アクティブ TCP コネクションを用いたインラインネットワーク計測

Inline Network Measurement with Active TCP connections

Cao Le Thanh Man　　長谷川 剛　　村田 正幸

大阪大学情報科学研究科

1　はじめに

Akamai 等の CDN サービスにおいては、複数台の実 Web サーバと、プロキシ (キャッシュ) サーバがネットワーク内に配置される。Web クライアントからプロキシサーバに対してドキュメントの転送要求が発生した場合、プロキシサーバは、自身のドキュメントキャッシュに要求されたドキュメントが存在すればそれを転送し、存在しなければ対応する実 Web サーバからそのドキュメントを取得し、Web クライアントへ転送する (この時発生するトラヒックをここではフォアグラウンドトラヒックと呼ぶ)。プロキシサーバは、ミスキャッシュ転送に加えて、Web クライアントが近い将来に要求すると考えられるドキュメントをあらかじめ実 Web サーバから取得 (プリフェッチ) し、キャッシュに保存する (プリフェッチ動作によって発生するトラヒックをバックグラウンドトラヒックと呼ぶ)。フォアグラウンドトラヒックはキャッシュミスによって発生するため、その転送はできるだけ早く完了する必要がある。そのため、バックグラウンドトラヒックによってフォアグラウンドトラヒックが影響を受け、転送速度が低下することを避ける必要がある。

上記の問題に対する解決方法として、エンドホスト間 (CDN では実 Web サーバ/プロキシサーバ間) のパス上において利用可能な帯域を計測し、計測結果に基いてフォアグラウンドトラヒックに悪影響を与えないようにバックグラウンドトラヒックの転送速度を調整することが考えられる。例えば、計測された利用可能帯域を基に、バックグラウンドトラヒックを転送している TCP コネクションの最大ウィンドウサイズを設定することで、バックグラウンドトラヒックが不要に高いレートで転送されることを防止することができる。しかし、既存の利用可能帯域計測方式 [1-3] は、計測に長い時間がかかる、多くの計測用のパケットを用いるため外部トラヒックに与える影響が大きいなどの特徴を持つ。CDN 等のサービスオーバレイネットワークにおいては、常に最新の利用可能なネットワーク資源量をネットワーク内の他のトラヒックに悪影響を与えることなく取得することが重要であるため、既存の方式をそのまま適用することはできない。

そこで本稿では、サービスを提供しているエンドホスト間の TCP コネクションを直接用いて、データ転送中に得られる情報からエンドホスト間の利用可能帯域を随時推測するインラインネットワーク計測方式の提案を行う。この方式により、計測用のパケットをネットワーク内に送出することなく計測を行うことができるため、計測負荷を最小限に抑えることができる。まず、少ない計測パケット数で計測の初期段階から利用可能帯域の計測結果を導出することのできる、利用可能帯域の計測方式を提案する。また、提案した利用可能帯域の計測方法を、データ転送中の TCP コネクションを用いて行うインライン計測方式に関しても検討を行う。

2　提案方式

提案方式は、送信側エンドホスト、受信側エンドホスト間の現在の利用可能帯域値 A を導出する。提案方式においては、まず送信側エンドホストが計測パケットを送出し、受信側エンドホストは受信した計測パケットをそのまま送信側エンドホストへ返送する。送信側端末は返送されたパケットの到着間隔から利用可能帯域の推測を行う。

提案方式において利用可能帯域を計測する際には、図 1 に示すように、現在の利用可能帯域値が含まれると考えられる帯域の上限と下限を設定し、その間の探索区間の中から利用可能帯域を探す。探索区間を設定することで、不必要に高いレートで計測パケットを送出することを避けることができるため、外部トラヒックに与える影響を最小限に抑えることができ、計測パケット数を現象させることができる。提案する利用可能帯域計測アルゴリズムにおける、送信側エンドホストの動作概略を以下に示す。図 1 は、計測アルゴリズムの概略を示している。詳細については [4] を参照されたい。

