ネットワークトラフィックの時間的局所性を利用したブロードバンドネットワーク向けキャッシュ型パケット処理技術

奥野通貴[†] 西村信治[†]
石田慎一[††] 西宏章[††]

急速なネットワークのブロードバンド化にともない，2007年以降頃のバックボーンルータは回線あたり現在比10倍の100 Gbps（Gigabit per second）のパケット処理スループットを大幅な消費電力の増加なしで実現することが求められる．筆者らは，ネットワークトラフィックで短時間に同一のヘッダを持つパケットが大量に出現することを利用するキャッシュ型パケット処理技術を提案した．本技術を用いるパケット処理エンジン（PPE）は，消費電力増加の原因となる内蔵プロセッサを増加させず，専用ハードウェア部にパケット処理履歴を記録し同一ヘッダを持つパケットに適用するProcess Learning Cache（PLC）と，PLCに未登録パケットの処理を行うCache-Miss Handler（CMH）を用いて高スループット化と低消費電力化を実現する．本稿では，キャッシュ型PPEを実現する際の技術課題と対応策を示し，FPGAを用いた規模縮小版の試作，実トレースを利用した評価を示す．そして，キャッシュ型PPEが100 Gbpsスループットを実現可能であることと，小規模プロセッサを多数搭載する従来型PPEに比べ面積，消費電力ともに46%程度に削減できる見込みであることを示す．

Cache-based Packet-processing Techniques for Broadband Network Exploiting Temporal Locality of Network Traffic

Michitaka Okuno,[†] Shinji Nishimura,[†] Shin-ichi Ishida[††] and Hiroaki Nishi[††]

The cache-based packet-processing enginge (PPE) that consists of Process-Learning Cache (PLC) and Cache-Miss Handler (CMH) is proposed to achieve 100-Gbps (Gigabit per second) wire-rate throughput which would be needed for backbone routers in 2007 or later to support broadband network. In this paper, several technical requirements and those solutions for cache-based PPE are shown. Also details of a prototype of the cache-based PPE are shown. By using real network traffic traces, we show the cache-based PPE can achieve the wire-rate throughput. As another aspect, PPEs are required low power consumption. In our estimation, the cache-based PPE can be constructed about 46% LSI-die size and power consumption of a conventional PPE.

1. はじめに

近年，数10 Mbps程度の安価なADSL（Asymmetric Digital Subscriber Line）やFTTH（Fiber To The Home）がエンドユーザに普及し，基幹インフラにおいても従来のSONETに加え，低コスト大容量通信可能なイーサネット[1]が普及しつつある．このようにネットワーク全体のブロードバンド化が進行中であり，1990年の10 Mbps以降4年ごとに10倍の高速化がなされてきたイーサネットにおいては，早ければ2007年頃には100 Gbps回線の登場が予測でき，基幹インフラのハイエンドルータでの利用が期待される．

ここで，現在のハイエンドルータは10 Gbps回線を16程度持ち，最大消費電力が4 KWクラスに達しているが[2),3)]，100 Gbps回線を搭載する次世代装置では，冷却や運用コスト面から消費電力を大幅に上昇させられない．本稿では，特にルータの主要構成要素であるパケット処理エンジン（PPE: Packet-Processing Engine）☆について，文献4)，5)で提案したキャッシュの手法を利用して消費電力の増加を抑え

† 日立製作所中央研究所
Central Research Laboratory, Hitachi, Ltd.
†† 慶應義塾大学理工学部
Faculty of Science and Technology, Keio University

☆ ネットワークプロセッサとも呼ばれる．

つつ 100 Gbps 回線処理を実現する方式の詳細について説明する．また，試作機による動作検証と実ネットワークトラフィックにおいて本方式が実用可能であることを示す．

2. 従来型のパケット処理エンジン

現在のハイエンドルータでは，10 Gbps から 40 Gbps 程度のパケット処理スループットを持つ PPE が利用されている．**図 1** に従来型 PPE を模式化した構成を，また，**表 1** に代表的なハイエンド PPE を示す[2),6),7)]．

PPE は，通常 64 byte から 1,518 byte のイーサフレームもしくはパケットと呼ぶ単位で転送されるデータを処理する．最小の 64 byte パケットを処理するのに与えられる時間は 10 Gbps イーサネットでは，わずか 67 ns である☆．パケットの解析，宛先検索，ヘッダ修正等のパケット処理をこの短時間で実現するために，多くの PPE は多数の小規模プロセッサ（PU）をチップ内に集積し，処理を細分化してパイプライン処理（X10q-w 等）や，並列処理（SPP 等），また両者の併用（NP-1c 等）によりスループットを確保している．

将来の 100 Gbps 処理に従来型 PPE で対応するには，内蔵 PU 数を増加させて並列処理数を増やすか，周波数向上が必要であるが，いずれも消費電力を比例して増加させるため，代替策が必要となる．

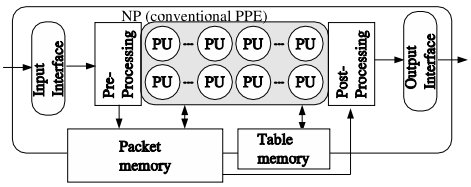


図 1 従来型パケット処理エンジン
Fig. 1 Conventional packet processing engine.

