インラインネットワーク計測:
TCPを用いたエンド間パスの帯域推定に関する研究

Inline Network Measurement:
TCP Built-in Techniques for Inferring End-to-end Bandwidth

大阪大学大学院情報科学研究科
村田研究室
CAO LE THANH MAN

学位論文公聴会

2006年 12月 4日

内容

- 研究の背景
  - ネットワーク帯域計測の必要性
  - 既存計測手法の欠点
- インラインネットワーク計測
  - インライン計測の概念
  - 関連研究
- 提案手法
  - 利用可能帯域の計測
  - 物理帯域の計測
  - 高速ネットワーク環境における計測
- まとめ

研究の背景

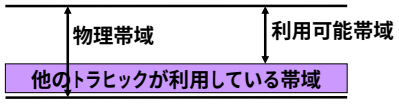
背景

- インターネットを用いたサービスの普及
  - 人々の生活に必要不可欠なものになりつつある
    - ✓World Wide Web
    - ✓インターネットを用いた電話･テレビ会議
  - インターネットの資源状態がサービスの品質を左右する
- インターネットはベストエフォート型
  - パケット交換ネットワーク
  - 資源の保証はできない
  - 事前に資源の利用状況を知ることもできない
- サービス品質向上のためには、ネットワーク資源状態の把握が重要
  - 空き帯域、パケット損失率、伝播遅延…

研究の背景

帯域の概念

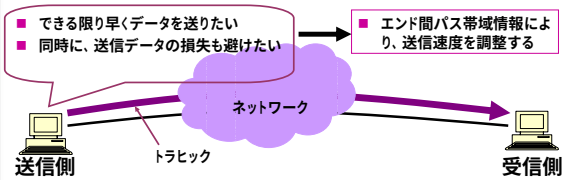
- ネットワークの資源状況を示す重要な指標
- 物理帯域
  - 最大でデータを送信できる速度 (物理的な上限)
    - ✓ADSL: 10 Mbps, 24 Mbps
- 利用可能帯域
  - 利用されていない帯域



エンド間帯域: エンド間パスの中でもっとも小さいリンク帯域

研究の背景

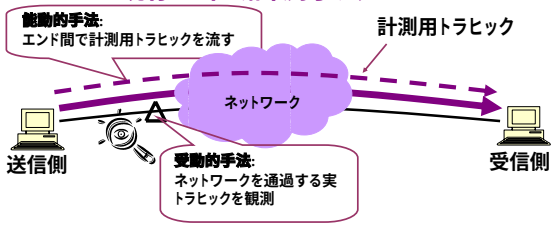
帯域情報の必要性



- 高度なトランスポートプロトコル
  - 高速転送、バックグラウンド転送、一定スループットの確保、…
- ルーティング
  - サーバ選択、経路選択
- ネットワーク障害場所の特定
- ISPの課金システム

研究の背景

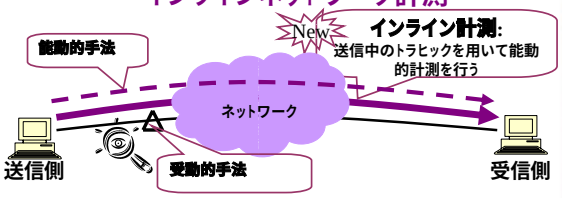
既存の帯域計測手法



- 受動的手法 (Nettimer, ABEst, PPrat…)
  - ネットワークに影響を与えない
  - 計測データを集めるのに時間がかかる
- 能動的手法 (Bprobe/Cprobe, Pathload, CapProbe…)
  - 正確に素早く計測結果を導出
  - 計測用トラヒックがネットワークの負荷となる

