

ハードウェアTCP-NIC (TCPオフロードエンジン)の 性能評価とTCP動作解析, その考察

NEC ネットワーキング研究所

ネットワーク・アーキテクチャ

長谷川洋平

2003年5月23日

本日の発表内容

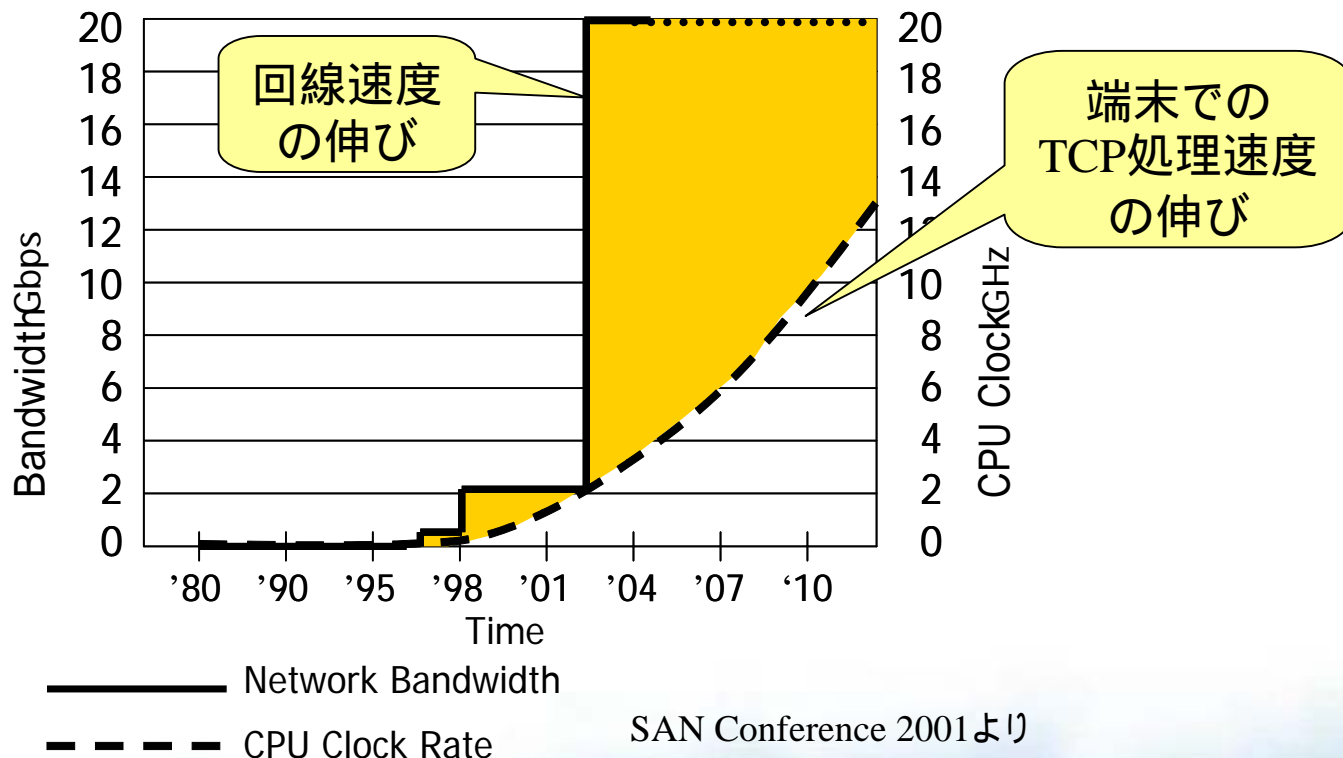
- 背景
- TCPオフロードエンジン (TOE) の評価
 - デバイスがボトルネックとなる状況での性能評価
 - 送信スループット
 - 受信スループット
 - 送信・受信同時スループット
 - ネットワークがボトルネックとなる状況での性能評価
 - 送信スループット
- TOEの特性と、適用領域の考察
- まとめ

背景 ~ 回線速度 と CPUクロック

1GHzのCPUは およそ
1Gbps のTCPを処理可能

10Gbps Ethernetの登場
(双方向で20Gbps)

