

スーパーコンピュータを用いた大規模グラフ解析と Graph500 ベンチマーク

藤澤 克樹

九州大学 マス・フォア・インダストリ研究所

[アブストラクト]

新しいスーパーコンピュータの応用として大規模なグラフ解析が注目を集めている。新しいスーパーコンピュータ(スパコン)の応用として大規模なグラフ解析やデータ処理が注目を集めている。グラフ解析の応用分野としては大規模災害等での避難誘導計画、社会公共政策や企業経営等のためソーシャル・ネットワーク等の大規模データの有効活用等が想定されているが、非常に計算量やデータ量さらに電力使用量などの規模が大きく従来の手法では処理が困難である。そのためハイパフォーマンスコンピューティング分野の技術を用いて大規模なグラフ解析を行う研究がさかんに行われており、その中から Graph500 ベンチマークとグラフ解析の性能について解説を行う。これまで国内のスパコン(京コンピュータ、東工大 Tsubame, 東大 FX10 等)を活用して、以下の図 2 のように京コンピュータで世界 1 位(第 8 回 Graph500 ベンチマーク)等の優れた成果を挙げた。

[キーワード] 最適化問題, グラフ解析, Graph500 ベンチマーク, HPC, オープンソース

1. はじめに

新しいスーパーコンピュータの応用として大規模なグラフ解析が注目を集めている。グラフは点集合と枝集合から構成される。具体的には図 1 のように様々な応用分野において解析の対象とする事象の関係(Relationships)を点と枝で表現していく。例えば道路交通ネットワークでは点は交差点、枝は交差点間の道路に該当する。また Twitter などのソーシャル・ネットワークの解析では、点はユーザ、枝はユーザ間のフォロー関係(あるいはメッセージ送信)などに関連させることが多い。さらに各枝を連結させてグラフを構成して(Step 1)、目的に応じて最短経路検索などのグラフ解析を行う(Step 2)。またグラフ解析の結果は元の応用問題の分析や理解のために使用される(Step 3)。実際にカーナビゲーションシステムでは道路ネットワークがグラフデータとして内蔵されていて、ユーザの指示に応じて出発地点と目的地点間の最短経路検索を行っている。このように社会における実データをグラフデータに変換して、計算機で高速処理する需要が高まっている。

2. 大規模グラフ探索と Graph500 ベンチマーク

スパコンのベンチマークテストでは Top 500 が有名である。Top 500 では主に数値計算能力が測定される。今日ではスパコンの応用が幅広い分野に及び、交通ネットワーク分析やソーシャル・ネットワーク解析で注目されている大規模グラフ解析などのビッグデータ処理用途においては Top 500 の結果を用いて計算機評価を行うことが難しくなっている。そのため、スパコンの大規模データ処理性能を計測する Graph 500 (注 1) ベンチマークテストが 2010 年より開始された。また、Green Graph500 (注 2) は Graph500 と同じくグラフ探索性能を競うベンチマークテストであり、昨今の厳しい電力事情を考慮すると、性能と同時に省電力性も極めて重要であり、今回の省電力グラフ探索の技術の多方面への応用が期待されている。