表 1 代表的なハイエンド PPE
Table 1 List of high-end packet-processing engines.

Product (Vendor)	スループット 周波数	プロセス (fab) 消費電力	PU 数
SPP (Cisco 内製)	40 Gbps 250 MHz	0.13 μm (IBM) 未公表	188
X10q-w (Xelerated)	40 Gbps 200 MHz	0.13 μm (TSMC) 9.5 W (typ)	200
NP-1c (EZchip)	20 Gbps 250 MHz	0.13 μm (IBM) 15 W (max)	64
IXP2800 (Intel)	10 Gbps 1,400 MHz	0.13 μm (Intel) 25.5 W (typ)	16 (128 スレッド)

☆ パケットの先頭に 8 byte のプリアンブル，パケット間に最低 12 byte のインターフレームギャップが入る．(64 + 8 + 12)/10 Gbps = 67.2 ns.

3. ネットワークトラフィックの特性とその利用

筆者らが注目しているのは，ネットワークトラフィックの特性を利用し内蔵プロセッサを増加させずに PPE のスループット向上を図る手法である．ネットワークを流れるデータは複数のパケットの連なりとしてネットワークを通過するため，文献 8)，9) でも示されるように，短時間に同一ヘッダ☆☆を持つパケットが多数出現しやすい．すなわち高い確率で時間的局所性が存在するといえる．筆者らも文献 4) においてネットワークのコアからエッジに近いところまで代表的な 4 サイトのトレースを調査した．宛先 IP アドレスと送信元 IP アドレスのペアが同じパケットを同一ヘッダを持つパケットと見なした場合，約 4,000 エントリのキャッシュメモリを利用すると，エッジ近くのサイトでは 98%以上の，またコアに近いサイトでも 80%近いヒット率が得られ，高い時間的局所性が存在していることをソフトウェアシミュレーションにより確認している．

時間的局所性の利用にはキャッシュが有効であり，キャッシュメモリを利用して，PPE のボトルネック処理である宛先検索処理において，宛先 IP アドレスに対応する次ホップ情報をキャッシュし，メモリ参照時間の平均値を小さくして高速化，ひいてはスループット向上を図るための研究がこれまでにも多数なされてきた．たとえば，ゲートウェイネットワークアドレスをフルアソシアティブキャッシュに蓄える手法[10)]，IP アドレスキャッシュ[11)]，通常のプロセッサのキャッシュを利用する手法[12),13)] 等があげられる．ここで，ネットワーク上では多数のパケットがつねに流れ続けているため，先行パケット処理がキャッシュミスを起こすと後続パケット処理がブロックされてしまい，最悪の場合処理が間に合わず廃棄されてしまう．先にあげたいずれの手法も，初回のメモリ参照と 2 回目以降のキャッシュメモリ参照の処理を適切に行いパケット廃棄を回避する手法が明示されていないため，高スループットを実現するために PPE にそのまま適用するのは困難である．そこで，筆者らが提案しているパケット廃棄が起こらないようキャッシュ処理を適切に行い PPE のスループット向上を図る手法について説明し，提案手法が実現可能であることを試作機により示す．

☆☆ たとえば，宛先 IP（Internet Protocol）アドレスと送信元 IP アドレスが同じヘッダを同一ヘッダと見なす．設定によってはさらに多くの情報が一致している場合を同一ヘッダと見なす．

4. キャッシュ型パケット処理エンジン

ここでは，キャッシュの概念を利用したPPEをキャッシュ型パケット処理エンジンと呼ぶ．

4.1 キャッシュ型 PPE 構成上の課題

キャッシュ型 PPE の課題は次の 3 点であり，それぞれについて対応策を示す．

(1) 高スループット化可能であり，キャッシュ論理追加が内蔵プロセッサ数増加より十分小さいこと
(2) キャッシュが十分高いヒット率を出せること
(3) キャッシュ処理によるパケット廃棄をなくすこと

4.2 高スループット化への対応策

高スループット部としてキャッシュメモリを備えた多ビット幅の専用パイプラインハードウェア，低スループット部としてプログラム性のあるプロセッサ群をキャッシュ型 PPE の基本構成とする．そして，低スループット部で実施した検索結果，およびヘッダ修正情報をデータおよび命令として，高スループット部のキャッシュメモリに記録し利用する．

