

B-6-59

# イーサネットを用いたシステム仮想化技術 ExpressEther の提案 (3) パケット再送方式

## ExpressEther – Ethernet-Based Virtualization Technology for Reconfigurable Hardware Platform (3) Packet Retransmission Scheme

下西 英之 鈴木 順 飛鷹 洋一 樋口 淳一 柳町 成行 吉川 隆士 岩田 淳  
Hideyuki Shimonishi Jun Suzuki Yoichi Hidaka Junichi Higuchi Nariyuki Yanagimachi Takashi Yoshikawa Atsushi Iwata

NEC システムプラットホーム研究所  
System Platforms Research Laboratories, NEC Corporation

### 1. はじめに

本稿では ExpressEther [1]技術における、end-to-end パケット再送制御のシミュレーション評価結果を示す。

ExpressEther では CPU-I/O 間非常に広帯域(数 Gbps)で低レイテンシ(数 usec)の通信を提供する必要があり、レイヤ 2 でのハードウェアによるパケット再送制御が必須である。本稿では CPU-I/O 間通信で想定される環境条件において、レイヤ 2 での再送制御についてのシミュレーション結果を示し、低コストのハードウェアを用いた再送制御でも十分な性能が得られることを示す。

### 2. イーサネットを利用した I/O 接続の課題

CPU と I/O を接続する技術としては、例えば Advanced Switch [2]のような hop-by-hop で再送制御・輻輳制御を行う技術が存在する。それに対して ExpressEther では、end-to-end で再送制御および輻輳制御を行うことにより、イーサスイッチに対して特別な機能を要求せず、低コストに CPU-I/O 間の接続ネットワークを提供する。そのため、レイヤ 2 ハードウェアでの再送制御が必要であり、ハードウェアコストと性能のトレードオフが課題となる。

### 3. ExpressEther における再送制御方式

再送制御方式としては一般的に Go-back-N や Selective-Repeat などがあり、ハードウェアコストの観点からは受信側でパケット並べ替えのためのバッファが不要な前者の方式が優れる。一方、Selective-Repeat 方式には再送パケット数が少ないという利点があるが、ExpressEther では同一拠点内、あるいは同一シェルフ内に CPU と I/O が存在することを前提とするため、往復伝播遅延が非常に小さく、Go-back-N でも無駄に再送されるパケットが少ないと考えられる。そこで、本稿では以下のような再送制御方式における ExpressEther の性能を評価する。

- 再送制御：Go-back-N
- 高速再転送：2-Duplicate ACK (受信側から同じシーケンス番号の ACK パケットが 2 個連続すると、次のシーケンス番号のパケットが廃棄されたと判断)

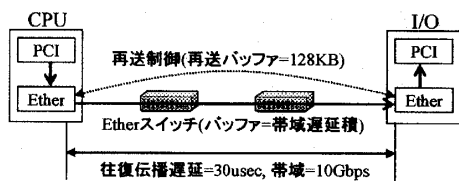


図 1: シミュレーションモデル

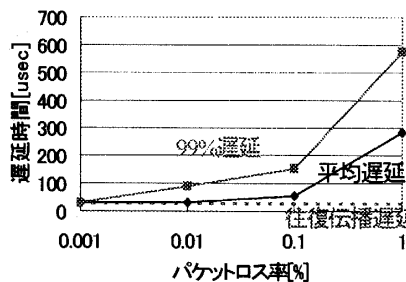


図 2: 1セッション時の end-to-end 遅延

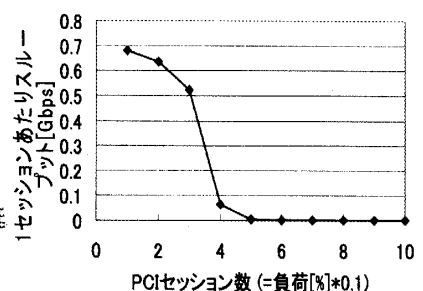


図 3: 複数セッション時のスループット

- 再送タイムアウト：一定時間パケット送信が滞るとパケット廃棄があったと判断

### 4. シミュレーション評価

図 1 のような条件でシミュレーション評価を行った。PCI-Express パケット長を 128 バイトとし、これにイーサのヘッダ等を付加して全体の長さを 170 バイトとした。

図 2 では、7Gbps の PCI セッション 1 本に対して、スイッチでパケットロスランダムに発生させた場合の、パケットレベルの end-to-end 遅延時間を示す。パケットロス率が  $10^{-5}$  以下では 99%遅延・平均遅延共に往復伝播遅延とほぼ等しくなり、通常想定されるパケットロス率の範囲では Go-back-N でも十分な性能が得られることがわかる。

次に、1MB のデータを平均 11.4msec(指数分布)間隔で発生する PCI セッション(平均レート 0.7Gbps)を複数本入力した場合の各セッションのスループットを図 3 に示す。この図より、複数の CPU-I/O ペアがイーサスイッチで多重される場合、高い利用率を実現するためには輻輳制御などのアクセス制御が必要であることがわかる。本トラフィック条件では、90%以上のスループットを出すためには利用率を 20%以下で運用する必要がある。

### 5. まとめ

ExpressEther による CPU-I/O 間通信において、低コストのハードウェアを用いた再送制御でも十分な性能が得られることを示した。また、複数 CPU-I/O 通信を同一スイッチで多重化する場合には、輻輳制御が必要である。

### 参考文献

- [1] J. Suzuki, et. al., "ExpressEther – Ethernet-Based Virtualization Technology for Reconfigurable Hardware Platform", to be presented at HOT Interconnects 14, 2006. [2] J. Gudmundson, "PCI express and advanced switching: evolutionary path to building next generation interconnects", High Performance Interconnects, 2003

謝辞:本研究の一部は、総務省の委託研究「次世代バックボーンに関する研究開発」プロジェクトの成果である。