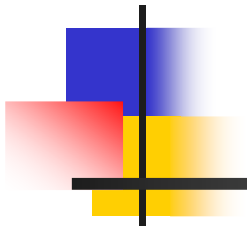


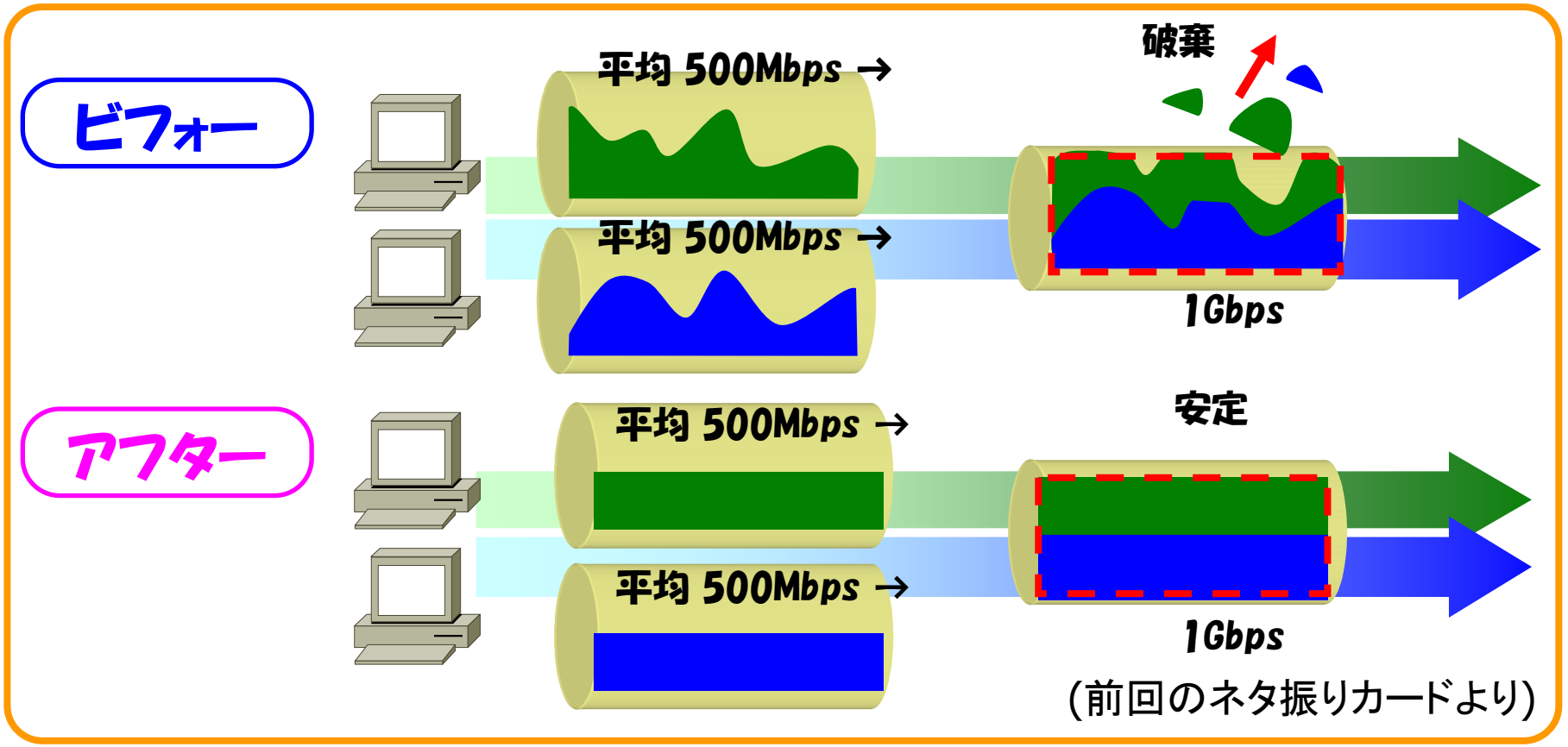
ネットワークの利用効率を高めるソフトウェア PSPacer

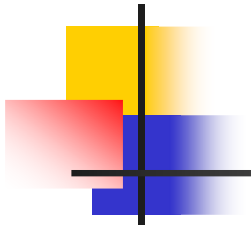
産総研グリッド研究センター/
株式会社アックス　高野了成

2005年11月25日 CELF Japan Technical Jamboree #5

ネットワークを効率よく使うには？