ここで，キャッシュヒット率☆を h，高スループット部の最大スループットを $th(high)$，低スループット部の最大スループットを $th(low)$ とする．高スループット部がキャッシュヒットパケットを処理し，低スループット部が高スループット部の補助をするため，全体の最大スループット $th(max)$ は式 (1) で表現できる．

$$th(max) \leq th(high) * h + th(low) \qquad (1)$$

なお，すべてのパケットは高スループット部を通過するため，$th(max)$ の最大値は $th(high)$ に制限される．さらに，$th(max) = th(high)$ とするためには，低スループット部は式 (1) から，$th(high) * (1 - h)$ 以上の $th(low)$ が必要となる．

図 2 にキャッシュ型 PPE の構成例（以後，P-Gear と呼ぶ）を示す．P-Gear は高スループット部に BSP（Burst Stream Path）と呼ぶパイプラインハードウェア，低スループット部に P-Engine（Programmable Engine）と呼ぶプロセッサ群を利用する．P-Engine は従来型 PPE に似た構成とし，並列処理によりスループットを確保する．入力パケットはパケットメモリ（PMEM）に格納し，パケットのヘッダ情報からトークンと呼ぶ内部情報を生成して処理を行う．トークンの内容概略および，本試作での bit 幅を**表 2** に示す．BSP は，上流から順に，入力パケット解析用の A-Engine（Analysis Engine），キャッシュ処理用の C-Engine（Cache Engine），出力パケット処理用の R-Engine（Rebuilding Engine）と呼ぶ 3 つのブロックにより構成する．C-Engine は，P-Engine で実施した処理を BSP だけで行うためのデータを記録する Process Learning Cache（PLC）と，PLC にミスしたトークンの処理および P-Engine へのスケジューリングを行う Cache Miss Handler（CMH）により構成する．筆者らは文献 5) において 0.13 μm プロセスでの PPE の面積と消費電力に関する見積りを行い，従来型 PPE の PU 部は 113.3 W，一方，20 Gbps 分の PU を集積するキャッシュ型 PPE の消費電力は PU 部と BSP 部を合わせて 27.8 W との結果を得た．これより，キャッシュ型 PPE は従来型 PPE の 1/4 程度（27.8 W/113.3 W）の消費電力で実現できる見通しを得ている．

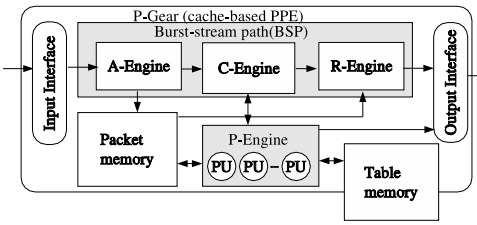


図 2 キャッシュ型パケット処理エンジン
Fig. 2 Cache-based packet processing engine.

表 2 トークンの構成
Table 2 Token specification.

名称	詳細
U-Info	23 bit （PMEM アドレス等）
A-Info	12 bit （レイヤ 2，3，4 ヘッダ解析情報）
E-Info	336 bit （パケットヘッダ抽出情報）
E'-Info	336 bit （新規パケットヘッダ情報）
P-Info	29 bit （出力ポート，パケット修正情報等）
U/A/E-Info	371 bit A-Engine token
U/A/E'/P-Info	400 bit R-Engine token

4.3 高キャッシュヒット率への対応策

十分な PLC ヒット率を確保するために PLC エントリ数は多いほど良いが，実装可能な LSI 面積とのトレードオフの解消を図る必要がある．文献 4) に示すように，4 K 程度のエントリ数を目安として得た．また，PLC の同一エントリだけが利用されることを防ぐため，トークンに CRC ハッシングをかけて PLC 上でのトークン格納位置を散らすことが有効である．さらに，PLC 登録の完了前に同一 PLC ミスが発生しうるため，CMH に PLC ミストークンをフロー☆☆ごとに管理する CMT（Cache Miss Table）を設け，同一 PLC ミスは CMT の各エントリに対応する CMQ

☆ (キャッシュに登録済のパケット数)/(受信パケット数)

☆☆ 同一のヘッダを持つ一連のパケット

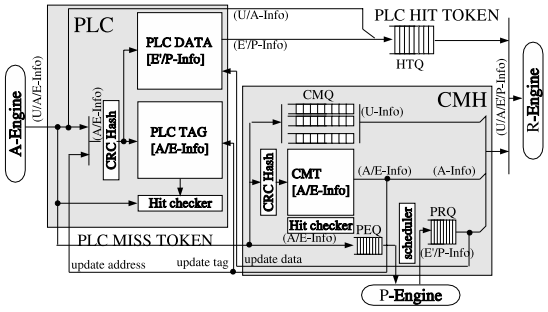


図 3 C-Engine の構成
Fig. 3 Structure of C-Engine.