TCP/IP処理がホストCPU
の負荷を圧迫

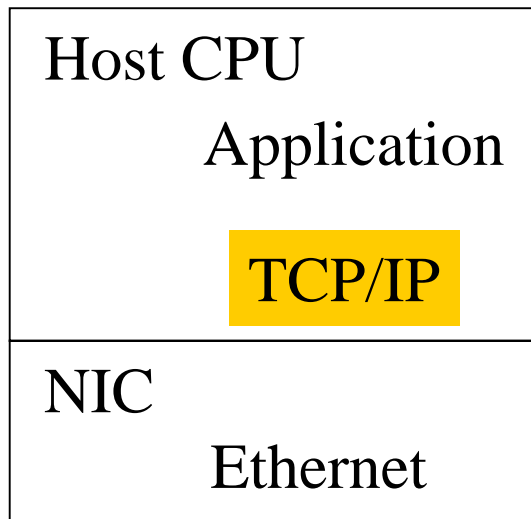


1Gbps以上の高速データ通信が必要な場合は
CPUからTCP/IP処理を切り離す

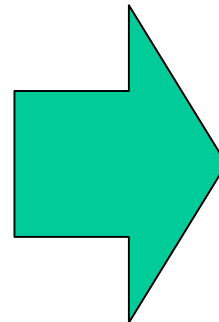
TCP処理エンジンが必要

背景 ~ TCP Offload Engine (TOE)とは？

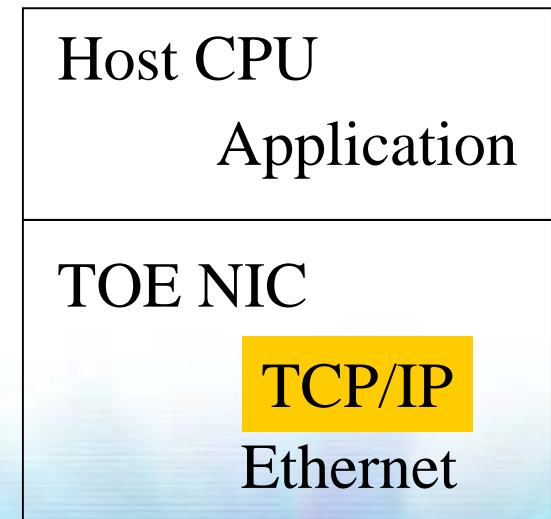
- TCP/IP処理 (+ 簡単な上位レイヤ処理)をNIC上に実現



従来のシステム



TCP処理のオフロード
CPU負荷の低減
TCP処理の高速化



TOEを組み込んだシステム

特性の複雑化 ~ TOEを正しく適用するためには

- TOE
 - 各社異なるアーキテクチャ, ハードウェア的な特性
 - 各社異なるTCP実装, ソフトウェア的な特性
 - = 特性の複雑化が予想される.

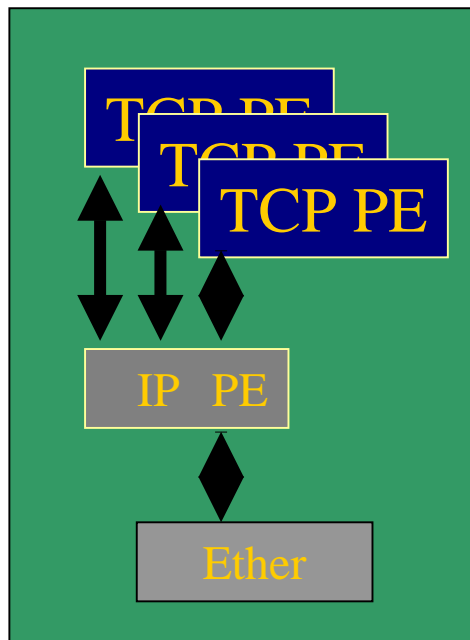


代表的な構造を持つ2つのTOEを評価し,
その性能, 特性の傾向と, 有望な適用領域を明らかにする.

評価対象としたTOEの概要

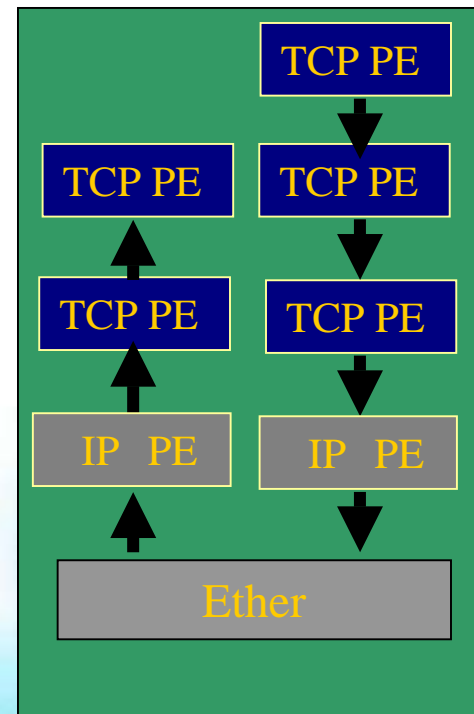
並列型

3つのPEによる
並列処理



パイプライン型

受信4つ, 送信6つのPEによる
パイプライン処理



評価環境 (使用マシン)

– TOE

- 並列型 OS: Windows NT4.0 Server
- パイプライン型 OS: Red Hat Linux 7.3

– TOE搭載マシン

- PentiumIII 933MHz 128MBメモリ

– 計測用マシン × 3

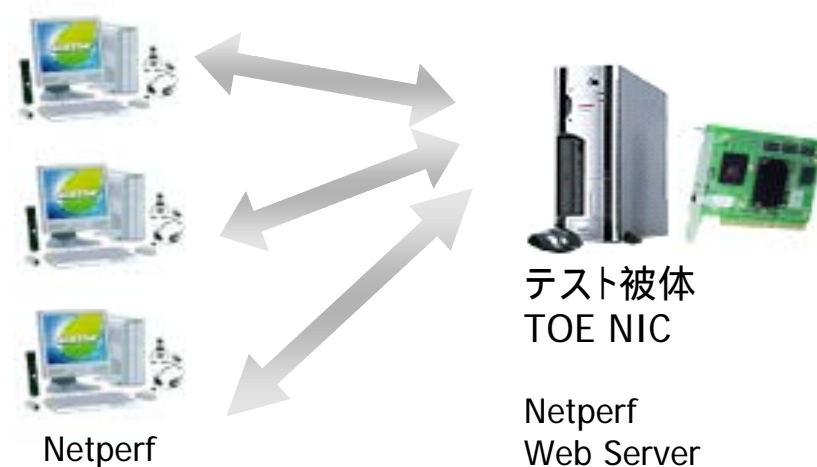
- PentiumIII 933MHz 128MBメモリ
- 1000Base/TX × 1
- FreeBSD4.3 , FreeBSD4.5

