

構造化オーバーレイネットワークによるタイムスロットを用いた正確なデータ集計手法

武田 敦志†

† 東北学院大学教養学部情報科学科

1 はじめに

携帯端末やセンサ端末が普及したことにより、広域に配置された多数の端末から膨大な量のデータを計測することが可能となった。これらの端末で計測されるデータを集計することにより、広域の環境変化を観測することや遠隔地の異常を検知することができる。従来よりグリッドシステムを対象としたデータ集計システムが開発されてきたが、多くのシステムは中央サーバでデータ集計を行うためスケラビリティや対故障性に問題があった。そこで、構造化オーバーレイネットワークのルーティングの特性を利用してデータを集計する手法が提案された [1]。この手法は高いスケラビリティと耐故障性を有するが、1回のデータ集計に必要な通信データ量が $O(N)$ (N はネットワーク内のノード数) となるため通信の効率性に問題がある。一方、構造化オーバーレイネットワークの各ノードが部分的なデータ集計を非同期に行うことにより、 $O(\log N)$ の通信データ量で集計を行うことができる効率的なデータ集計手法が提案されている [2, 3]。しかし、これらの手法では集計対象となるデータの計測時刻を統一できないため、正確な集計結果を得ることが難しい。

そこで、本稿では、構造化オーバーレイネットワークを用いて従来よりも正確な集計結果を得ることができるデータ集計手法を提案する。提案手法では、従来のデータ集計手法にタイムスロットの概念を導入することにより、集計対象となるデータの計測時刻を統一する。提案手法の特徴を以下に示す。

1. 効率的な集計方式を用いており、1回のデータ集計に必要な通信データ量は $O(\log N)$ となる。
2. データの計測時刻を考慮した集計方式を導入しており、従来よりも正確な集計結果を得られる。
3. データ集計に時間を要するためリアルタイムに集計結果を得ることはできない。

本稿では構造化オーバーレイネットワークを用いたデータ集計手順について述べ、提案手法であるタイムスロットの概念を導入した正確なデータ集計手法を提案する。さらに、プロトタイプシステムを用いて実験を行い、その結果を通じて提案手法の有効性を示す。

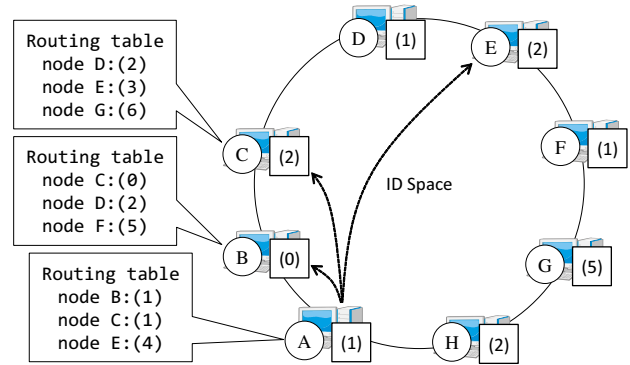


図 1: データ集計を行う構造化オーバーレイネットワーク

2 構造化オーバーレイネットワークを用いた効率的なデータ集計手順

構造化オーバーレイネットワークを用いて効率的にデータ集計を行う手順 [2, 3] は以下の通りである。

1. 各ノードは ID 空間上で自身とルーティング先ノードの間に存在するデータを非同期に集計する。
2. 各ノードが集計した部分的な集計結果を集めることで全体の集計結果を得る。

以下、このデータ集計手順を Chord# [4] に導入した場合のデータ集計手順について述べる。

図 1 に効率的にデータ集計を行う構造化オーバーレイネットワークの構造を示す。図 1 の括弧内の数字は集計対象となるデータの値である。各ノードは ID 空間上で自身とルーティング先のノードとの間に存在するデータを集計し、これをルーティングテーブルのレコードと関連付けて記録する。例えば、図 1 の場合、node A は ID 空間上で自身とルーティング先のノードである node B, C, E との間に存在するデータを集計し、これらを node A のルーティングテーブルのレコードに関連付けて記録する。この部分的なデータ集計はルーティングテーブルの更新時に実行される。全体の集計結果が必要な場合、各ノードのルーティングテーブルに関連付けて記録された部分的な集計データを集めることにより全体のデータ集計結果を得る。この全体の集計結果を得るために必要となる通信データ量は $O(\log N)$ (N はネットワーク全体のノード数) となる。しかし、従来手法では、データの計測時刻を考慮せずに部分的な集計データを非同期で更新するため、得られる集計データの正確性が低いという問題がある。