インライン計測

提案手法: インラインネットワーク計測



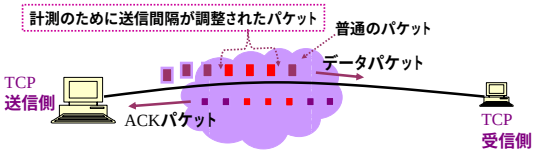
- 送信中のトラヒックを計測に用いる
  - 能動的手法、受動的手法双方の利点を持ち、欠点を克服
    - ✓計測結果を正確に、かつ素早く導出
    - ✓ネットワークへ影響を与えない
  - トランスポートプロトコルの動作を妨げないように工夫が必要

7

インライン計測

インライン計測の仕組み

- TCP: Transmission Control Protocol
  - 現在もっとも普及しているトランスポートプロトコル
  - 送信側がデータを複数のパケットを用いて受信側に送る
  - 受信側はパケットを受信時に確認パケット (ACK)を返す



- インライン計測の仕組み
  - TCP送信側で動作
  - 数個のデータパケットの送信間隔を調整
  - 対応ACKパケットの到着間隔からTCP送受信間の帯域を推定

8

インライン計測

インライン計測の関連研究

- 従来のTCP
  - パケット損失、パケットの転送遅延情報を監視し、送信レートを調整 (Reno TCP (‘90), Vegas TCP (‘94)…)
- TCPを計測ツールに変える手法
  - Sting (パケットロス, ‘99)、Sprobe (物理帯域, ‘01)、Abget (利用可能帯域, ‘04)
- TCPで受動的計測
  - TCP Westwood (‘01), TCP-Rab (‘04):利用可能帯域
- TCPで能動的計測
  - TCP Probe (‘05): 物理帯域のみ
- 提案手法: TCPで能動的に計測 (‘03〜)
  - 物理帯域、利用可能帯域を同時に計測する手法
  - 高速ネットワーク環境における計測手法

9

インライン計測

博士論文の構成

- Chapter 1 Introduction
- Chapter 2 ImTCP: ImTCP with an Inline Measurement Mechanism for Available Bandwidth
  - 研究1: 利用可能帯域のインライン計測手法
- Chapter 3 A Simultaneous Inline Measurement Mechanism for Capacity and Available Bandwidth
  - 研究2: 物理帯域のインライン計測手法
- Chapter 4 Inline Bandwidth Measurement Techniques for Gigabit Networks
  - 研究3: 高速ネットワーク向けインライン計測手法
- Chapter 5 Conclusions

10

研究1: 利用可能帯域計測

アプローチ

- TCP内で利用できる計測アルゴリズム
  - 既存の計測アルゴリズムから相応しいものを選ぶ
    - ✓PathLoad [1]
  - TCPに導入するための修正
    - ✓一回の計測に用いるパケット数を減らす
    - ✓精度よりも速度と頻度を重視
      - 利用可能帯域の時間的な変化を反映したい
- TCPに計測アルゴリズムを導入
  - 導入方法
  - パラメータ調整

[1] M. Jain, C. Dovrolis, “End-to-end available bandwidth: Measurement methodology, dynamics, and relation with TCP throughput”, *IEEE/ACM Transactions on Networking*, 11(4):537-549, August 2003.

11

1. 利用可能帯域計測

既存手法: PathLoad

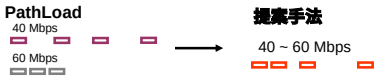
- 少しずつストリームの送信レートを上げる
  - 計測ストリーム: 連続して転送する複数のパケット
- ストリームの到着間隔から利用可能帯域を推定
  - 送信レート ＞ 利用可能帯域なら
    - ✓パケットの到着間隔が伸びる (大きくなる)

12

1. 利用可能帯域計測

提案計測アルゴリズム

- 一つのストリーム内で送信レートを変化させる
  - 一つのストリームで広い帯域区間を探索する
- ストリーム内のパケット数を減らす (100 から 5 に)