- 正確な帯域制御(ペーシング)が必要である
  - ストリーミング通信，高遅延環境でのTCP/IP通信



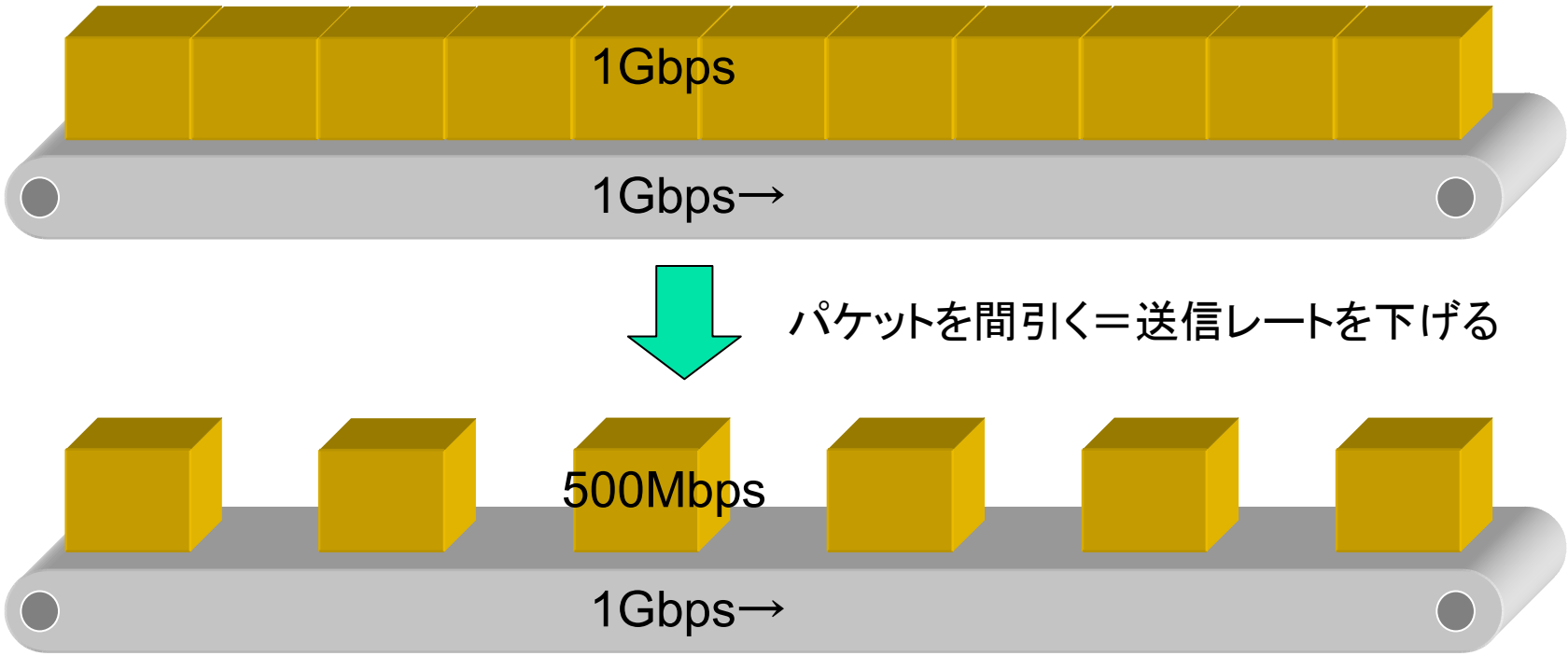


発表の流れ

- ペーシングとは？
- 既存の帯域制御方式の問題点
- ギャップパケットの提案
- PSPacerの実装
- 実験結果
- まとめ

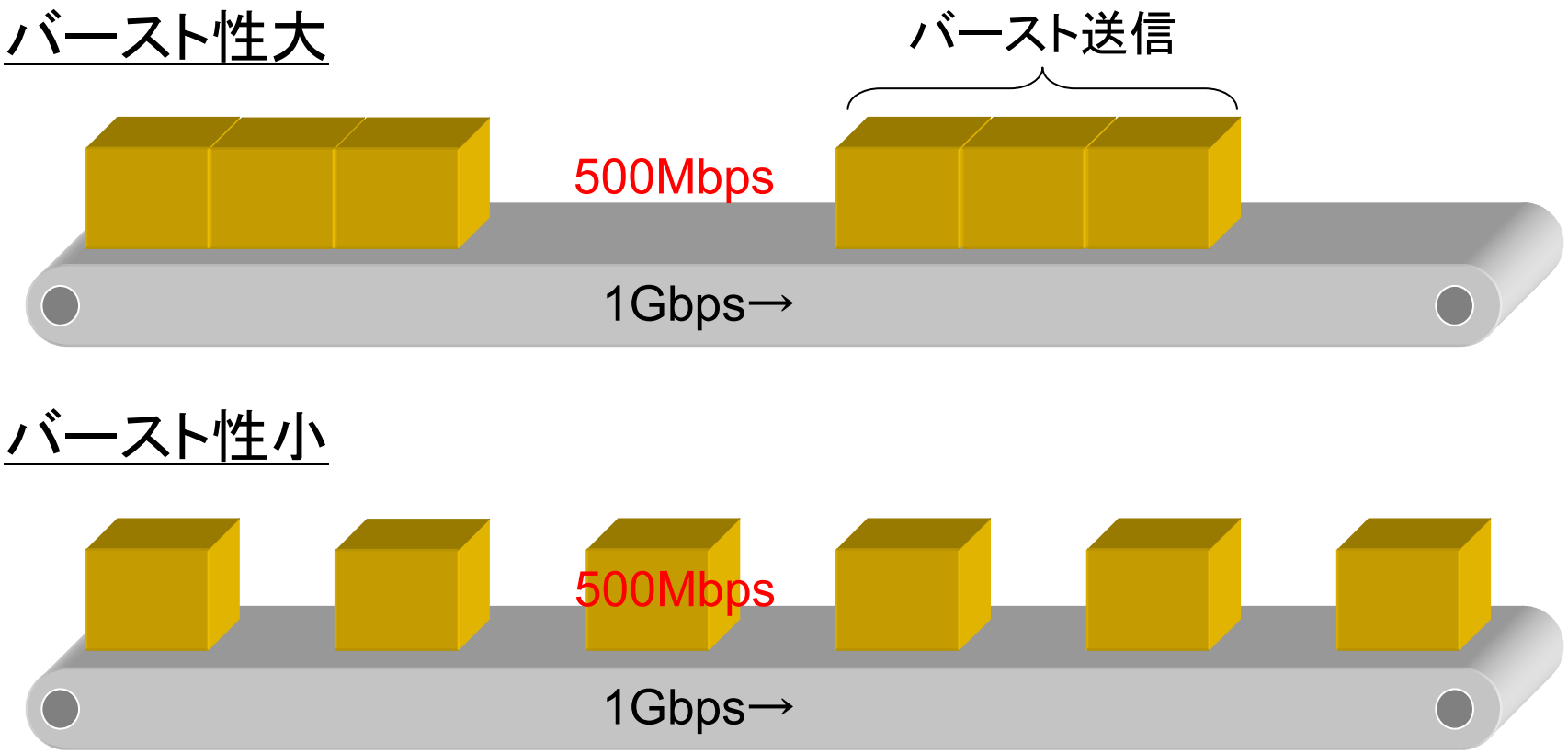
ベルトコンベアーのたとえ

- ネットワーク回線　＝　ベルトコンベアー