(Cache Miss Queue) と呼ぶキュー群に保持し，つねにフローの先頭トークンのみ P-Engine にスケジューリングし，同一フロートークン処理による P-Engine 資源の浪費を回避する．

4.4 キャッシュ処理起因のパケット廃棄回避策

C-Engine の構成詳細を表 2 のトークンの流れを含め，**図 3** に示す．C-Engine 起因によるパケット廃棄を防ぐためには次の 4 つの要求を満たせばよい．

(1) PLC の帯域確保：PLC は A-Engine からのトークン参照要求と CMH からの更新要求を賄える帯域が必要である．よって，トークン参照要求は最短で 2 サイクルに 1 回とする．

(2) PLC ヒットトークンの保持：C-Engine は R-Engine に対し，PLC ヒットパスと CMH 発行パスを持つ．CMH が R-Engine にトークン発行中も PLC 参照をやめないために，PLC ヒットトークンを保持するキュー（HTQ）を付加する．

(3) 十分な CMH 資源の確保：CMH 資源（CMT エントリ数，CMQ の深さ）が不足すると，CMH が PLC ミストークンを受信できず，後続トークンの PLC 参照をやめてしまう．理想的には，CMT エントリ数は，PU 数を N とした場合，最大 $N * th(high)/th(low)$，CMQ 深さは最大 $N * th(high)/th(low)$ となる☆．

(4) P-Engine 資源の活用：式 (1) 中の $th(low)$ のスループットを最大限に引き出すため，P-Engine 内部の PU 群の並列処理が滞らないよう，P-Engine の入力部に待ち行列キュー（PEQ），出力部に結果保持キュー（PRQ）を付加する．

4.5 試作機概要

式 (1) が実現可能であることを実機検証するため，規模縮小版のキャッシュ型 PPE を試作した．試作機には，Xilinx 社の FPGA（Field Programmable Gate Array）である Virtex-II Pro☆☆を 4 チップ利用し，インタフェースは Gigabit Ethernet 2 本，最大スループットは 2 Gbps とした．まず，P-Gear を構成する主要ブロックに関して概略を説明する．

(1) A-Engine：入力パケットを解析し，マイクロコードによりパケットヘッダの必要部を抽出してトークンを生成する．また，パケットをメモリに格納する．設定により宛先 IP アドレスだけの抽出，送信元/宛先 IP アドレスの抽出，5tuple☆☆☆抽出等が可能である．

(2) C-Engine：PLC と CMH により構成し，A-Engine から受信したトークンを R-Engine でパケット処理をするための処理済みトークンに置換する．PLC は，P-Engine の処理結果 E'/P-Info をデータ，A/E-Info をタグとし記録する．A/E-Info を CRC ハッシングして参照する．PLC 既登録のフローは PLC 参照がヒットし，P-Engine をバイパスし R-Engine に処理済みトークンを送信する．CMH は，PLC ミストークンを処理する．異なるフローを管理する CMT は A/E-Info を，同一フローを管理する CMQ は U-Info を保持する．また，P-Engine 処理済みトークンで PLC を更新し，対応する CMT，CMQ に適用して R-Engine へ送信する．CMH の CMT はフルアソシアティブメモリが理想だが，FPGA の内部資源量を考慮し，8way のフルアソシアティブメモリと CRC ハッシングで参照する総計 1,024 エントリの 4way アソシアティブメモリを組み合わせた．

(3) R-Engine：C-Engine から受信した処理済みトークンを参照してメモリから元のパケットを読み出し，処理済みトークンの指示でヘッダの追加/置換/削除，アライン，部分修正を実施しパケットを外部へ送信する．

(4) P-Engine：本試作においては，P-Engine の BSP に対するスループットを可変にした評価を行うために，P-Engine の各 PU は C-Engine から受信したトークンを一定のルールに基づき変換し，ディップスイッチにより設定した待機時間ののち C-Engine に返信する擬似論理とした．

☆ なお，文献 4) の調査では，同一トークンだけが恒常的に連続する可能性は非常に少なく，CMQ 深さは 32 程度が目安となる．

☆☆ Type: xc2vp70, Package: FF1517, Speed Grade: −6, 33088 slices, 66176 FFs, 5904 K block rams

☆☆☆ 送信元/宛先 IP アドレス，プロトコル，送信元/宛先ポートの 5 種類の情報

表 3 P-Gear 試作パラメータ
Table 3 Spec of P-Gear testbed.