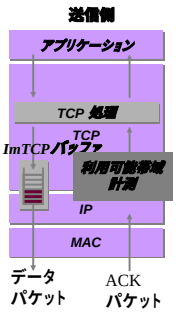


- 探索区間を導入
  - 以前の計測結果の統計データから今回の計測結果が含まれる確率が高い帯域幅の区間を絞る
  - 一回の計測に用いるストリーム数を少なくできる
    - ✓ 120 から 2～4に
  - 不必要に高いレートのストリームを送信することがない
    - ✓ 周りのトラヒックに迷惑を与えない

1. 利用可能帯域計測

TCPへの導入

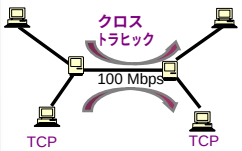
- ImTCP: Inline measurement TCP
  - Reno TCPへ計測アルゴリズムを適用
- 計測プログラム
  - TCPレイヤの最下層に挿入
  - ImTCPバッファ
    - ✓ 計測時、TCPパケットを一旦溜める
    - ✓ パケットの送信間隔を調整し、計測ストリームを構成
    - ✓ 計測しないときはパケットをそのまま通す
  - ACKパケットの到着間隔を監視して、利用可能帯域を計算する
- 計測オーバヘッドは小さい
  - 送信マシンの平均CPU負荷はほとんど増加しない [2]
    - ✓ ImTCP: 19.12%, RenoTCP: 18.62%



[2] T. Tsugawa, G. Hasegawa and M. Murata, "Implementation and evaluation of an inline network measurement algorithm and its application to TCP-based service," in *Proceedings of 4th IEEE/IFIP Workshop on End-to-End Monitoring Techniques and Services*, April 2006

1. 利用可能帯域計測

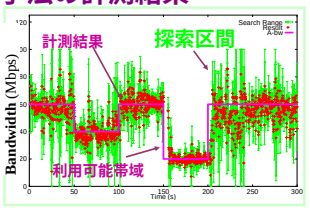
提案手法の計測結果

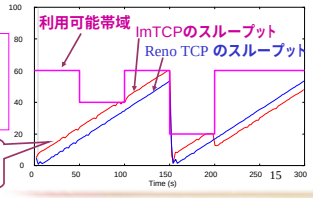


クロストラヒックの転送レートを変えることで、利用可能帯域を変える

- 計測結果が利用可能帯域の変化を反映できる
- ImTCPがReno TCPと同じぐらいの転送性能を持つ
  - 計測はTCPの動作を妨げない

TCP スループットが利用可能帯域に達していない場合でも計測できる





研究2: 物理帯域計測

研究の目標

- ImTCPに物理帯域計測機能を追加
  - 利用可能帯域と同時に、物理帯域を計測可能

■ImTCP

利用可能帯域計測機能

物理帯域計測機能 New

2. 物理帯域計測

既存の物理帯域計測アルゴリズム

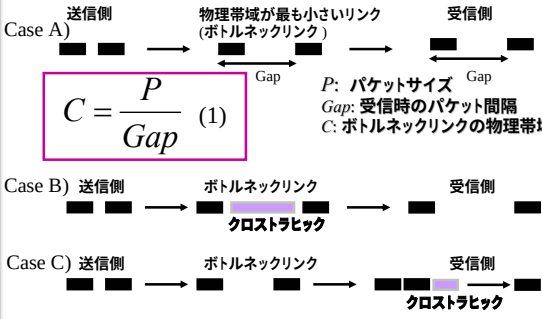
- 1パケットを用いる手法
  - 様々なサイズのパケットを利用
  - Packet Tailgating
- Hop-by-hop
  - 様々な生存時間のパケットを利用
  - Pchar, Pathchar
- パケットペア
  - 様々な間隔を持つパケットのペアを利用
  - CapProbe, Pathrate