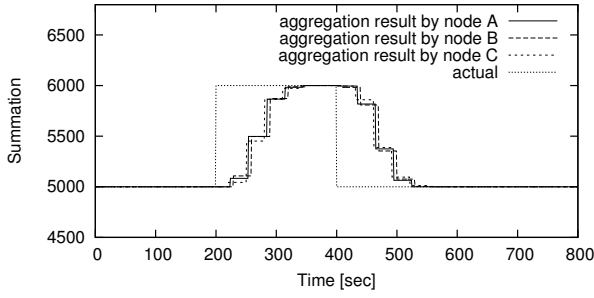


図 2: 従来手法で集計した計測データの合計値

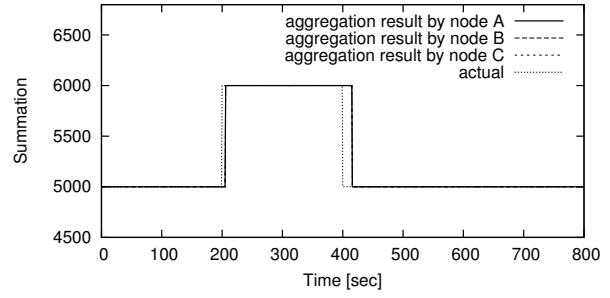


図 3: 提案手法で集計した計測データの合計値

3 提案: タイムスロットを用いた正確なデータ集計手法

提案手法では、2 で述べた構造化オーバーレイネットワークでのデータ集計手法にタイムスロットの概念を導入する。これにより、同一時刻に計測されたデータを集計できるため、従来よりも正確な集計結果を得ることができる。まず、構造化オーバーレイネットワーク上の各ノードは最新の計測データだけではなく、計測データの過去の履歴も保持する。ここで、時刻 t にノード n_i が保持する計測データの履歴を $V_i(t)$ とすると

$$V_i(t) = \{v_i(t), v_i(t-1), \dots, v_i(t-m)\}$$

となる。ここで、 $v_i(t)$ はノード n_i が時刻 t に計測したデータの値である。また、 m は計測データの履歴の長さであり、 d をルーティングテーブルの更新間隔とすると $m \geq d \log N$ (N はネットワーク上のノード数) となる。各ノードが持つ計測データの履歴を計測時刻を統一して集計することにより、ネットワーク全体の正確な集計結果を得る。時刻 t におけるノード n_i の計測データの履歴 $V_i(t)$ とノード n_j の計測データの履歴 $V_j(t)$ の集計結果を $V_{ij}(t)$ とすると

$$\begin{aligned} V_{ij}(t) &= \{v_{ij}(t), v_{ij}(t-1), \dots, v_{ij}(t-m)\} \\ v_{ij}(k) &= \text{aggregate}(v_i(k), v_j(k)) \end{aligned}$$

となる。上記の方法でデータ集計を行うことにより時刻 $t-m$ の正確な集計結果を得ることが可能となる。また、集計に必要な通信データ量は $O(\log N)$ (N はネットワーク上のノード数) である。一方、提案手法は集計結果を得るために m だけの時間が必要となるため、リアルタイムに集計データを得ることは難しい。

4 実験

提案手法を導入した構造化オーバーレイネットワークのプロトタイプシステムを実装し、このプロトタイプシステムを用いて実験を行った。この実験では単一のコンピュータ (CPU: ADM Opteron 4176 × 2, Memory: 48 Gbyte, OS: FreeBSD 9.2, VM: Java6 JRE) 上で 500 個のノードを動作させることで行った。500 個のノードはそれぞれ個別のプロセスとして動作し、相互の通信はループバックインタフェースを介して行った。集計対象となる各ノードの計測値の初期値は 10 (全体の合計は 5000) に設定し、実験開始から 200 秒後に各ノード

の計測値を 12 (全体の合計は 6000) に変更し、実験開始から 400 秒後に各ノードの計測値を 10 (全体の合計は 5000) に変更した。また、構造化オーバーレイネットワークの ID 空間の大きさは 2^{256} とし、ルーティングテーブルの更新間隔は 30 秒、計測データの履歴の長さは 480 秒に設定した。上述の条件で任意に選んだ 3 個のノードが集計した全体の計測値の合計値を測定した。

図 2 に従来手法 [3] で合計値を集計した結果を示す。従来手法では計測時刻を考慮せずに集計を行うため、計測値が変化した場合に正確に集計することは難しい。図 3 に提案手法で合計値を集計した結果を示す。提案手法ではデータの計測時刻を統一して集計を行うため、従来手法に比べて正確に集計することが可能となる。

5 おわりに

携帯端末やセンサ端末から得られる計測データを正確に集計する仕組みが必要とされている。そこで、従来の構造化オーバーレイネットワークを用いたデータ集計手法に対してタイムスロットの概念を導入し、従来よりも正確にデータ集計を行う仕組みを提案した。

謝辞

本研究は文部科学省科学研究費補助金若手研究 (B) (25730064) の助成を受けて実施したものである。

参考文献

- [1] Min Cai and Kai Hwang. Distributed aggregation algorithms with load-balancing for scalable grid resource monitoring. *Proc. 21th International Parallel and Distributed Processing Symposium (IPDPS 2007)*, pp. 1–10, 2007.
- [2] Kota Abe, Toshiyuki Abe, Tatsuya Ueda, Hayato Ishibashi, and Toshio Matsuura. Aggregation skip graph: A skip graph extension for efficient aggregation query over p2p networks. *International Journal On Advances in Internet Technology*, Vol. 4, No. 3, pp. 103–110, 2012.
- [3] Atsushi Takeda, Takuma Oide, and Akiko Takahashi. A structured overlay network for aggregating sensor data. *Proc. 7th International Conference on Broadband, Wireless Computing, Communication and Applications (BWCCA2012)*, pp. 684–689, 2012.
- [4] Thorsten Schütt, Florian Schintke, and Alexander Reinefeld. Range queries on structured overlay networks. *Computer Communications*, Vol. 31, No. 2, pp. 280–291, 2008.