項目	詳細
インタフェース	1 Gbps Ethernet x2
最大スループット	2 Gbps
コア動作周波数	33 MHz
PLC 構成	1 K entry x 4 way set associative CRC hashing ($X^{10}+X^7+1$) LRU replacement 349 bit タグ，365 bit データ
PLC タグ容量	175 K byte
PLC データ容量	183 K byte
HTQ 容量	6.25 K byte（400 bit × 128 depth）
CMH's CMT 構成	256 entry x 4 way set associative with 8 entry x full associative CRC hashing ($X^8+X^6+X^3+X^2+1$)
CMH's CMT 容量	45 K byte（349 bit × 1,032 entry）
CMH's CMQ 容量	92.7 K byte（23 bit × 1,032 entry）
CMH's PEQ 容量	10.9 K byte（348 bit × 256 depth）
CMH's PRQ 容量	10.9 K byte（348 bit × 256 depth）
P-Engine	32 PUs（スループット調整可能）
PMEM 本体構成	512 K byte（256 Byte × 2,048 block）
PMEM 管理部構成	7.25 K byte（18 bit × 2,048 entry mem + 11 bit × 2,048 depth queue）

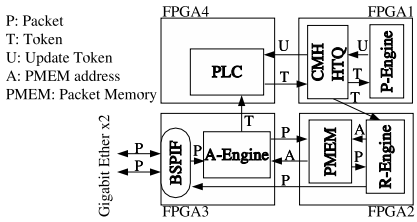


図 4 機能ブロック配置構成
Fig. 4 Functional block map.

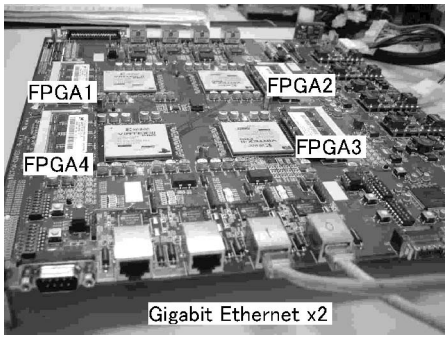


図 5 P-Gear 試作基板
Fig. 5 P-Gear testbed board.

表 3 に試作機の諸元，**図 4** に P-Gear の各機能ブロックの FPGA への割当て，**図 5** に試作機基板，**表 4** に FPGA の主要内部資源の利用率をそれぞれ示す．

表 4 FPGA 主要内部資源利用率
Table 4 FPGA main resource utilization.

機能	内部資源利用率		
	SLICE	LUT	RAM
BSPIF, A-Engine	65%	54%	18%
PLC	19%	10%	56%
CMH & P-Engine	42%	33%	47%
PMEM, R-Engine	6%	5%	72%

5. 評　　価

5.1 パケットジェネレータによる予備評価

5.1.1 予備評価条件

IXIA 社のパケットジェネレータ IxExplorer IxOS v3.65 で，**表 5** に示すキャッシュヒット率を作為的に調整した 6 種類の IPv4 イーサネットトレース☆1を試作機へ入力し，P-Engine 性能を変化させながらキャッシュヒット率とパケット通過率を測定し，従来型 PPE に対するキャッシュ型 PPE 性能に関し考察した．各トレースは PLC の全エントリ（4,096）を総入れ換えできるよう 16,000 種類のフローで構成した．パケット長は，PPE の処理負荷が最高となる最小イーサパケット長（64 byte）および，現実のインターネットトラフィックに見られるパケット長構成に似た比率のインターネットミックス（IMIX）☆2とし，約 10 秒間ずつ計測を行った☆3．

5.1.2 予備評価結果

最小パケット長での PLC ヒット率☆4を白抜きラベルで，また，PLC ヒット率および CMH ヒット率☆5を合わせた C-Engine 全体のキャッシュヒット率（以下，単にキャッシュヒット率と呼ぶ）を黒塗りラベルで**図 6** に示す．また，パケットの通過率を**図 7** に示す．横軸はいずれも搭載する P-Engine の BSP に対する相対スループットを示す．PLC off のラインは CMH 機能と PLC への登録機能を無効化した場合の trace A のパケット通過率で，P-Engine だけのスループット，すなわち従来型 PPE の場合のスループットに相当する．

☆1 たとえば trace A は，送受信 IP アドレスのホスト部またはネットワークアドレス部を一定の割合で変化させたパケット 50 個を 3 回繰り返し流し，次に IP アドレスが完全にランダムなパケット 50 個を流す．これを 80 セット分用意したトレースである．
☆2 パケット長の出現率は 64 byte : 570 byte : 1,518 byte = 7 : 4 : 1
☆3 64 byte 長で約 30 M パケット，IMIX で約 6.7 M パケット
☆4 (PLC にトークンが登録トークンがあった数)/(受信パケット数)
☆5 (PLC 未登録かつ，CMH の CMT に登録トークンがあった数)/(受信パケット数)

表 5 利用トレース
Table 5 traces for IxExplorer.