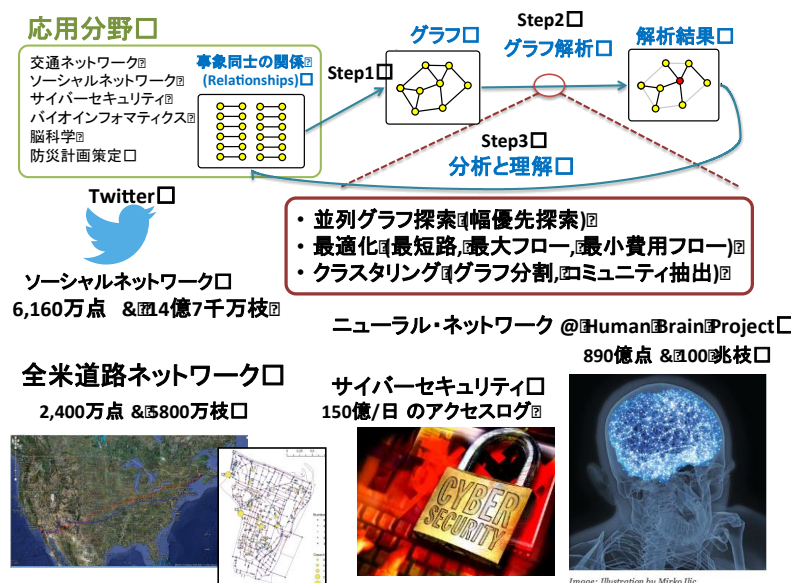


図 1: グラフ解析の利用方法と応用分野

共同研究チームは、次世代のスパコン上で大規模なグラフの高速な探索処理を行うソフトウェアの開発を2011年から進めてきた。冗長なグラフ探索を削減するアルゴリズムの利用、数千~数万台規模が高速なネットワークで接続された超並列計算機上での通信性能の最適化、マルチコアプロセッサ上でのメモリへのアクセス最適化、フラッシュデバイス利用による大規模グラフ処理と高性能化の両立などの先進的なソフトウェア技術を高度に組み合わせることにより、モバイルデバイスからスパコンまでの様々なコンピュータ上での高速かつ省電力なグラフ処理を可能にしている。最近(国際会議 ISC14, 2014年6月)、独自に開発したソフトウェアを用いて、大規模なグラフを解くことでスパコン上のビッグデータ処理を計測する Graph500、及び、その省電力性を計測する Green Graph500 ベンチマークを様々な研究機関や企業の協力と支援によって実施した結果、両者において世界第1位の高成績を達成した。当 CREST チームでは国内のスパコン(京コンピュータ、東工大 TSUBAME、東大 FX10 等)を活用して、以下の図2のように京コンピュータで世界1位(第8回 Graph500 ベンチマーク)等の優れた成果を得た。

Graph500ベンチマークにおける 当CRESTチームの成果

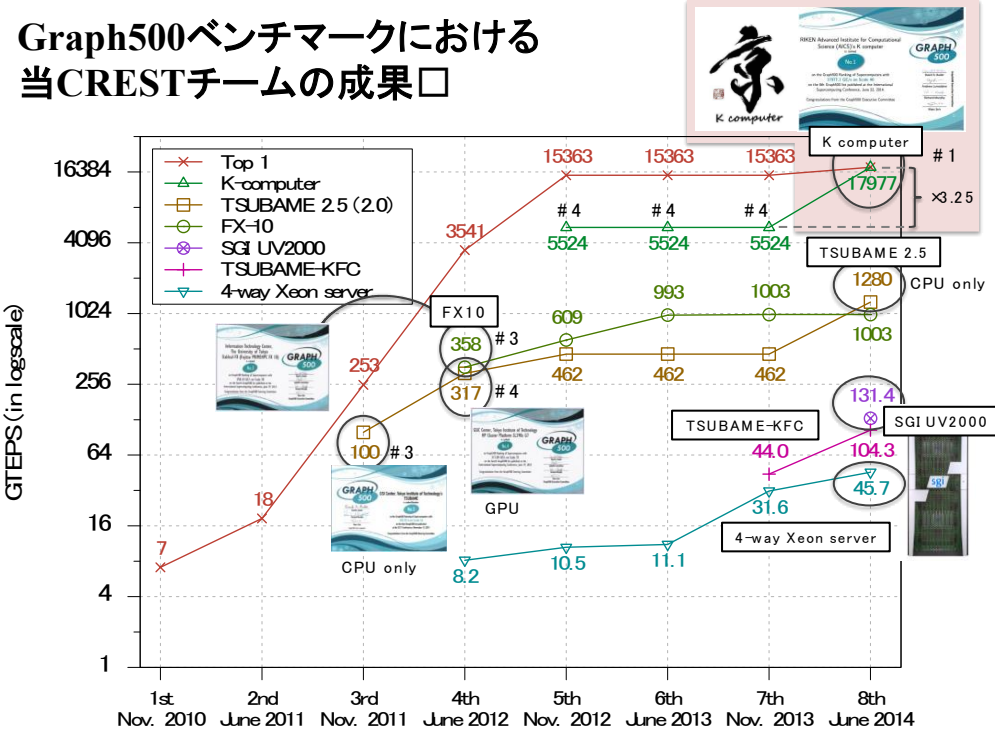


図2: Graph500 ベンチマークにおける当 CREST チームの成果

- (注1) Graph500 : 幅優先探索によって1秒間に探索したグラフのエッジ (枝) の数が指標として用いられる。具体的には単位指標は TEPS (Traversed Edges Per Second) が用いられ、TEPS 値が高ければ、高速にグラフ探索を行うことができることを意味する。Web サイト: <http://www.graph500.org/>
- (注2) Green Graph500 : 単位指標は TEPS / W (Traversed Edges Per Second / Watt) が用いられ、TEPS / W 値が高ければ、単位消費電力あたりのグラフ探索性能が高い、つまり省電力性能が高いことを意味する。Green Graph500 では処理する対象となるグラフの規模に応じてビッグデータ部門とスモールデータ部門に分かれている。Web サイト: <http://green.graph500.org/>