– スイッチ

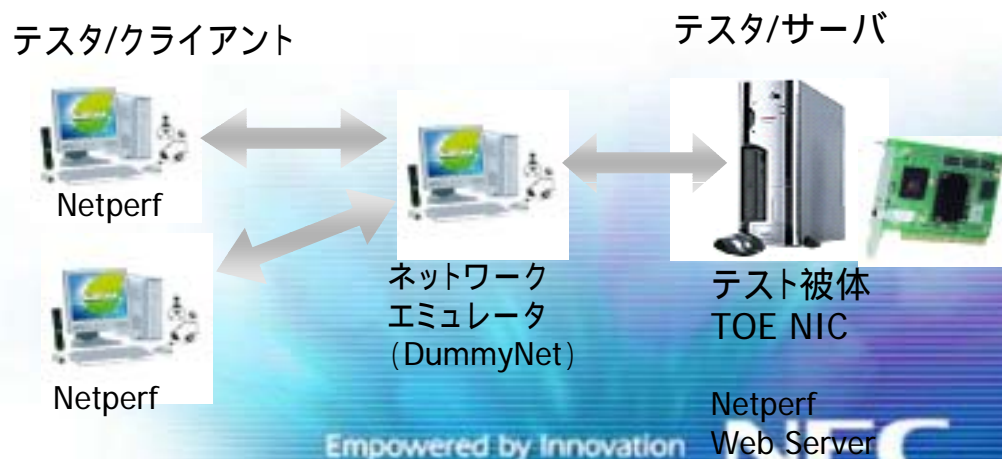
- 1000Base/TX × 4

TOE評価項目

- デバイスがボトルネックとなる状況での性能
(TOEの最大性能を観察)
 - スループット
 - コネクション開設レート
 - スループット + コネクション開設