トレース名	内容	hit 率
trace A	80 pairs of (50 random, 50 fix ×19)	90%
trace B	80 pairs of (50 random, 50 fix ×9)	80%
trace C	80 pairs of (50 random, 50 fix ×6)	71.4%
trace D	80 pairs of (50 random, 50 fix ×4)	60%
trace E	80 pairs of (50 random, 50 fix ×3)	50%
random	random trace	0%

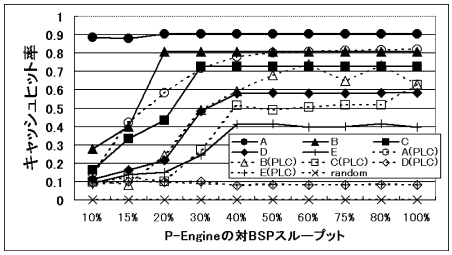


図 6 64 byte パケットキャッシュヒット率
Fig. 6 64 byte packet cache hit rate.

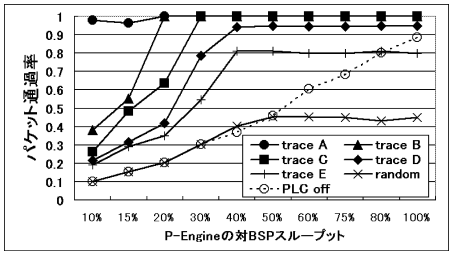


図 7 64 byte パケット通過率
Fig. 7 64 byte packet passing rate.

図 6 より，P-Engine の対 BSP スループットが低い領域では，PLC ヒット率がトレース本来のキャッシュヒット率より低く，CMH の存在によりキャッシュヒット率を向上できていることが確認できる．また，P-Engine の対 BSP スループットが 10%〜40%の低い領域では，trace B，C，D，E のキャッシュヒット率が著しく低下している．これは，P-Engine のスループットが不足して後続パケットの PLC 参照がブロックされた結果，パケットロスが発生したためと考えられる．図 7 の trace A，B，C より，キャッシュミス分を補う以上の対 BSP スループットを持つ P-Engine を搭載することで，式 (1) を満たす最大の $th(max)$，すなわちパケット通過率 1 を実現できることが確認できた．このとき，従来型 PPE より少ない P-Engine スループットで目標のパケット通過率を達成できることが分かる．なお，traceD，E，random では P-Engine の対 BSP スループットが十分高い領域でもパケットロスが発生している．これは，CMH の CMT エントリ競合により，CMH が PLC ミストークンを受信できなくなった結果，PLC 参照をブロックされてパケットロスが発生したためと考えられる．

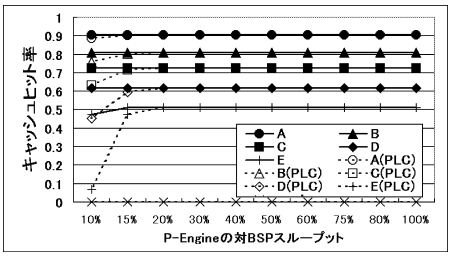


図 8 IMIX パケットキャッシュヒット率
Fig. 8 IMIX packet cache hit rate.

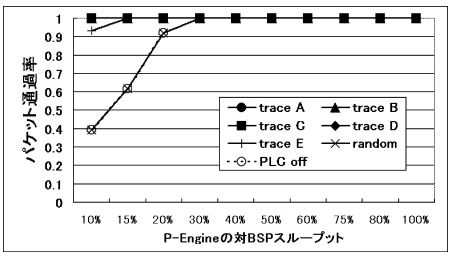


図 9 IMIX パケット通過率
Fig. 9 IMIX packet passing rate.

また，パケット長を IMIX に変えて同様の評価を行った結果を**図 8**，**図 9** に示す．図 8 より，IMIX では P-Engine の対 BSP スループットが小さい範囲でも PLC ヒット率は，トレース本来のキャッシュヒット率に近い価を示している．これは，平均パケット長が 354 byte の IMIX では，最小パケット長の場合に比べ，パケット受信間隔が約 4.4 倍長く☆，次のパケットを受信するときにはすでに PLC に当該トークン情報が記録されているためと考えられる．

また，図 9 より，CMH のトークン受信間隔も 4.4 倍長いため，図 7 で観測された P-Engine の対 BSP スループットが高い領域でのパケットロスも発生せず，50%以上のヒット率のトレースでは，P-Engine の対 BSP スループットが 15%以上の領域でパケット通過率 1 を達成できている．

5.2 実トレースによる評価

5.2.1 実トレース評価条件

ここでは，キャッシュ型 PPE が実ネットワーク上で利用できることを示すために，**表 6** に示す 3 サイトで採取されたパケットトレースを使ってキャッシュヒット率，パケット通過率を評価し考察した．WIDE,

☆ $(354 + 8 + 12)/(64 + 8 + 12) = 4.45$

表 6 トレース採取サイト一覧
Table 6 Site list.