- パケット　＝　荷物



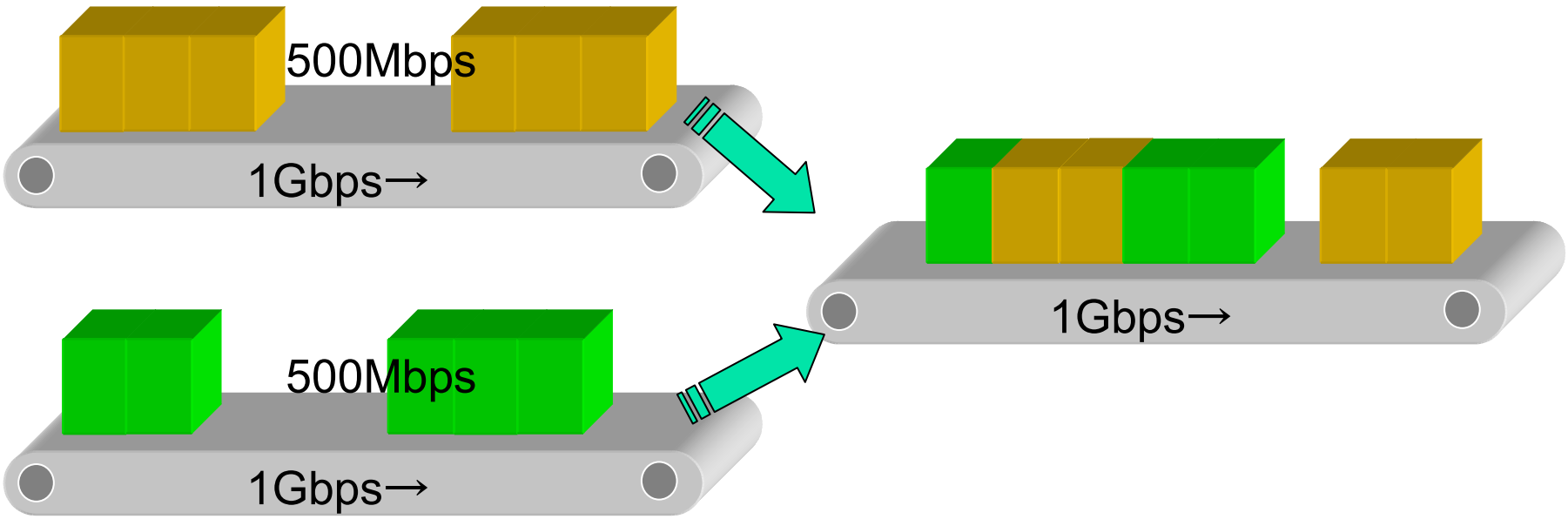
平均送信レートとバースト性

- 平均送信レートは同じでも…



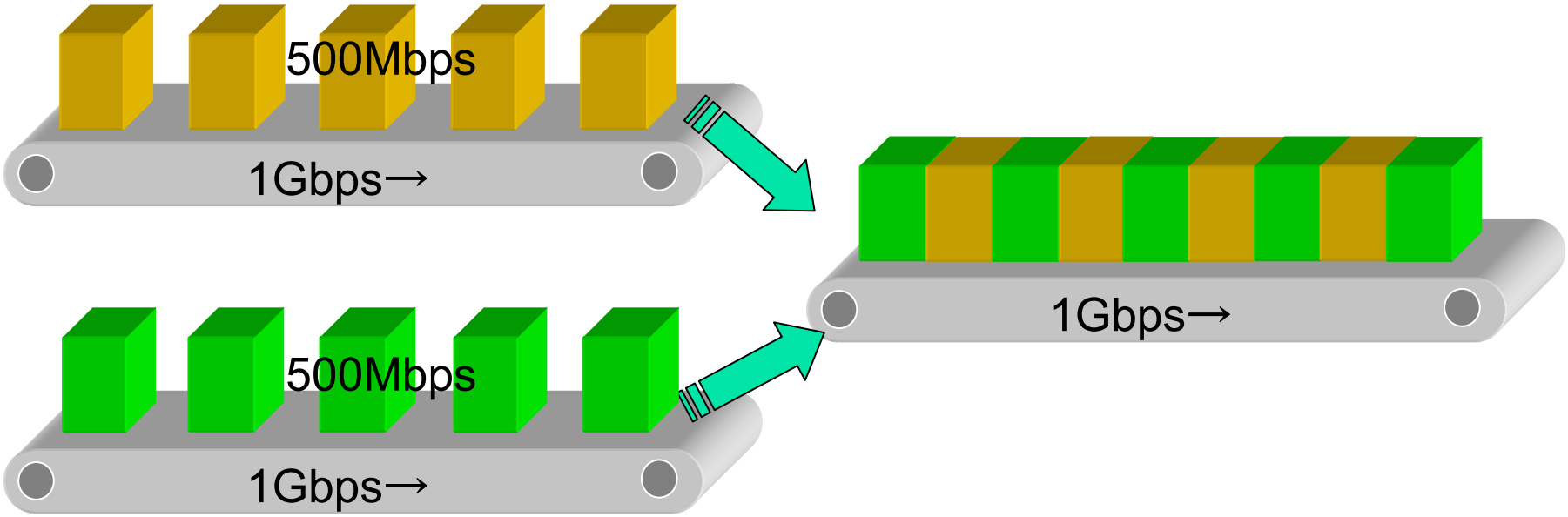
バースト送信が重なると…

- 500Mbps+500Mbpsが1Gbpsにならない！？
- 一時的に利用可能帯域を超過する
  - パケットロスの可能性



ペーシングとは？

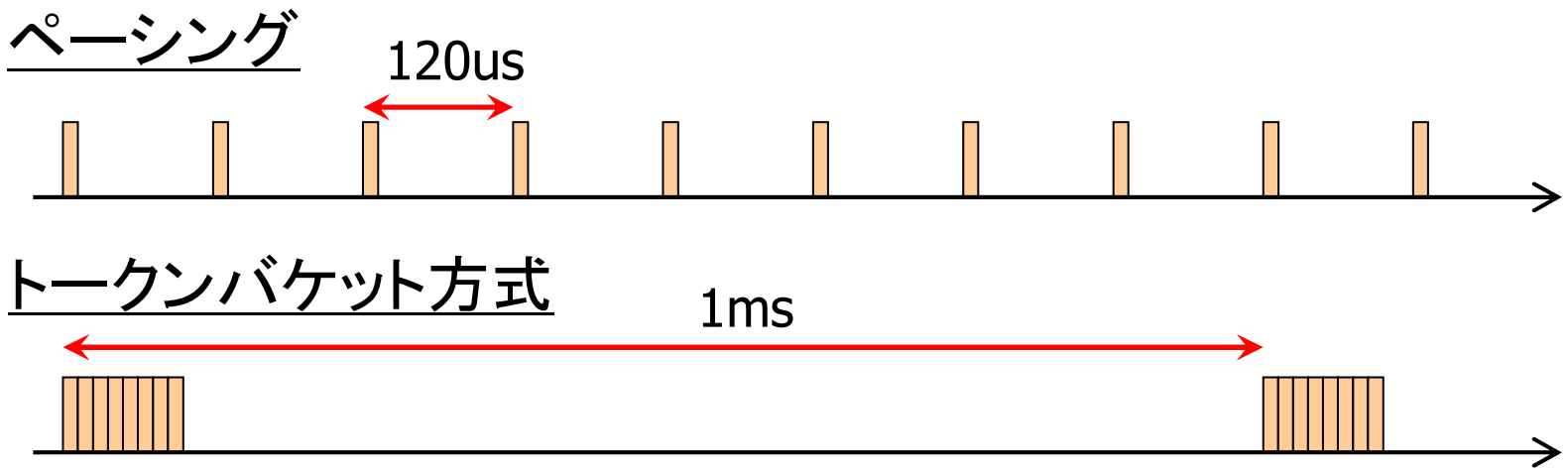