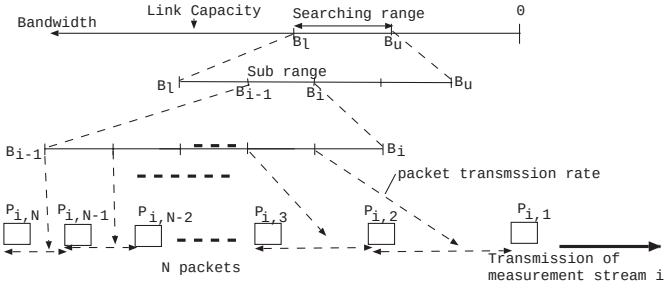
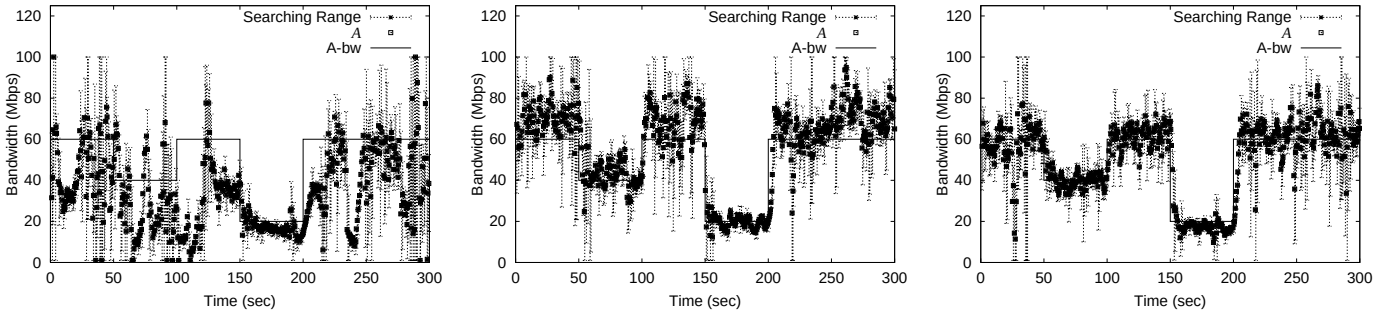


図 1: 計測アルゴリズム

1. **初期探索区間の決定**
   まず、Cprobe のアルゴリズムに基づい初期探索区間を設定する。
2. **探索区間の分割**
   探索区間を、大きさの等しい k 個の小区間に分割する。
3. **計測ストリームの送出及びパケット間隔の比較**
   分割した k 個の小区間それぞれに対して、計測ストリームをネットワーク内に送出する。その際、1 つの計測ストリームで様々な送信レートに対する計測を行うことで、計測パケット数を少なくし、短時間で計測結果を導出する。また、送出したストリームの受信結果から、PathLoad のアルゴリズムを用いて、パケット間隔の増加傾向を調べる。
4. **小区間の選択**
   全ての計測ストリームのパケット間隔の増加傾向から、現在の利用可能帯域値が含まれると考えられる小区間を選択する。
5. **利用可能帯域を算出**
   ステップ (4) において探索区間内のある小区間内に利用可能帯域が存在すると判断された場合は、線形回帰法を用いて利用可能帯域を導出する。
   探索区間内に利用可能帯域が存在しないと判断された場合には、利用可能帯域が大きく変化していると判断し、探索区間の両端の値を利用可能帯域とする。
6. **探索区間の再計算を行い、(2) へ戻る**
   今回の計測結果と、前回までの計測結果を用いて、次回の計測で用いる探索空間を決定し、ステップ (2) へ戻る。

3　評価結果

本章では、3 章で提案した利用可能帯域推測方式を、ns を用いたシミュレーションによって評価した結果を示し、提案方式の有効性を検証する。シミュレーションでは、帯域が 100 Mbps、伝搬遅延時間が 30 msec のボトルネックリンク上に、ルータを介して背景トラヒック (Cross traffic) を発生させる送受信エンドホスト、および利用可能帯域の計測を行う送受信エンドホスト (Sender、Receiver) が接続されているネットワークを用いる。各ホストとルータ間のリンクは全て帯域が 100 Mbps、伝搬遅延時間が 30 msec である。背景トラヒッ



(a) N=3 (b) N=5 (c) N=8

図 2: シミュレーション結果

クは、指定したレートで UDP によるデータ転送を行うことで生成している。

図 2 は、背景トラヒック量を時間によって変動させ、ボトルネックリンクの実際の利用可能帯域 (A-bw) が 0 sec から 50 sec までが 60 Mbps、50 sec から 100 sec までが 40 Mbps、100 sec から 150 sec までが 60 Mbps、150 sec から 200 sec までが 20 Mbps、200 sec から 300 sec までが 60 Mbps と変化した時の、利用可能帯域の計測結果 (A) およびその時の探索区間の大きさ (Searching Range) を示している。図 2(a)- 2(c) は、1 つの計測ストリーム内の計測パケット数 N をそれぞれ 3、5、8 とした時の結果である。図から、1 つの計測ストリーム内の計測パケット数が少ない場合には、計測の精度が低下し、利用可能帯域の推定がうまく行えていないことがわかる。またその反面、計測パケット数が多くなると、計測精度が向上しており、急激な利用可能帯域の変化がある場合にも、素早く対応して新たな計測結果を導出している。これは、計測パケットが少なくなると、アルゴリズムのステップ (3) におけるパケット間隔の増加傾向の判断が困難になり、ステップ (4) における適切な小区間の選択が不正確になるためである。しかし、N が 5 以上であれば、利用可能帯域をほぼ正確に計測することができている。しかし、計測に必要な計測パケット数 N は、さまざまなネットワーク状況 (利用可能帯域の大きさ、変動の大きさ、背景トラヒックの性質等) によって変化すると考えられる。適切な N の設定方法に関しては今後の課題としたい。