KEY	SITE	bps
WIDE	WIDE trans-Pacific line B	100 M
A-IV	Univ. of Auckland uplink	155 M
IPLS	Abilene Indianapolis router	2.5 G

KEY	トレース名	ave. length
WIDE	20030227{1600,1615}	520 byte
A-IV	20010327-010000-1	1,019 byte
IPLS	IPLS-KSCY-20020814-100000-{0,1}	742 byte

WIDE: http://tracer.csl.sony.co.jp/mawi/
A-IV: http://wand.cs.waikato.ac.nz/wand/wits/auck/4/
IPLS: http://pma.nlanr.net/PMA/

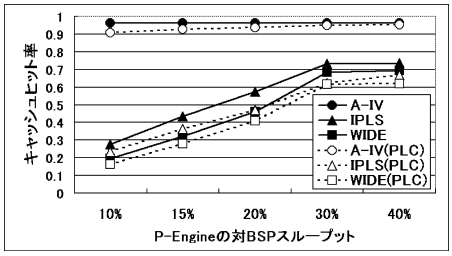


図 10 最小パケット長での実トレースキャッシュヒット率
Fig. 10 minimum length packet cache hit rate (real trace).

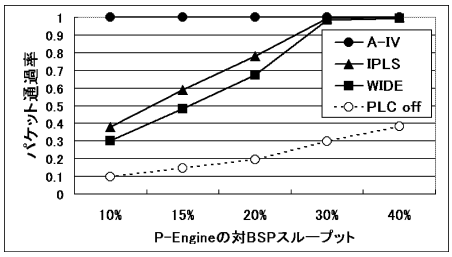


図 11 最小パケット長での実トレースパケット通過率
Fig. 11 minimum length packet passing rate (real trace).

IPLS はコア網近く，A-IV はアクセス網もしくはミドルマイル網と考えられる．表 6 に回線速度と採取トレースの平均パケット長を示す．今回，試作機に対しワイヤレートで採取トレースを入力する手段がなかったため，試作機の Verilog-HDL コードを NC-Verilog シミュレータ上で動作させ，各 200,000 パケットを最小インターフレームギャップ（12 byte）間隔で与え評価した．パケット長は，実パケット長と処理負荷を高めるために最小長両方を利用した．

5.2.2 実トレース評価結果と消費電力を含めた考察

図 10 に最小パケット長での PLC ヒット率を白抜きラベルで，C-Engine 全体のキャッシュヒット率を黒塗りラベルで，またパケット通過率を**図 11** に示す．実パケット長での同様の評価結果を**図 12**，**図 13** に

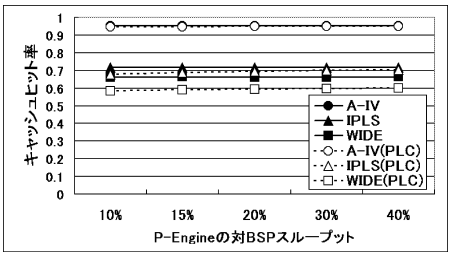


図 12 実パケット長での実トレースキャッシュヒット率
Fig. 12 real length packet cache hit rate (real trace).

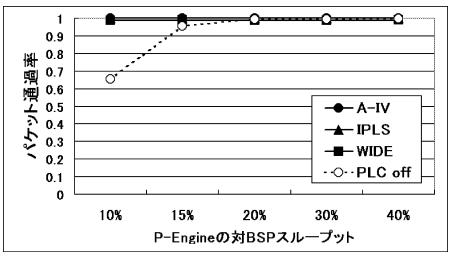


図 13 実パケット長での実トレースパケット通過率
Fig. 13 real length packet passing rate (real trace).

示す．PLC off の評価には平均パケット長が 3 サイトの中で最短の WIDE トレースを用いた．

図 10，12 から，A-IV，IPLS，WIDE のキャッシュヒット率はそれぞれ 95.2%，71.5%，66.0%であり，コア網に近い WIDE でもトラフィックに時間的局所性があることが分かる．図 11 から，最も処理負荷の高い最小パケット長の場合でもキャッシュミス率分を補う，対 BSP スループットがそれぞれ 10%，30%，40%以上の P-Engine の利用によりパケット通過率 1 を達成できることを確認した．なお，従来型 PPE の場合は，予備評価同様，対回線速度比が 100%の P-Engine が必要である．また，実パケット長を利用した図 13 からは，P-Engine は，10%程度の最小パケット長ケース時と比較して少ない対 BSP スループットで十分であるが，十分なヘッドルームを残す意味では，40%の対 BSP スループットを確保したほうが良いと考える．つまり，100 Gbps の PPE を構成するには 40 Gbps 相当の P-Engine を集積する．

筆者らは文献 5) において，0.13 μm プロセスにおける処理スループット 1，2.5，10，20，40，80，100 Gbps のキャッシュ型 PPE（ただし 100 Gbps 時の集積 PU は 20 Gbps 分）と従来型 PPE の面積と消費電力の見積りを行った．そして，100 Gbps 用 BSP は 31.0 mm^2，5.18 W，PU は 20 Gbps あたり 110.15 mm^2，22.66 W との価を得た．これより，40 Gbps の PU を集積する 100 Gbps キャッシュ型

PPE は 250.52 mm^2，51.7 W 程度であり，100 Gbps の従来型 PPE（550 mm^2，113.3 W）の 46%程度の面積，消費電力で効率良く実現できる見通しである☆.