- パケット送信間隔を均一に送信する
  - 最小限のバッファで十分
  - キューイング遅延が小さい
- 正確なパケット送信間隔制御が必要



既存方式の問題点（1）

- トークンバケット方式による帯域制御では，「送信レート/HZ」のバースト送信が発生する
- パラメータの設定も面倒

GbEで100Mbpsに帯域制限した場合 (HZ=1000)



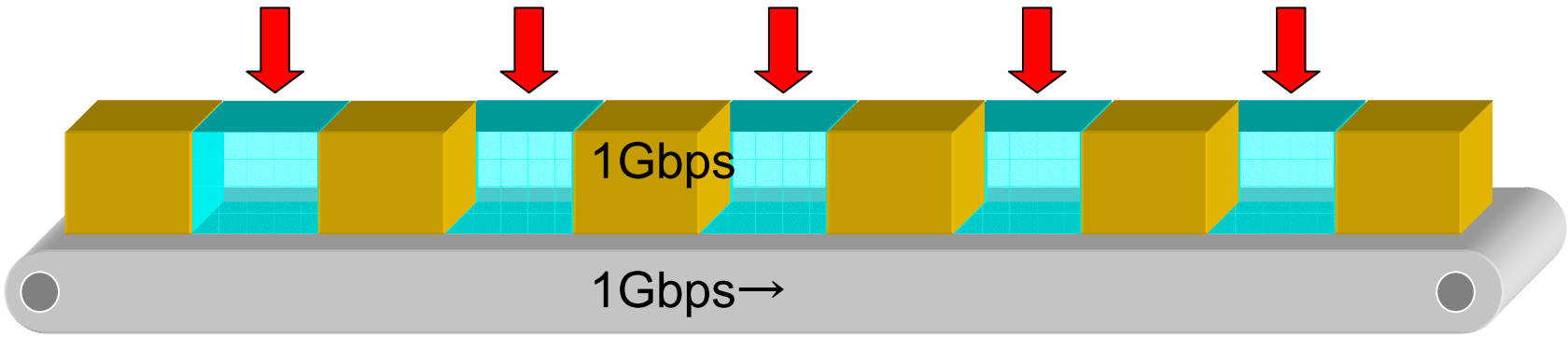
既存方式の問題点(2)