4 TCP コネクションによるインライン計測

本章では、アクティブな TCP コネクションを用いたインライン計測を実現する際に、問題となる点を挙げ、その解決方法に関する指針を示す。

4.1 ウィンドウサイズ

TCP コネクションが一度に送出できるデータパケット数は、ウィンドウサイズ W によって制限される。したがって、$W < N$ (N は提案方式における 1 つ計測ストリーム内のパケット数) の場合には計測を行うことができない。しかし、3 章で示したシミュレーション結果から、$N = 5$ 以上であれば計測を行うことが可能であることが明らかとなったため、転送データサイズが小さく、ウィンドウサイズがあまり大きくならない場合においても、計測を行うことが可能であると言える。

また、$N < W$ の場合には、(1)1 つの計測ストリーム内のパケット数を増加させる、(2) パケット数はそのままで、計測ストリーム数を増加させる、ということが考えられる。(1) は対応する小区間で用いる計測パケット数が増加するため、その小区間の計測精度が向上する。(2) は、1 ラウンドトリップ時間で送信する計測ストリーム数が増加するため、計測速度が向上する。提案するインライン計測方式においては、W 個のパケットから、それぞれ N 個のパケットから成る $\lfloor W/N \rfloor$ 個の計測ストリームを生成し、1 ラウンドトリップ時間で送出することで、計測速度の向上を優先させる。

4.2 受信側 TCP の Delayed ACK オプション

Delayed ACK オプションを用いる受信側 TCP は、1 個のデータパケットを受けとる毎に ACK パケットを送出するのではなく、2 個のデータパケットを受けとる毎に ACK パケットを 1 個送出する。3 章で示したアルゴリズムは、計測パケットが全て返送されることを前提としているため、受信側 TCP が Delayed ACK オプションを用いる場合には、そのまま適用することができない。

この場合には、提案アルゴリズムのステップ (3) で行う計測パケットの送出間隔と到着間隔の比較を、1 パケット毎に行うのではなく、2 パケット毎に行う必要がある。しかし、これにより 1 つの計測ストリーム内の計測パケット数が N から $\lfloor N/2 \rfloor$ に減少するため、計測誤差が大きくなる。したがって、N を大きくする、あるいは 1 つの小区間に対して複数の計測ストリームを用いる等の修正が必要になると考えられる。

4.3 提案方式のパラメータ設定

2 章で提案した計測方式は、探索区間を小区間に分割する際の分割数 k をパラメータとして持つ。3 章におけるシミュレーションでは、探索区間の大きさに応じて分割数を変化させている。しかし、TCP コネクションによるインライン計測を行う場合は、k は探索区間の大きさだけではなく、TCP コネクションの現在のウィンドウサイズを考慮して決定する必要がある。そこで、ウィンドウサイズ W が大きく、1 ラウンドトリップ時間で送出することのできる計測ストリーム数 $\lfloor W/N \rfloor$ が十分大きい場合には、$k = \lfloor W/N \rfloor$ とする。これにより、探索区間内の小区間数が大きくなるため、計測の精度が向上する。また、1 ラウンドトリップ時間で 1 回の計測が完了するため、ネットワーク状況の変化に対応しやすくなる。

また、探索区間の最小幅を、十分な計測精度が得られる小区間の大きさと、k の値を基に設定することにより、計測精度が高い場合にはより広い探索区間を用いることが可能となり、大きなネットワークの変動に対応することができる。計測精度の算出方法等、詳細については今後の課題としたい。

謝辞

本研究の一部は、総務省戦略的情報通信研究開発推進制度における特定領域重点型研究開発プロジェクト「ユビキタスインターネットのための高位レイヤスイッチング技術の研究開発」、および及び平成 13 年度文部科学省科学研究費奨励研究 (A)(13750349)によっている。ここに記して謝意を表す。

参考文献

[1] R. L. Carter and M. E. Crovella, “Measuring bottleneck link speed in packet-switched networks,” Tech. Rep. TR-96-006, Boston University Computer Science Department, Mar. 1999.

[2] B. Melander, M. Bjorkman, and P. Gunningberg, “A new end-to-end probing and analysis method for estimating bandwidth bottlenecks,” in *Proceedings of IEEE GLOBECOM 2000*, Nov. 2000.

[3] M. Jain and C. Dovrolis, “End-to-end available bandwidth: Measurement methodology, dynamics, and relation with TCP throughput,” in *Proceedings of ACM SIGCOMM 2002*, Aug. 2002.

[4] Cao Le Thanh Man, 長谷川 剛, 村田 正幸, “サービスオーバーレイネットワークのためのインラインネットワーク計測に関する一検討,” **電子情報通信学会技術研究報告**, Jan. 2003.