

研究領域「実用化を目指した組込みシステム用
ディペンダブル・オペレーティングシステム」
平成18年度採択研究代表者

H22 年度 実績報告

佐藤 三久

筑波大学 システム情報工学研究科 教授
計算科学研究センター センター長

「省電力でディペンダブルな組込み並列システム向け計算プラットフォーム」

§1. 研究実施の概要

ユビキタス情報社会における高度な情報処理の要請に対し、これからの高性能組込みシステムはマルチコア・マルチチップになることが想定される。本研究では、ディペンダブルOSの一部として、並列システムの高信頼化機構および、省電力高性能高信頼通信機構を研究開発する。

当該年度においては、組み込み向け省電力・高信頼・高性能通信コミュニケーションハブであるPEACHチップのプロトタイプを開発した。これを用いて、既存のPCサーバ上で利用可能とするためのPEACHテストベッドボードを開発し、汎用PCサーバ上にPCI-Expressカードとして装着し、PEACHプロトタイプチップの基本機能の検証と、各種通信レート及びPCI-Expressレーン数に応じた通信性能と消費電力制御のトレードオフ制御が可能であることを確認した。これにより、DEOSが目指すディペンダビリティのうち、ソフトウェアだけでは対処できないリンク故障への対処や、ネットワーク並びにシステム全体の省電力化、冗長化などのディペンダビリティの向上に寄与できることめどを得た。

さらに、DEOSの基本コンセプトを実現するための支援ツールの一つとして開発した、仮想マシンを用いたfault injection ツール FaultVM について、ユーザが任意のデバイスを追加する機能を加えた。デバイスをシステム記述言語である SpecC 言語を用いて記述し、組み込むことができる。組み込みシステムではユーザが設計した様々なデバイスが利用されることが多く、この機能により、ユーザが独自のデバイスを加えた場合にも対処できる。

一方、各研究チームメンバーからなるサブコアチームにおいて、ディペンダブルシステム全体を評価するためのベンチマークシステム DS-Bench について検討している。これまで、並列分散システムの検証・開発環境のために仮想マシンを柔軟に管理するシステムとして開発してきた D-Cloud を、テストおよびベンチマーク実行環境(TEST-ENV)として利用できるよう、DS-Benchと統合するための検討を行った。

§2. 研究実施体制

(1)「電力制御・高信頼並列システム」グループ

① 研究分担グループ長: 佐藤 三久 (筑波大学・計算科学研究センター、教授・センター長)

② 研究項目

・並列組込み向け高信頼共有メモリ機構および省電力実時間並列実行制御機構

(2)「通信システムアーキテクチャ」グループ

① 研究分担グループ長: 朴 泰祐 (筑波大学・計算科学研究センター、教授)

② 研究項目

・並列システム内高信頼高性能通信機構

(3)「高速ネットワーク」グループ

① 研究分担グループ長: 有本 和民 (ルネサス エレクトロニクス(株)、統括部長)

② 研究項目

- ・低電力高速インターコネクトと省電力高密度並列ハードウェアプラットフォームの開発

§3. 研究実施内容

(文中に番号がある場合は(4-1)に対応する)

1. 並列組込み向け高信頼共有メモリ機構および省電力実時間並列実行制御機構(電力制御・高信頼並列グループ)

当該年度においては、高信頼ソフトウェアを開発するためのフォルトインJECTION環境として開発してきた、仮想マシンを用いた **fault injection** ツール **FaultVM** について、ユーザが任意のデバイスを追加する機能を組込むことに注力した。組み込みシステムにおいては、その用途によって、ユーザが設計した様々なデバイスが用いられることが多いが、通常の仮想マシンにはサポートされていないため、そのままではこのようなケースに対処できない。このようなケースに対処するために、ユーザのデバイスをシステム記述言語である **SpecC** 言語を用いて記述し、組み込むことができる機能を加えた。具体的には、これまで用いてきた仮想マシン **QEMU** を拡張して、独自デバイスを持つ組み込みシステム全体をシミュレーションし、フォルトインJECTION機能を組込む方法について検討し、**SpecC** システム記述言語で記述されたデバイスモデルを利用してソフトウェアまで含めたシミュレーションを可能にする **FaultVM-SpecC** の開発を行った。

また、当該年度も継続して、大規模な並列分散システムの検証・開発環境を提供するための仮想マシンの管理システムである **D-Cloud** を実装および改良を行ってきた。特に、**D-Cloud** が管理する、多くの仮想マシンの計算資源を利用して、多くのテストを加速する仕組みについて検討した。テストツールの一つである、**DejaGnu** を取り上げて、**DejaGnu** で記述されたテストケース群を **D-Cloud** で実行する仕組みを試作した。これにより、システム開発におけるテストケースの実行を加速することができる。