- 正確な帯域制御には高解像度のタイマが必要
  - GbEにおける1パケットの送信時間＝12μ秒
- タイマ割込みを細粒度にするとオーバヘッド大

- 専用ハードウェアを利用する(負け)
- 発想の転換
  - タイマ割込みを使わない方法はないか？

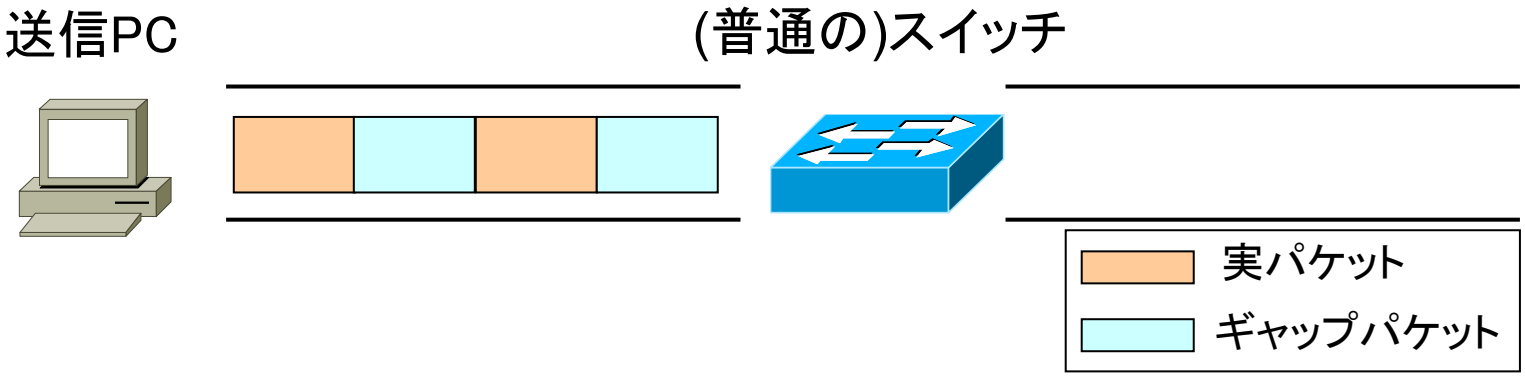
ギャップパケット(1)

- 実パケット間にギャップパケットというダミーのパケットを送信する
  - 送信レートを500Mbpsにするなら…



ギャップパケット(2)

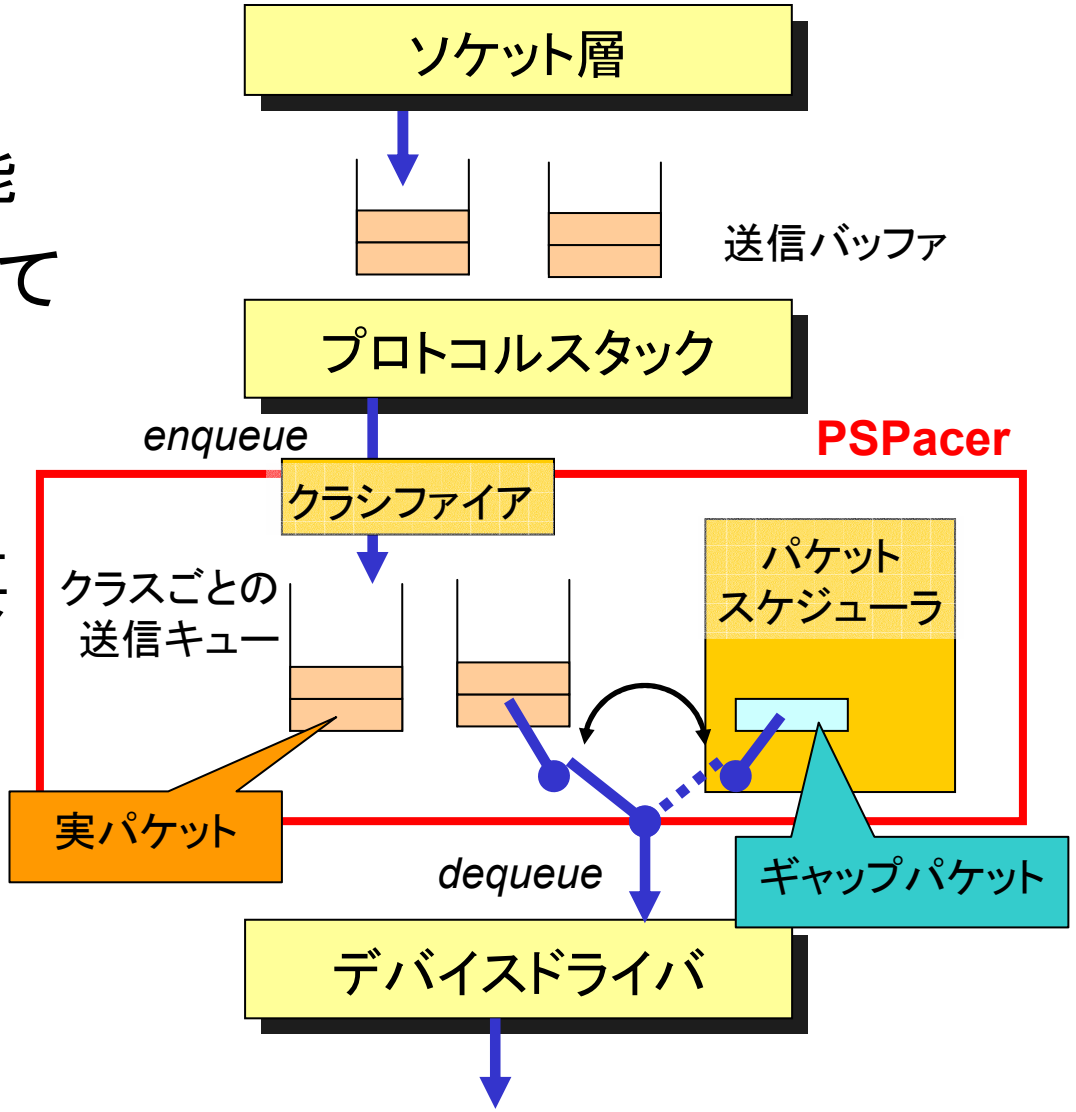
- PAUSEフレーム(IEEE 802.3x)を利用
  - スイッチ/ルータの入力ポートで破棄されるので，外部ネットワークへの影響はない
- ギャップパケットサイズを調整することで，送信レートを正確に制御可能
  - 1バイト　(8n秒)単位