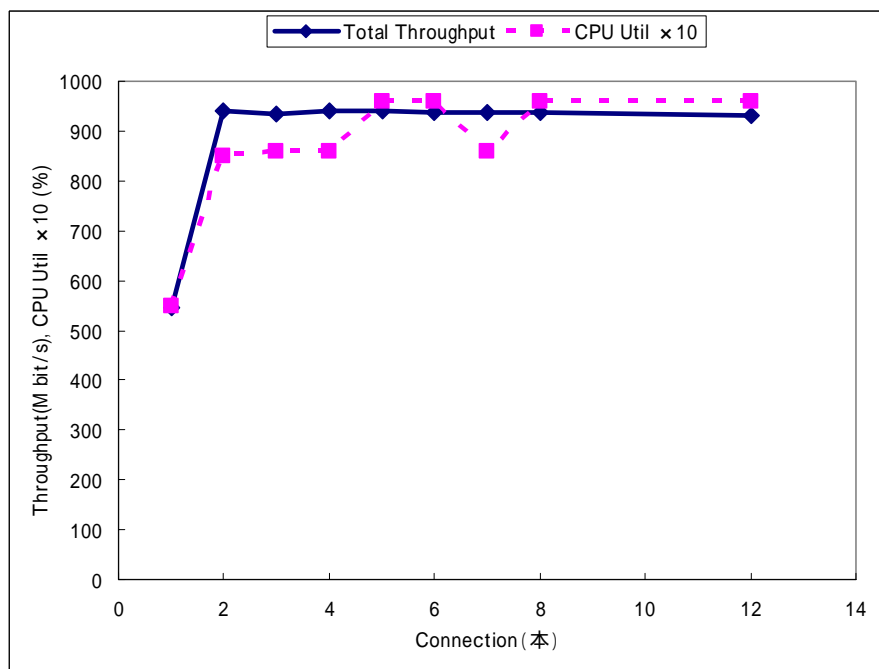


- ネットワークがボトルネックとなる状況でのスループット
(TOE上に実現されているTCPの性能を観察)
 - スループット

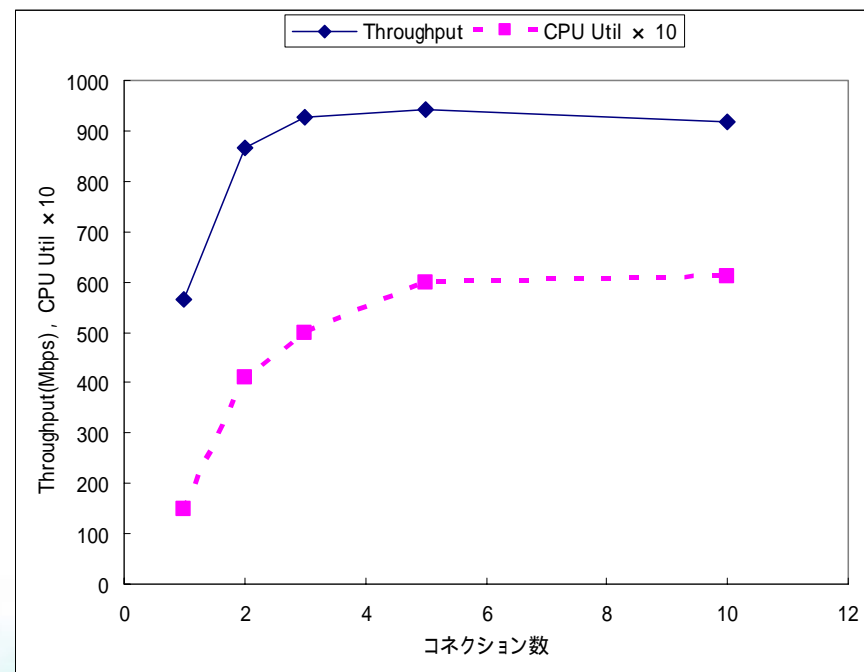


送信スループット

- 送信スループットでは, 2つのTOEともに, 良好な性能
 - コネクション1本の時の性能はFreeBSD側がボトルネック



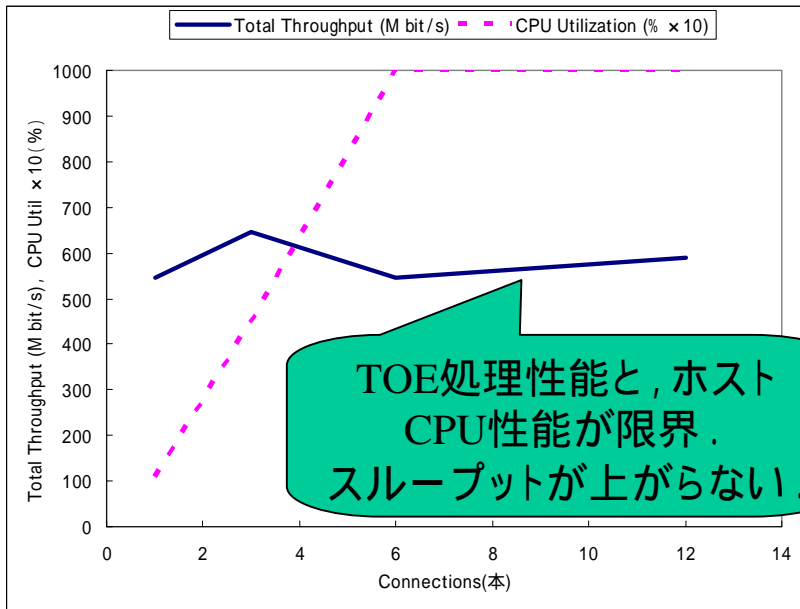
並列型TOEの送信スループット



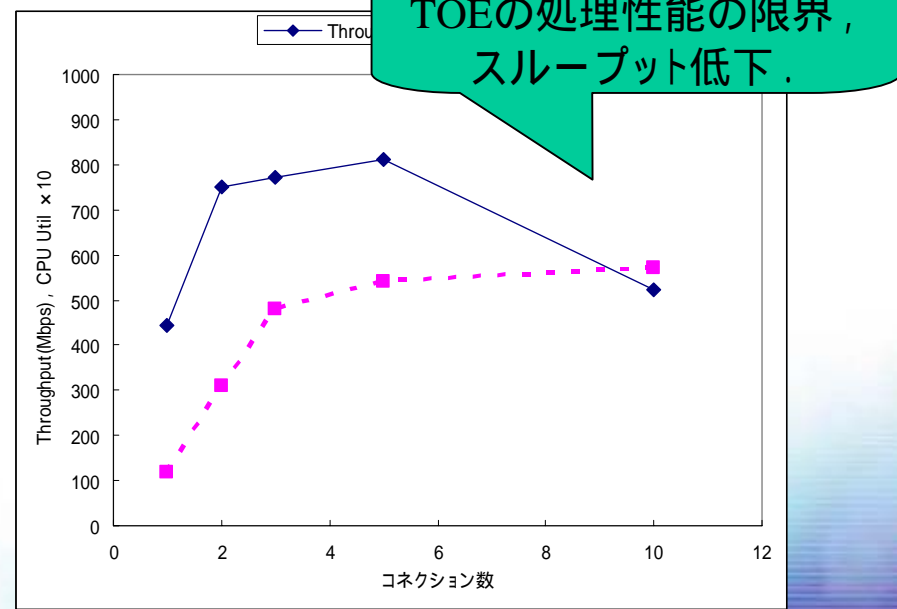
パイプライン型TOEの送信スループット

受信スループット

- 受信スループットは, 送信スループットに対して低い性能
 - 並列型TOE
 - コネクション数が増えると, スループットは変わらないが, CPU負荷だけ上昇.
 - パイプライン型TOE
 - コネクション数が増えるにつれて性能低下.
- 参考: 両方のTOEとも受信バッファサイズが小さい
 - 並列型 = 64KByte, パイプライン型 = 18KByte