6. おわりに

ネットワークトラフィックに存在する時間的局所性を活用し，低スループットのプロセッサ部と高スループットのキャッシュメモリ搭載ハードウェアを組み合わせたキャッシュ型パケット処理エンジン（PPE）の技術課題と対応策を示し，試作，評価を行った．特に実トレースを用いた評価により，100 Gbps スループットのキャッシュ型 PPE を構成する場合，現在の技術で実現可能な 10 Gbps〜40 Gbps 程度のプロセッサ部を集積することで 100 Gbps スループットを達成可能であり，40 Gbps のプロセッサ部を集積する場合には従来型 PPE に比べて面積，消費電力ともに 46%程度以下に削減できる可能性を示した．

キャッシュ型 PPE は，ブロードバンド化するネットワークを支えるために高スループット，低消費電力を要求されるバックボーンルータでの利用が期待できる．

謝辞 本研究の一部は（独）情報通信研究機構の委託研究「テラビット級スーパーネットワークの研究開発」の一環として実施された．

参考文献

1) IEEE802.3_task_force. http://grouper.ieee.org/groups/802/3/10G_study/public
2) Cisco. http://www.cisco.com
3) Hitachi. http://www.hitachi.co.jp
4) Okuno, M. and Nishi, H.: Network Processor Accelerator Using Temporal Locality of Traffic (in Japanese), *IPSJ Trans. Advanced Computing System*, Vol.45, No.SIG 6, pp.45–53 (2004).
5) Okuno, M., Ishida, S. and Nishi, H.: Low-Power Network-Packet-Processing Architecture Using Process-Learning Cache for High-End Backbone Router, *IEICE Trans. ELECTRON.*, Vol.E88-C, pp.536–543 (2005).
6) Xelerated. http://www.xelerated.com
7) Gwennap, L. and Wheeler, B.: *A Guide to NETWORK PROCESSORS Fifth Edition*, The Linley Group (2003).
8) Jain, R.: Characteristics of Destination Address Locality in Computer Networks: A Comparison of Caching Schemes, *Computer Networks and ISDN Systems*, Vol.18, pp.243–254 (1990).
9) Chvets, I. and MacGregor, M.: Multi-zone Caches for Accelerating IP Routing Table Lookups, *High Performance Switching and Routing* (*HPSR 2002*), pp.121–126 (2002).
10) Feldmeier, D.: Improving Gateway Performance with a Routing Table Cache, *INFOCOM'88*, pp.298–307 (1988).
11) Talbot, B., Sherwood, T. and Lin, B.: IP Caching for Terabit Speed Routers, *Global Communications Conference* (*Globecom'99*), Vol.2, pp.1565–1569 (1999).
12) Chiueh, T.C. and Pradhan, P.: High-Performance IP Routing Table Lookup Using CPU Caching, *INFOCOM'99*, pp.1421–1428 (1999).
13) Chiueh, T.C. and Pradhan, P.: Cache Memory Design for Network Processors, *IEEE High Performance Computer Architecture Conference* (*HPCA 2000*) (2000).

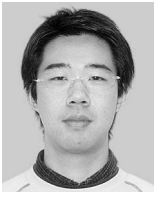
(平成 17 年 5 月 9 日受付)
(平成 17 年 11 月 1 日採録)

奥野 通貴

1998 年慶應義塾大学大学院理工学研究科修士課程修了．（株）日立製作所中央研究所勤務．スイッチ・ルータアーキテクチャ，ハードウェアの研究開発に従事．

西村 信治

1991 年東京大学大学院工学研究科修士課程修了．（株）日立製作所中央研究所勤務．ネットワークシステムの研究に従事，主任研究員，工学博士．



石田 慎一

2005 年慶應義塾大学理工学部卒業．現在，同大学院修士課程在席．インターネットアーキテクチャ，コンピュータネットワークの研究に従事．

☆ 利用する ASIC 種やプロセスルールにより面積，消費電力は変化するが，キャッシュ型 PPE と従来型 PPE の相対的な価は大きく変わらないと考える．なお，P-Engine に利用するプロセッサ群のうち未使用プロセッサへの電力供給を止めるなどしてさらなる消費電力削減も可能である．

西　宏章（正会員）

1999年慶應義塾大学大学院理工学研究科後期博士課程修了．同年技術研究組合新情報処理開発機構，2002年（株）日立製作所中央研究所，2003年慶應義塾大学理工学部システムデザイン工学科助手，工学博士．