TCPで利用可能

Question: 従来のパケットペア手法よりさらによい手法は？
- ImTCPの利用可能帯域情報を使うことにより実現

2. 物理帯域計測

パケットペアを用いた計測の原理及びその問題点



$$C = \frac{P}{Gap} \quad (1)$$

P: パケットサイズ
Gap: 受信時のパケット間隔
C: ボトルネックリンクの物理帯域

Case B), Case C)の場合、式 (1) が不正確な値を出す

2. 物理帯域計測

提案手法

- 既存手法
  - Case A) の パケット ペアのみを使う

$$C = \frac{P - \delta A}{Gap - \delta} \quad (2)$$

- 提案手法
  - 利用可能帯域の値を用いてクロストラヒックの影響を推定
  - Case A) , Case B) のパケットペアの両方を使う

$$C = \frac{P + L}{Gap}$$

L: クロストラヒック量の平均

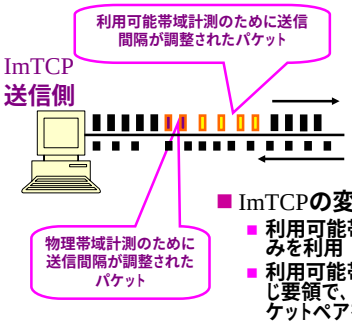
$$L = \delta\,(C - A)$$

A: 利用可能帯域

δ: パケットの送信間隔

2. 物理帯域計測

提案手法を ImTCPに導入



- ImTCPの変更は少ない
  - 利用可能帯域計測のための仕組みを利用
  - 利用可能帯域計測ストリームと同じ要領で、物理帯域計測用のパケットペアを作成

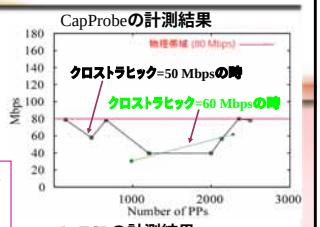
2. 物理帯域計測

シミュレーション評価結果

- 物理帯域: 80 Mbps
- クロストラヒック
  - 平均: 60Mbps

- 比較手法: CapProbe [3]をTCPに導入した手法
  - クロストラヒックが多い場合に結果が不正確
    - ✓ Case A) のパケットが少ないため
- 提案手法
  - クロストラヒックが多い場合でも正確な計測が可能
  - 計測結果に統計的な信頼区間を与えることができる

CapProbeの計測結果



ImTCPの計測結果

[3] R. Kapoor, L. Chen, L. Lao, M. Gerla and M. Sanadidi, "CapProbe: A simple and accurate capacity estimation technique", in *Proceedings of ACM SIGCOMM*, August 2004

研究3: 高速ネットワーク向け計測

高速ネットワークにおける既存の帯域計測手法の問題点

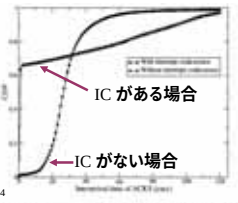
- 必要となるパケット間隔が小さい
  - 高い転送速度を実現するには、小さい送受信間隔が必要
    - ✓ 1 Gpsのリンク: 12マイクロ秒 (パケットサイズ 1500 B)
  - 小さいパケット間隔の生成・観測には、高いCPU能力が必要
    - ✓ 汎用マシンでは他の処理に影響を与える
- Interrupt Coalescence (IC, 割り込み調整機能)の影響
  - 高速パケット処理、CPUの負荷削減のために、高速 ネットワークインターフェースに導入される機能
  - 複数のパケットをひとつのグループにまとめてから、オペレーティングシステムに渡す [4]
  - パケット間隔情報を正確に読み取ることができない

[4] Intel, "Interrupt moderation using Intel Gigabit Ethernet Controllers" available at *http://www.intel.com/design/network/applnots/ap450.pdf* (2003)

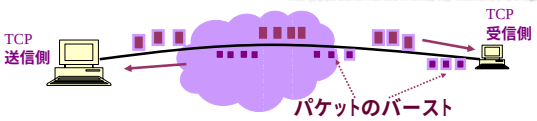
3. 高速ネットワーク向け

Interrupt CoalescenceのTCPへの影響

- バースト的なパケット転送を助長
  - 65% のパケットの到着間隔が1マイクロ秒より小さい [5]
    - ✓ 1ギガビットネットワークでの実験結果



[5] R. Prasad, M. Jain, and C. Dovrolis, "Effects of interrupt coalescence on network measurements," in *Proceedings of Passive and Active Measurement Workshop*, April 2004

