

高速ネットワークにおけるインラインネットワーク計測手法の実装に関する検討および性能評価

津川 知朗[†], Cao Le Thanh Man[†]
長谷川 剛^{*}, 村田 正幸[†]

[†]大阪大学 大学院情報科学研究科
^{*}大阪大学 サイバーメディアセンター

2006/07/14

IN研究会

1

発表内容

- 研究の背景と目的
 - インラインネットワーク計測
- 計測手法のカーネルシステム実装時における問題点
 - カーネルシステムの時間粒度
 - 割り込み削減機構 (IC: Interrupt Coalescence)
 - 受信側TCPの動作
- 計測手法のアルゴリズムおよび実装指針
 - ImTCP: パケット間隔に基づく計測手法
 - ICIM: バースト間隔に基づく計測手法
- 実験ネットワークを用いた性能評価
- まとめと今後の課題

2006/07/14

2

IN研究会

研究の背景

- インターネットの発展に伴いサービスが多様化
 - e.g., CDN, P2P, Grid, IP-VPN
- IPネットワークの資源状況を把握することが重要
 - 資源を有効に利用することができるようになる
 - ネットワークサービスの品質を向上させることができる
- ★ エンドホスト間で利用可能帯域を計測する技術が注目される
- 既存の利用可能帯域計測手法に存在する問題点
 - 計測するために多くの計測用パケットを必要とする
 - ネットワークに大きな影響を与える
 - 一度の計測に長い時間を必要とする

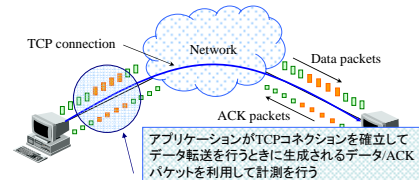
2006/07/14

3

IN研究会

インラインネットワーク計測

- 我々の研究グループが提案している計測概念
 - 計測アルゴリズムをTCPに組み込む
- インラインネットワーク計測の利点
 - 余計な計測用トラフィックを必要とせずに計測することができる
 - 送信ホストの修正のみで計測手法を実装することができる



2006/07/14

4

IN研究会

研究の目的

- インラインネットワーク計測手法には利点が存在する
 - カーネルシステムへ実装する際には問題が発生する
- カーネルシステム実装時における問題点を明らかにする
 - カーネルの時間粒度
 - 割り込み削減機構 (IC: Interrupt Coalescence)
 - 受信側TCPの動作
- 問題点に対する解決策を示す
- インラインネットワーク計測手法の実装を行う
 - FreeBSD 4.10カーネルシステムへ実装する
- 実験ネットワークを用いて実装実験を行う

2006/07/14

5

IN研究会

利用可能帯域の計測方法

- パケット間隔に基づいた利用可能帯域の計測方法



- パケット転送時にパケット間隔を調節
- パケット受信時にパケット間隔の変化を観察
- ★ 利用可能帯域を推測

2006/07/14

6

IN研究会

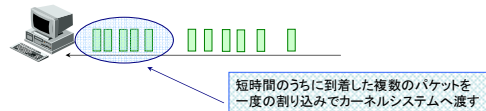
カーネルシステムの時間粒度

- **カーネルシステムの時間粒度はアプリケーションよりも荒い**
 - パケット間隔の調節も荒くなり計測精度が低下する
- **FreeBSDでは時間粒度はHZによって決定される**
 - HZによって計測可能上限値も決定される
 - HZ=100: 時間粒度10msec, 計測可能上限値※1.2Mbps
 - HZ=10,000: 時間粒度100μsec, 計測可能上限値※120Mbps
 - 計測可能上限値に近づくほど計測精度が荒くなる
- **広帯域ネットワークではHZを大きく設定する必要がある**
- **HZの値が大きすぎると実行速度に影響を与える**

※ パケットサイズが1,500Byteである場合

割り込み削減機構 (IC: Interrupt Coalescence)

- **ICはギガビットネットワーク用のNICに採用されている機構**
 - 短時間のうちに到着した複数のパケットをまとめる
 - 一度の割り込みでカーネルシステムへ渡す
- **ICはCPUへの負荷を抑えるための重要な技術**
- **カーネルシステムで観測されるパケット間隔が変化する**
 - 一度の割り込みで渡されたパケットの間隔はほとんどゼロとなる
- **パケット間隔に基づく計測手法は計測精度が低下する**



受信側TCPの動作

- **受信側TCPの遅延ACKオプションによる問題**
 - 遅延ACKオプションが有効な場合計測することができない
 - パケット間隔の変化を観測することで利用可能帯域を計測している
 - 受信したデータパケットに対して直ちにACKパケットを返す必要がある
- **遅延ACKオプションの問題を解決する必要がある**
 - 帯域計測手法をTCPへ組み込む場合
- **解決策のひとつとして考えられる方法**
 - 遅延ACKがオプションが有効な場合
 - 計測結果を利用する上位アプリケーションに計測結果が不正確であることを知らせる

インラインネットワーク計測手法の実装および実験

- **Inline measurement TCP (ImTCP) [6]**
 - パケット間隔に基づく計測手法
 - 低速ネットワークにおけるHZの設定指針を示す
- **Interrupt Coalescence-aware Inline Measurement (ICIM) [7]**
 - 高速ネットワークを想定して提案された計測手法
 - 割り込み削減機構 (IC) による問題を解決している
 - 1Gbpsを超えるような高速ネットワークでも計測できることを示す

[6] Cao Le Thanh Man, Go Hasegawa, and Masayuki Murata, "Available bandwidth measurement via TCP connection," in Proceedings of IFIP/IEEE E2EMON 2004, Oct. 2004.
 [7] Cao Le Thanh Man, Go Hasegawa, and Masayuki Murata, "ICIM: An inline network measurement Mechanism for high-speed networks," in Proceedings of IFIP/IEEE E2EMON 2006, Apr. 2006.

