そこで、重複する返信データを防ぐTop-k検索手法を提案する。この手法では、クエリメッセージに自身が保持するデータの識別子を添付することにより、自身よりクエリ発行端末に近い端末が保持するデータを把握する。さらに、全ての端末にメッセージを送信しなくても、一部の端末からのみで必要なデータを取得できる可能性がある。そのため、提案手法では、これまでにクエリメッセージを受信した端末のみから既に上位k個のデータを取得できるかを各端末が自律的に判断する。取得できると判断した場合、クエリメッセージの転送を停止し、データの返信をすぐ

を開始する。これにより、トラヒック、および検索時間の削減が可能である。

3.3 Top-k 検索のための複製配置およびメッセージ処理手法

モバイルアドホックネットワークにおける Top-k 検索では、近くの端末から上位 k 個のデータを取得することが理想的である。そこで、効率的に複製を配置した上で、近くの端末からデータを取得する方法を実現する必要がある。しかし、端末が移動すると、アクセス可能な端末が変化するため、周囲の状況に合わせて、検索の範囲を動的に決定することも必要とされる。本研究では、複製配置手法とメッセージ処理手法を同時に考えることにより、上記の課題を解決する。ここで、Top-k 検索では、スコアが最も大きいデータが常に検索対象となる等、アクセスされるデータに大きな偏りがある。このとき、データのアクセス頻度のみを考慮して、複製の配置頻度を決定すると、配置される複製数に大きな偏りが発生してしまう。さらに、Top-k 検索では、順位が近いデータは、同時にアクセスされる可能性が高いため、近い順位のデータをまとめて保持することにより、アクセスしなければいけない端末数を減少させることができる。そこで、各端末が保持するデータの複製の組合せ、および複製データの多様性を保つ複製の配置頻度を決定する複製配置手法を提案する。また、近くの端末から必要なデータを取得するまで、検索範囲を広げながら、繰り返しメッセージを送信するメッセージ処理手法も提案した。提案した複製配置、およびメッセージ処理手法は、既存の複製配置手法およびメッセージ処理手法よりも、高い性能を達成した。

4. おわりに

本論文では、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 検索処理手法を提案した。今後、ますます発展するであろうモバイルアドホックネットワークにおける検索処理の効率化を実現した。

佐々木 勇和 Yuya SASAKI

2014 年大阪大学大学院情報科学研究科博士後期課程修了。
モバイル環境におけるクエリ処理に関する研究に従事。情報処理学会学生会員。