PSPacerの実装

- iproute2 + tc
  - tcコマンドで設定可能
- Qdiscモジュールとして実装

- カーネル再構築不要
- デバドラ非依存
- プロトコル非依存



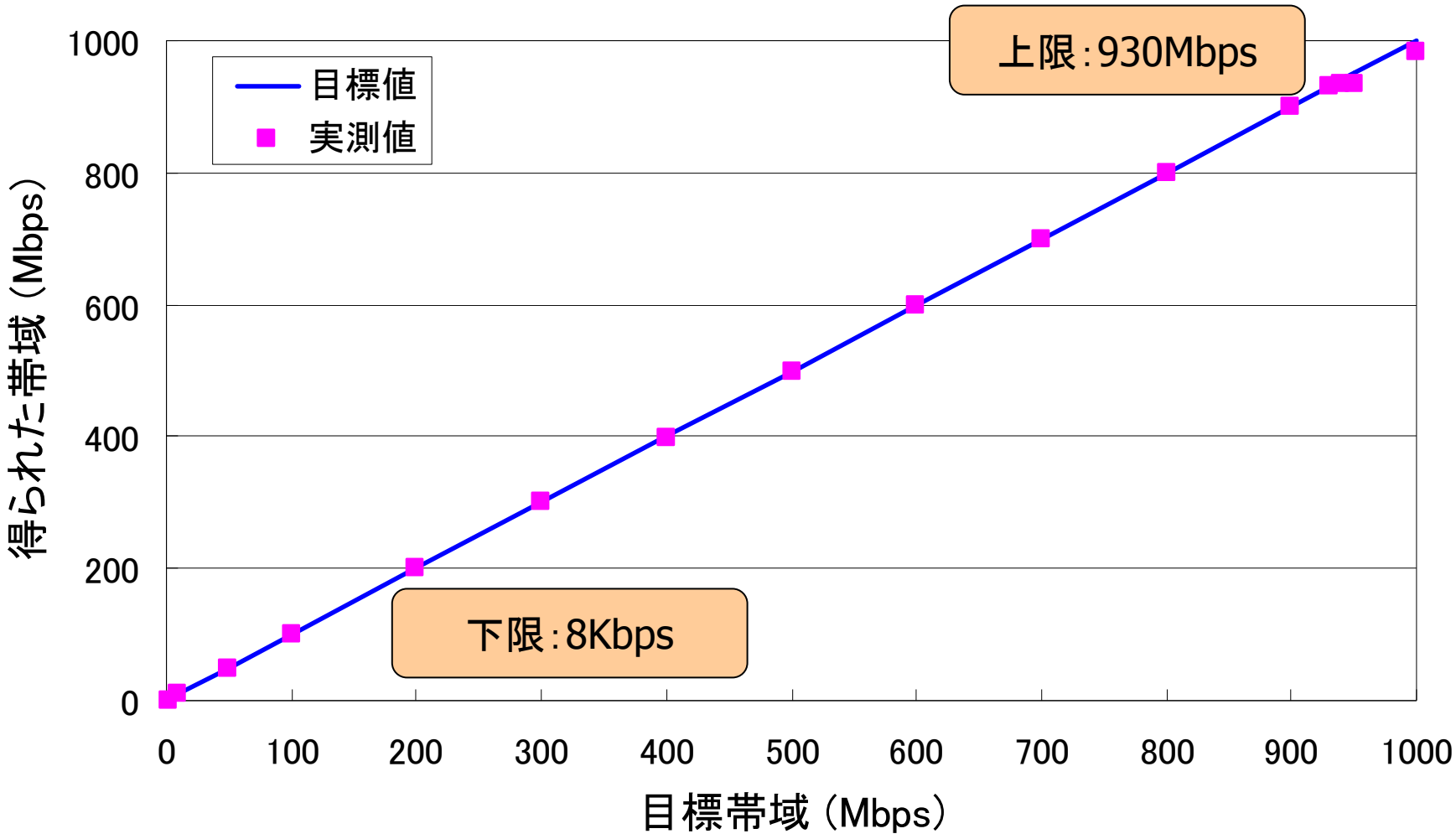
基本的な実験

- 目標帯域と実際の結果が一致するか？
- クラスごとに帯域制御できているか？

PC環境
- CPU: Intel Xeon/2.4GHz dual
- メモリ: 2GB (DDR400)
- NIC: Intel PRO/1000 (82545EM)
- OS: Fedora Core 3
  Linux 2.6.11.10
- BIC TCP