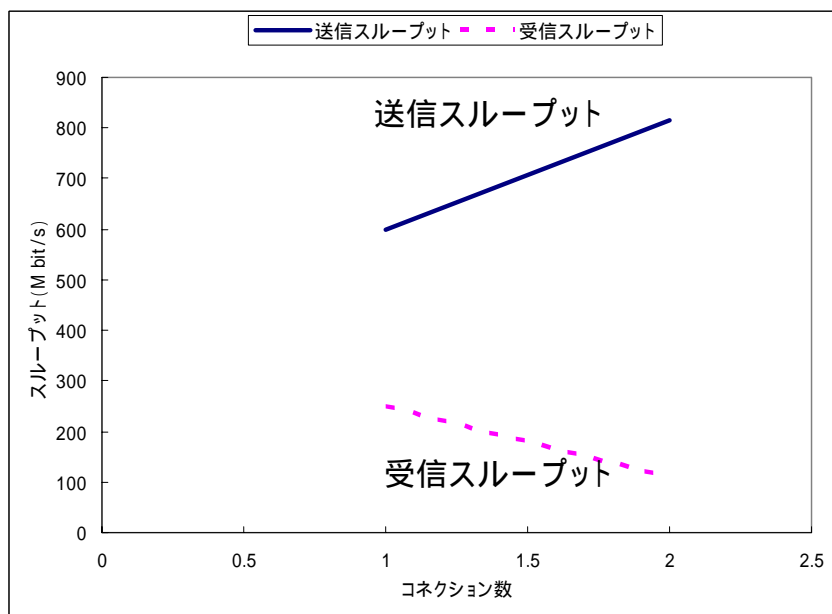
並列型TOEの受信スループット



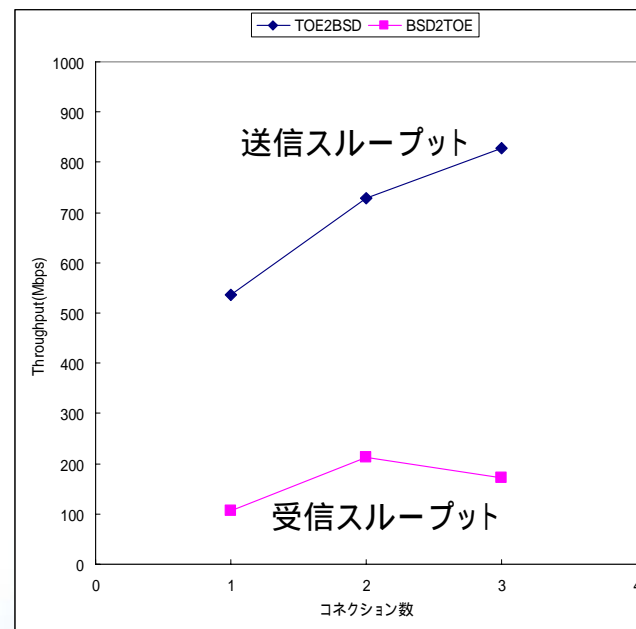
パイプライン型TOEの受信スループット

送信と受信の同時処理

- 送出力トラフィックが受信トラフィックを圧迫, 性能差は更に大きくなる.



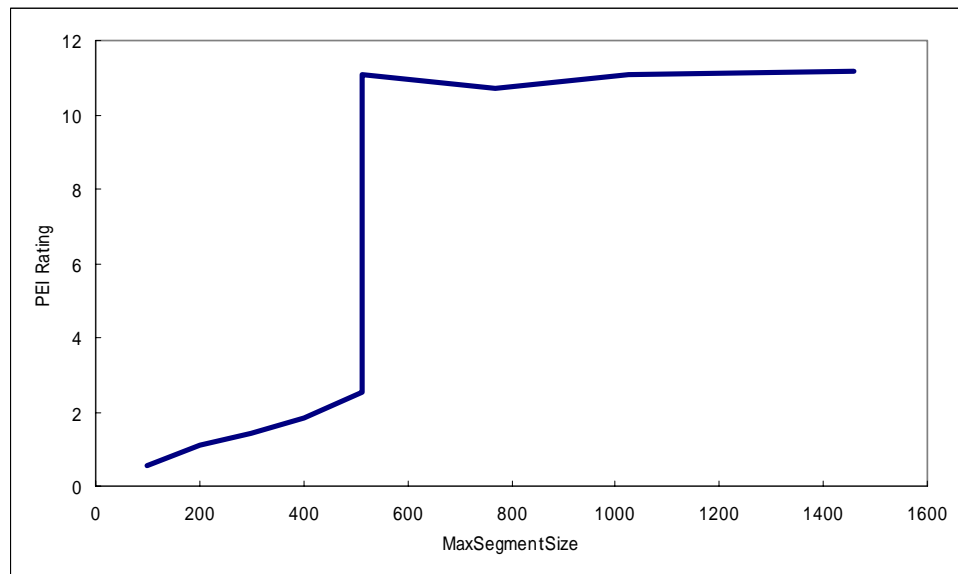
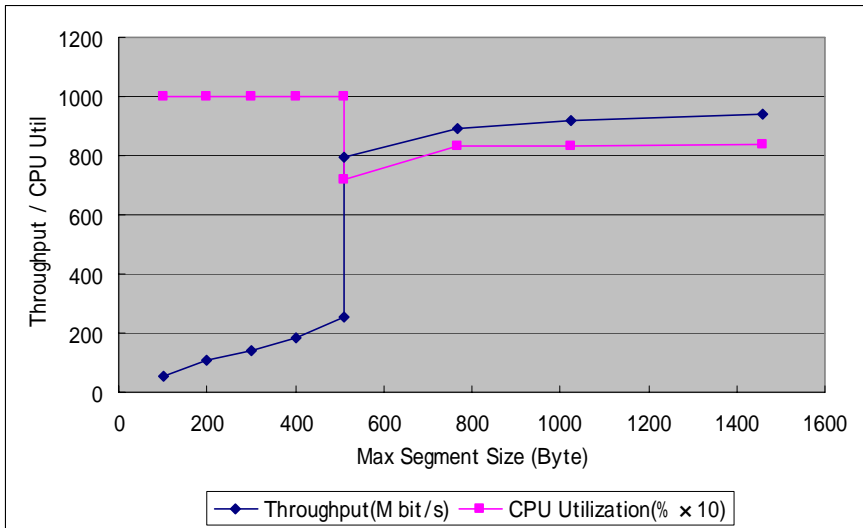
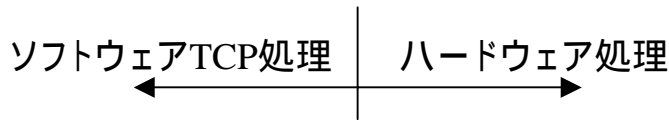
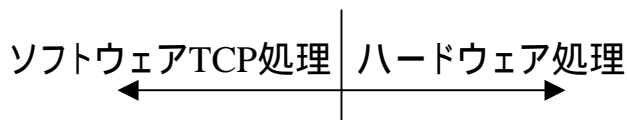
並列型TOE: 送信と受信の競合