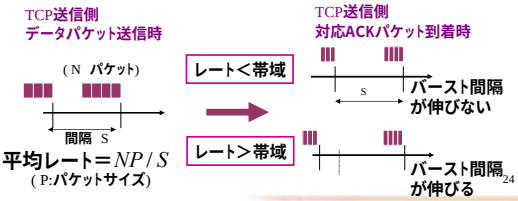


3. 高速ネットワーク向け

提案手法: ICIM

Interrupt Coalescence-aware Inline Measurement

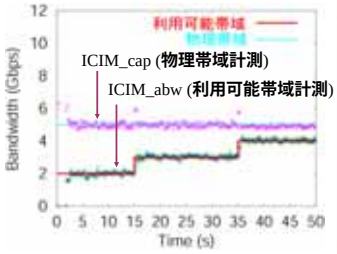
- 新しいインライン計測手法
  - 利用可能帯域 (ICIM_abw) 及び物理帯域 (ICIM_cap) を計測
  - ICが用いられる高速ネットワーク環境で動作する
  - 帯域計測にパケット間隔を用いない
- アプローチ
  - ICがもたらすTCPパケットのバーストを用いる
    - ✓ パケットバーストを大きなパケットと見なす
  - パケット間隔の代わりに、バースト内のパケット数を調整



3. 高速ネットワーク向け

シミュレーション評価

- 設定
  - 物理帯域
    - ✓5 Gbps
  - 利用可能帯域
    - ✓0-15 (s): 2 Gbps
    - ✓15-35 (s): 3 Gbps
    - ✓35-50 (s): 4 Gbps



- 結果
  - 数Gbpsのネットワーク環境で物理帯域・利用可能帯域を計測可能

25

実装コード

- 提案手法の実装コードの入手方法
  - URL: http://www.anarg.jp/imtcp/
  - FreeBSD 4.10 カーネルでの実装
    - ✓ImTCPの利用可能帯域計測機能 (‘05/3〜)
    - ✓高速ネットワーク用利用可能帯域計測機能 (ICIM_abw) (‘05/12〜)
- 利用状況
  - 2006年4月14日から11月28日までの記録
  - ImTCP: 85回のダウンロード
  - ICIM_abw: 39回のダウンロード

26

まとめ及び今後の課題

- TCPに新しい機能を提案: 計測機能
  - 利用可能帯域の計測
  - 物理帯域の計測
  - 高速ネットワーク向けの計測
- インターネットが提供しない帯域情報を容易に把握
  - 多くのインターネットトラヒックはTCPトラヒック
  - 計測ツールが不必要、計測の負荷がない、様々な場所で利用可能
- 本研究で提案したImTCPが使われている研究
  - 高速転送のためのTCP (TCP symbiosis) [6]
  - バックグラウンド転送のためのTCP (ImTCP_bg) [7]
  - 一定スループットを確保するTCP [8]
- 今後の課題
  - 他のネットワーク資源指標の計測: BTC (Bulk Transfer Capacity) など
  - インライン計測結果を用いた転送プロトコルの性能向上

[6] G. Hasegawa and M. Murata, “TCP symbiosis: congestion control mechanisms of TCP based on Lotka-Volterra competition model,” in *Proceedings of Inter-Perf 2006*, October 2006
[7] T. Tsugawa, G. Hasegawa and M. Murata, “ Background TCP data transfer with inline network measurement,” in *IEICE Transactions on Communications, Vol.E89-B, No.8, pp.2152-2160*, August 2006
[8] K. Yamanegi, G. Hasegawa and M. Murata, “ Congestion control mechanism of TCP for achieving predictable throughput, ”in *Proceedings of Australian Telecommunication Networks and Applications Conference*, December 2006.

27