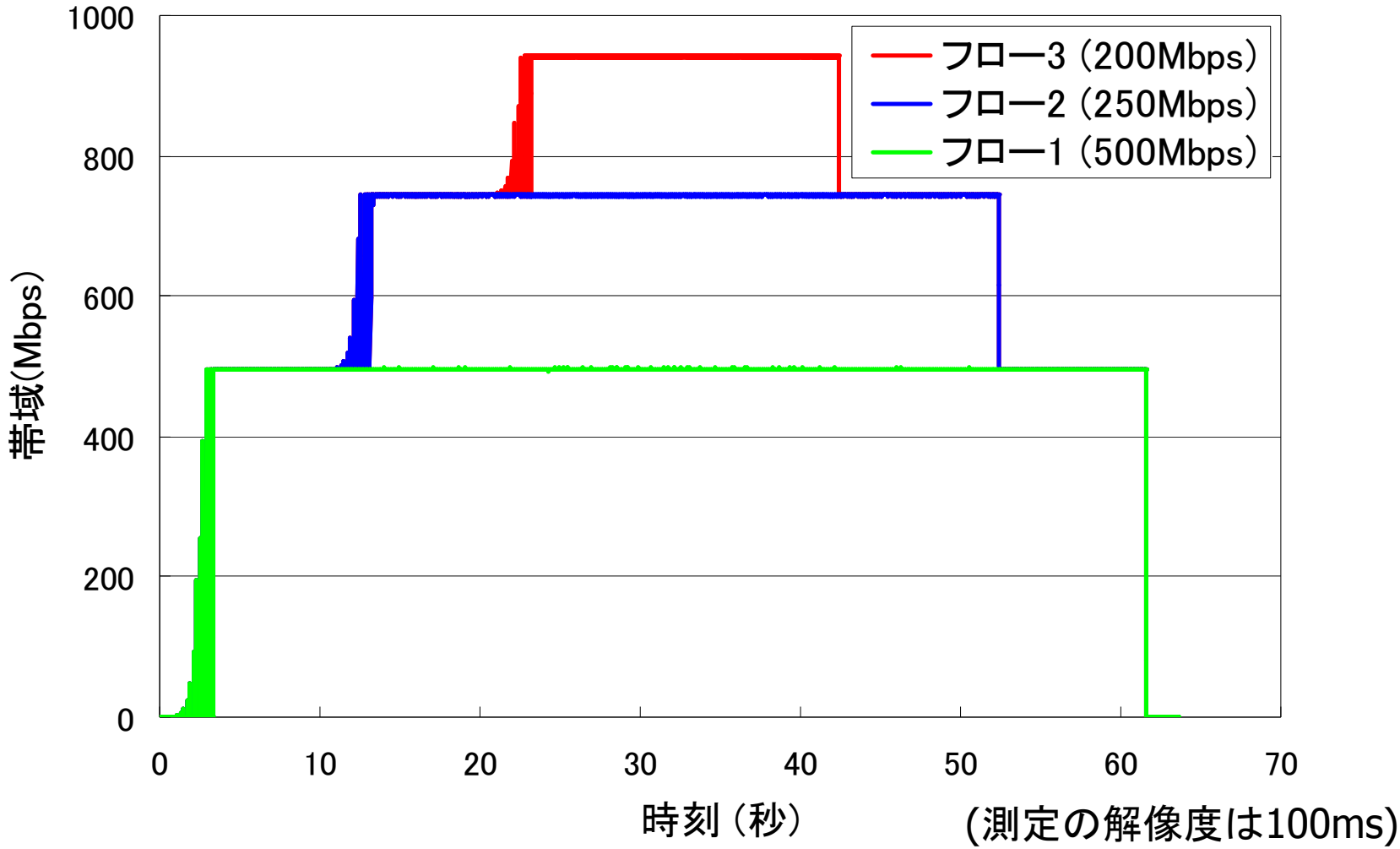
精密な帯域制御の実現

8Kbps～930Mbpsでの帯域制御に対応



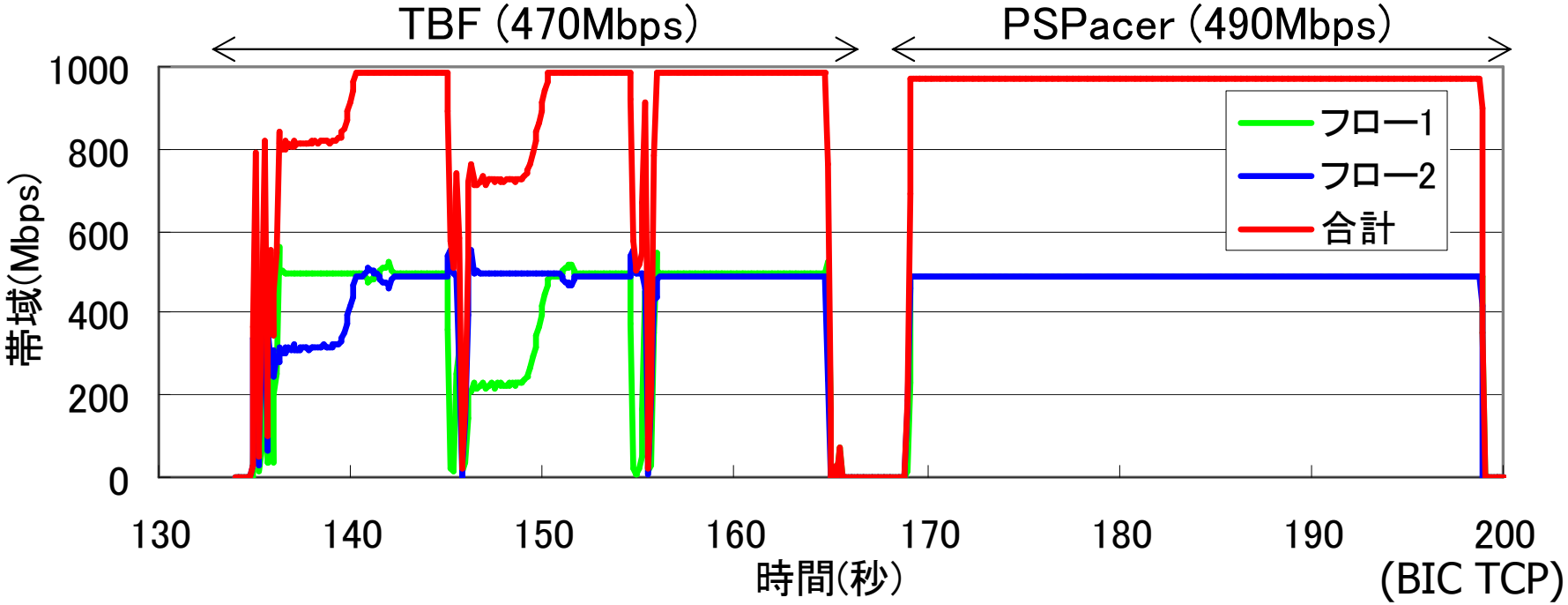
クラスごとの帯域制御

100クラス程度での動作は確認済み

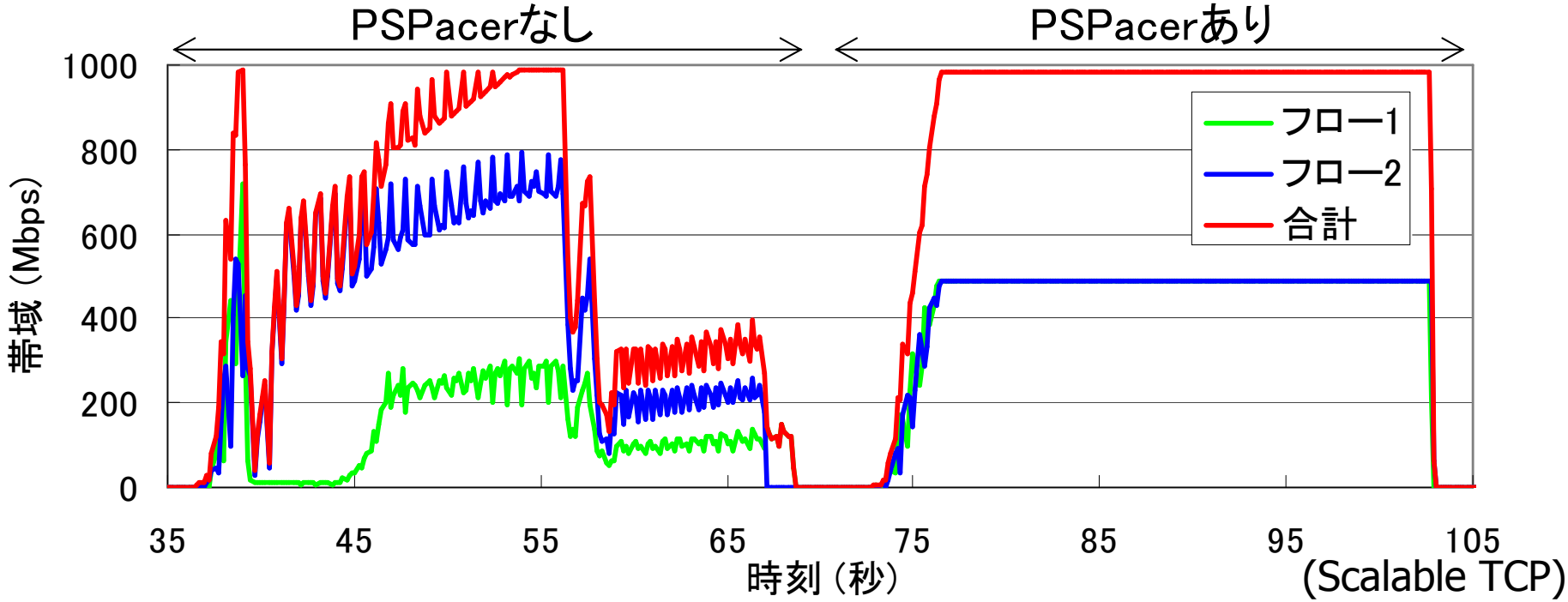


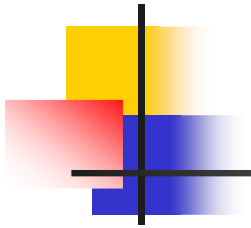
(測定の解像度は100ms)

高遅延環境でのTCP/IP性能(1)



高遅延環境でのTCP/IP性能(2)





組込みLinuxで使えるの？

- PSPacerの制限
  - バスボトルネックなどで送信性能が制限される場合は，正確な帯域制御ができない
    - 例：GbE + PCI (32bit/33MHz)
- CPU，メモリバスへの負荷
  - ギャップパケットは，プロトコル処理は不要だが，DMA転送を伴う
    - CPU負荷：比較的小さい，メモリ負荷：比較的大きい

まとめ

- 従来専用ハードウェアが必要だった精密なペーシング機構をソフトウェアだけで実現
- 組込み分野での適用例はない
- 興味を持たれた方は，ぜひ使ってください！

- GPLライセンスにて公開
  - http://www.gridmpi.org/

なお、本研究の一部は文部科学省「経済活性化のための重点技術開発プロジェ クト」の一環として実施している超高速コンピュータ網形成プロジェクト（NAREGI：National Research Grid Initiative）による。

AIST

NAREGI

GtrcNET-1: Programmable Gigabit Network Testbed