パイプライン型TOE: 送信と受信の競合

パケットサイズ – スループット特性

- MSS < 512Byteではオフロードが無効となり, 急激にCPU負荷増大, スループット低下.
 - (この領域でオフロードが無効となる仕様となっている.)



パケットサイズ スループット特性

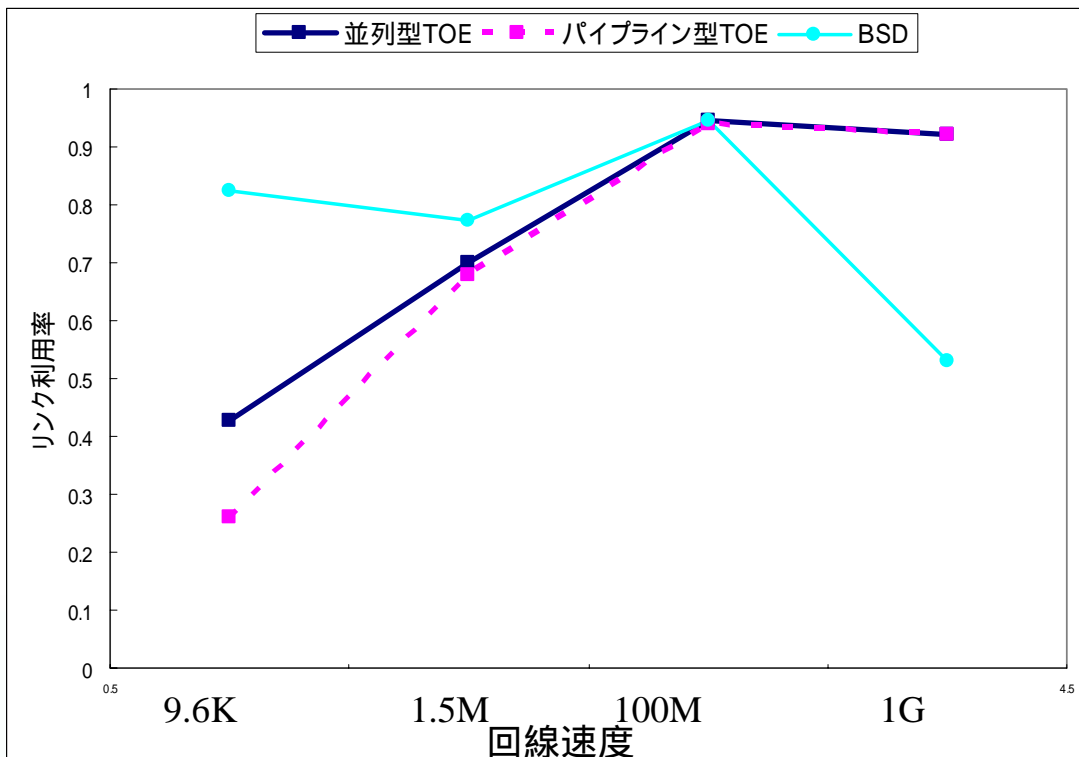
パケットサイズ – PEI特性
(CPU負荷あたりのスループット)

- リアルタイム性を重視するためにパケットサイズを小さくした場合には問題になる.
- ソフトウェアだけでは, 到底考えられない特性.