Inline measurement TCP (ImTCP)

- **パケット間隔に基づく計測手法**
 - データパケットの間隔を調節する
 - 対応するACK間隔の変化を観察し利用可能帯域を推測する
 - 探索区間を設定しその探索区間で利用可能帯域を検索する

計測結果を利用する上位アプリケーションに計測結果が不正確であることを知らせる

TCP層の下部にFIFOバッファを作成しTCPプロトコル処理の終了したパケットを格納する

格納したパケットを送信アルゴリズムに基づいてIP層へ渡す

データパケットの送信時刻
ACKパケットの受信時刻を記録する

Interrupt Coalescence-aware Inline Measurement (ICIM)

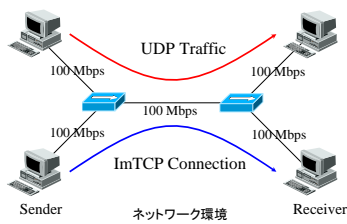
ACKパケットのバーストの到着間隔を観察する
利用可能帯域を推測する

バーストに含まれるパケット数を調節する

- パケット間隔に基づく計測手法とは異なり遅延ACKの影響を受けない

- **実装指針**
- ImTCPの実装を引き継ぎ修正を加える
 - FIFOバッファに格納されているパケット数を調節する
 - 同時にバーストに含まれるパケット数を記録する
 - 一度の割り込みで受信したACKパケットの数を記録する

低速ネットワークにおける実験

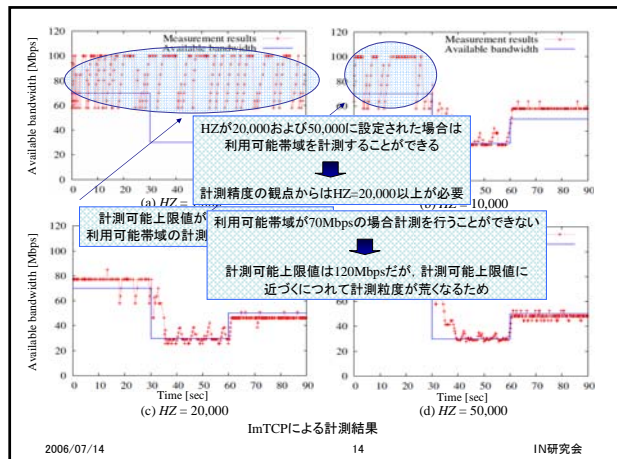


- クロストラヒックの量を時間の経過とともに変動させる
 - ボトルネックリンクの利用可能帯域を変化させる
 - 70Mbps (0 - 30sec), 30Mbps (30 - 60sec), 50Mbps (60 - 90sec)
- HZを変化させたときのImTCPの計測精度について評価する

2006/07/14

13

IN研究会

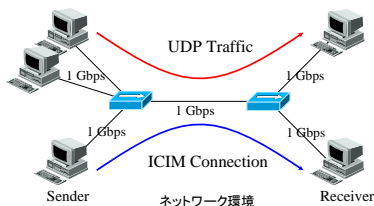


2006/07/14

14

IN研究会

高速なネットワークにおける実験

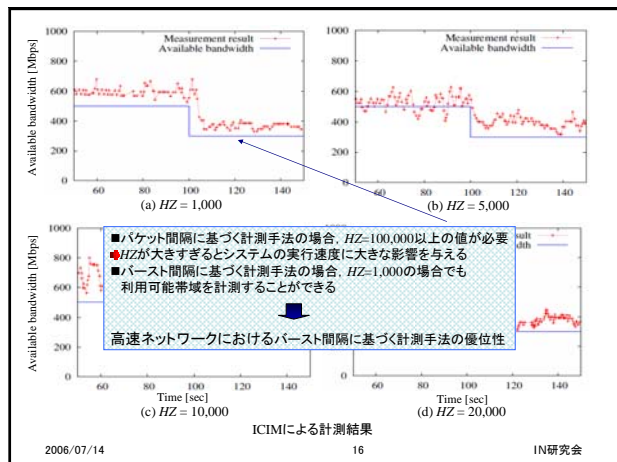


- クロストラヒックの量を時間の経過とともに変動させる
 - ボトルネックリンクの利用可能帯域を変化させる
 - 500Mbps (50 - 100sec), 300Mbps (100 - 150sec)
- HZを変化させたときのICIMの計測精度について評価する

2006/07/14

15

IN研究会



2006/07/14

16

IN研究会

まとめと今後の課題

- インラインネットワーク計測手法を実装する際の問題
 - 実装する際に問題となる点について明らかにした
 - それらに対する解決策を示した
- 実ネットワーク上での実装実験を行った
 - カーネルシステムのパラメータ設定の指針を示した
 - バースト間隔に基づく計測手法が高速ネットワークでも利用可能帯域を計測することができることを示した
- ImTCPおよびICIMの実装コードを公開しているWebサイト
 - <http://www.anarg.jp/imtcp/>
- 今後の課題
 - エンドホスト間で計測できる利用可能帯域以外の帯域情報
 - e.g., Bulk Transfer Capacity (BTC), Capacity
 - ▶ インラインネットワーク計測の概念に基づく計測手法の提案

2006/07/14

17

IN研究会

- ご清聴ありがとうございました

2006/07/14

18

IN研究会