ボトルネックリンクがある状況でのスループット

- インターネットなど、回線がボトルネックになる状況ではスループットが發揮されない。

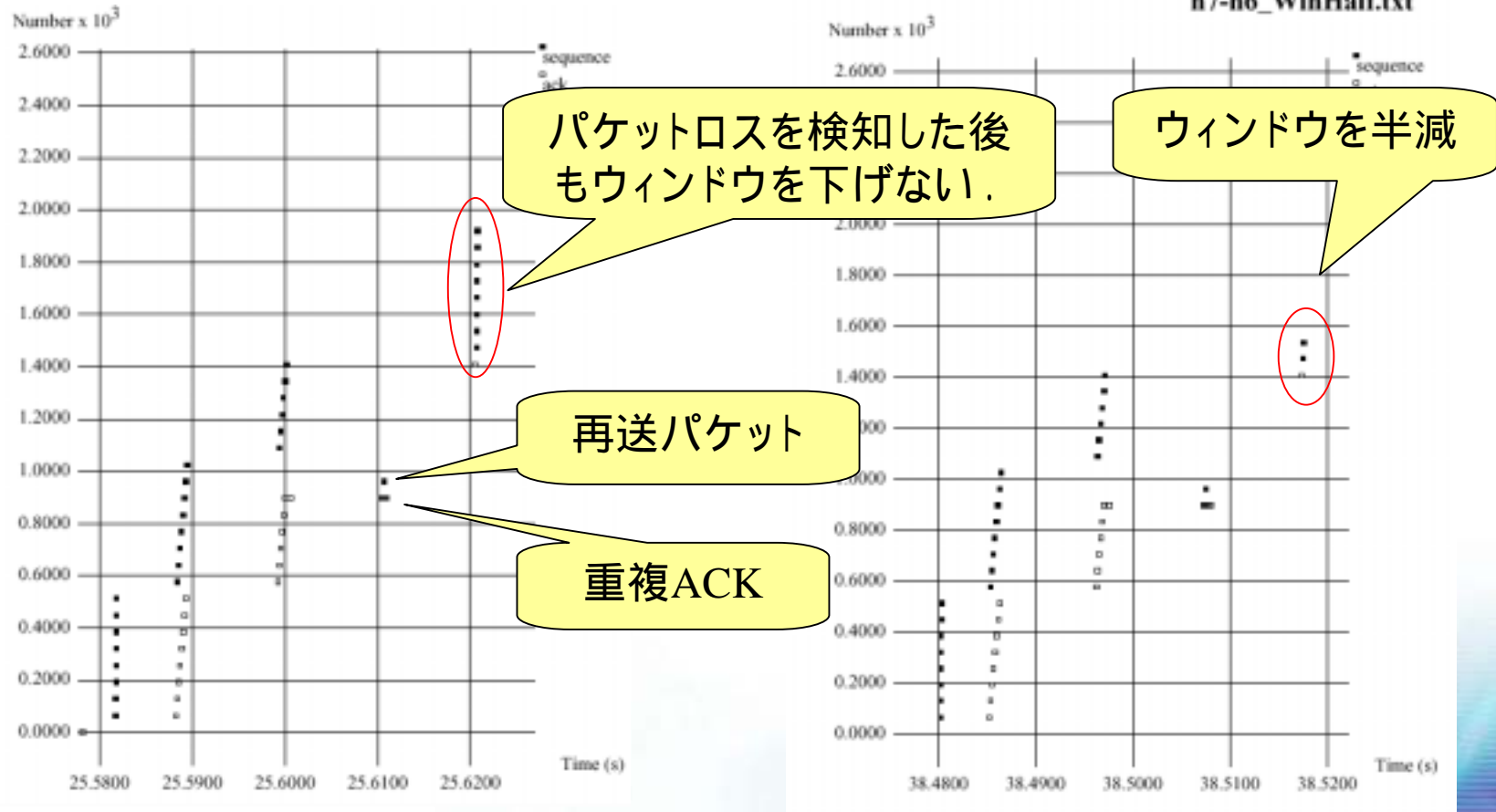
ネットワークエミュレータを使用し
回線速度を制限,
10msの遅延を付加



回線速度ごとのリンク利用率

並列型TOEのレート制御

並列型TOEはパケットロス後に輻輳ウィンドウを下げない場合がある。
重複ACKの検出と Fast Retransmit は正常に機能。



並列型TOEのレート制御

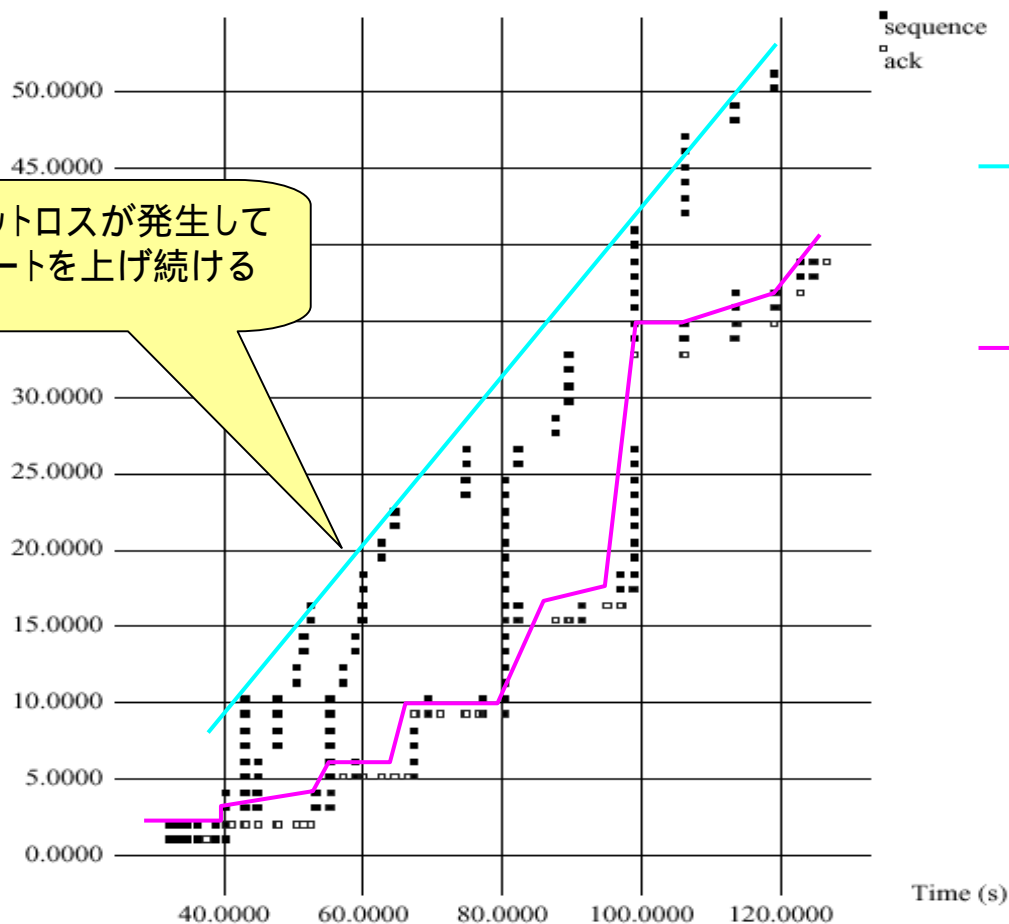
参考: FreeBSDのレート制御

Empowered by Innovation

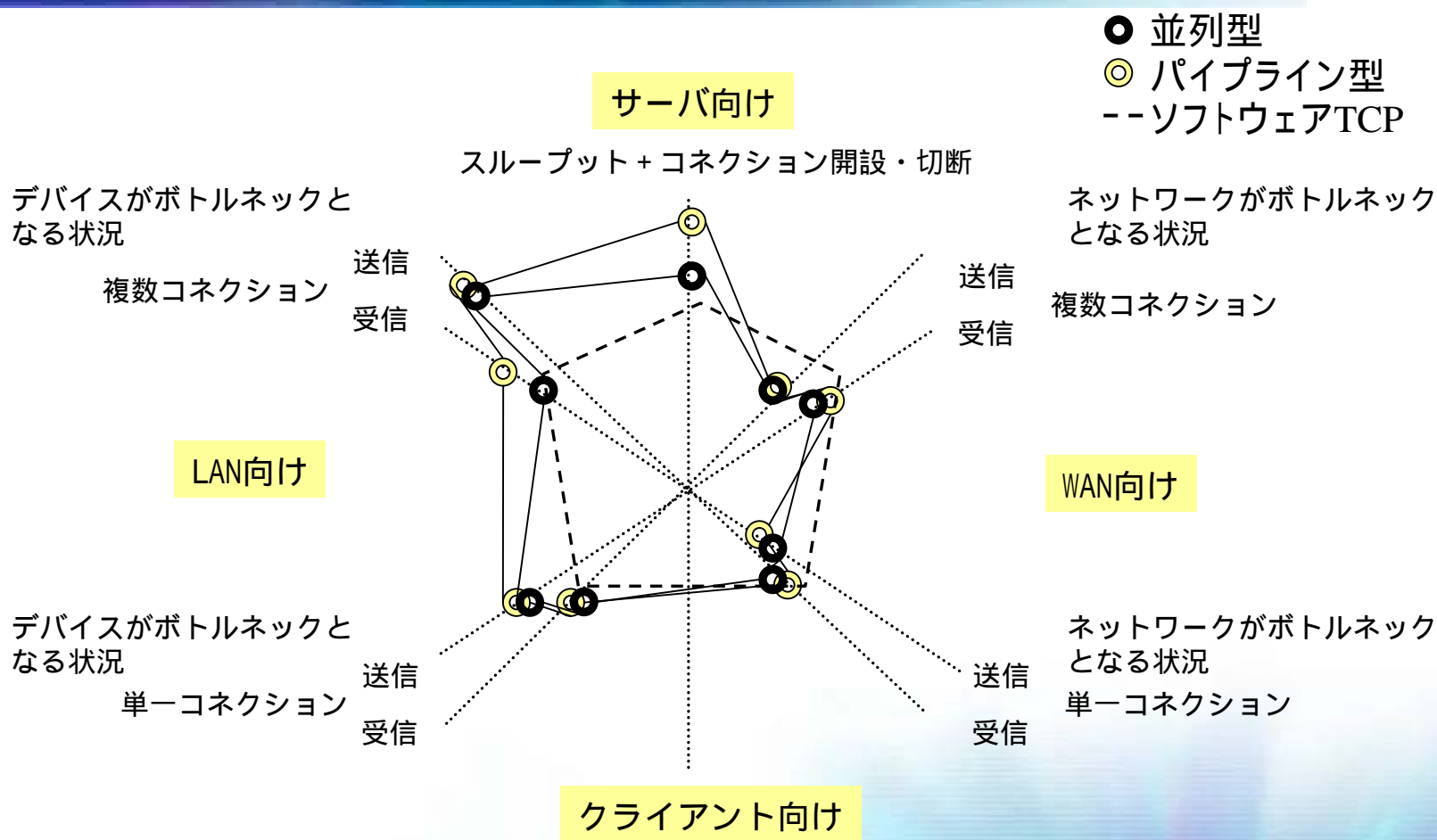
NEC

パイプライン型TOEのレート制御

パケットがロスしていても、レートを上げつつ再送を繰り返す。
多くのパケットを数回再送。 通信効率の悪化



TOEの特性グラフ



主に送信性能に優れサーバへの適用が考えられるが、レート制御能力に問題があるため、WANに直接つなぎ使用するにはリスクがある。

まとめ

- TOEには、ソフトウェアでは考えられない特性があり、適用には注意が必要。
- 評価した 並列型、パイプライン型のTOEは、どちらも
 - 送信性能は高いが、受信性能は相対的に低い
 - 遅延とパケットロスがあるネットワークでは、レート制御に問題があるため通信効率が悪化。
- これら特性のため、
 - LAN内で使用されるサーバなどでは、TOEの効果が期待できる。
 - WANに直接接続されるサーバ(例えばWebサーバ)として適用した場合は、性能が発揮されない危険性がある。特にクライアントからみた性能(コネクション1本あたりの性能)が低下する危険性がある。