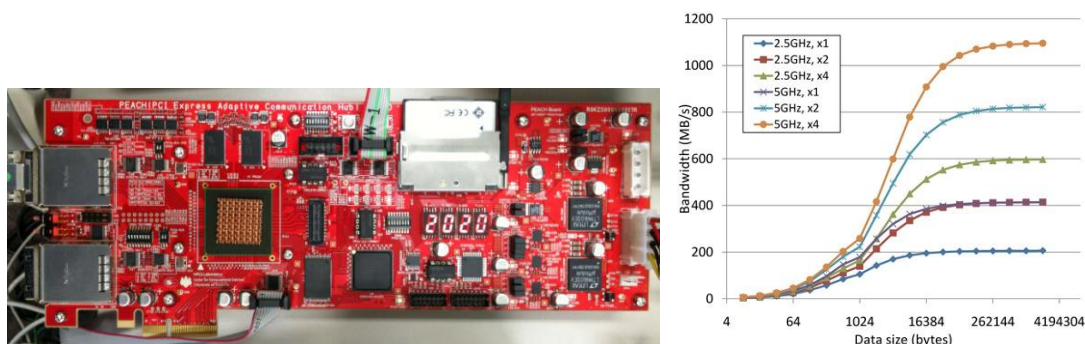
また、ディペンダブルシステム全体を評価するためのベンチマークシステム **DS-Bench** において、**D-Cloud** をベンチマーク実行のための環境 **TEST-ENV** として用いることを検討した。東大石川チームと共同で、**DS-Bench** に **D-Cloud** を統合することとし、検討を行った。

また、実証システムに向けてはマルチノード構成の組み込み向けシステム上で動作する、外付けハードディスクに依存しない耐故障ソフトウェア分散共有メモリである **SCASH/MCAPI** を開発した。今年度の開発では、別途高性能通信システムグループで開発中の、**Ethernet** 上で動作する **MCAPI** ライブラリである **XMCAPI** を用いて実装し、2ノード上で **SCASH** 機能が実現できていることを確認した。

2. 並列システム内高信頼高性能通信機構(通信システムアーキテクチャグループ)

組み込み向け省電力・高信頼・高性能通信リンク **PEARL** を実装するためのコミュニケーションハブである **PEACH** チップのプロトタイプが高速ネットワークグループによって実装完了したことを受け、このチップを各種汎用 PC サーバに応用し様々な実験を行うための **PEACH** テストベッドボードを開発した。**PEACH** テストボードを、**ATOM** プロセッサを用いたホストノードに **PCI-Express** カードとして装着し、これを用いて **PEACH** の基本機能がハードウェア実装できていることを確認

した。さらに、同テストベッド上で PEARL ネットワークを機能させるためのファームウェア開発、ホストノード上の CPU 上で動作する PEARL デバイスドライバを開発した。これらの上で、PEACH プロトタイプチップが設計通りの通信性能を持つこと、各種通信レート及び PCI-Express レーン数に応じた通信性能と消費電力制御のトレードオフ制御が可能であることを確認した。図1に、PEACH テストベッドボードと性能評価の結果を示す。これらの成果は国際会議 SC10 及び国内技術展示会 ET10 においてデモ発表した。



(図 1) PEACH テストベッドボード(左)とレーン及び通信レートに応じた通信性能(右)

3. 低電力高速インターコネクトと省電力高密度並列ハードウェアプラットフォームの開発(高速ネットワークグループ)

省電力・高性能・高信頼通信機構 PEARL の詳細設計及びこれを実現する PEACH チップの作成を行った。通信システムアーキテクチャグループと共同で、PEACH チップの機能及び内部バス性能、割り込み制御回路、そして最も重要な中央制御用プロセッサである 4 core M32R に付随する IP を確定し、テープアウト向け仕様作成を完了し、PEACH チップをルネサステクノロジの 45nm8 層配線、マルチ Vth トランジスタ対応の Low power CMOS プロセスで試作した。チップサイズは 11x11mm²、パッケージは 1008 ピンの BGA、電源電圧はコア部の 1.2V、DDR3 インターフェース部の 1.5V、周辺 IO 部の 3.3V の 3 電源である。試作結果は良好で、PCI Express Gen2 での 4 lane×4 port の通信が確認された。図 2 は PEACH のブロック図とチップ写真、表 1 は PEACH のチップ諸元である。さらに、PEACH チップ完成後、一般の PC 環境において PEACH 及び PEARL のテストと評価を容易に行うための PCI-e カードとして PEACH ボードを作成した。一般の PC サーバ上で PEACH のテストと PEARL としてのノード間接続機能・性能評価を行えるよう、PEACH を搭載した PCI-E ボードである PEACH ボードを開発した。

また、PEARL システム構築の為にソフトウェア作成のハードウェア面での支援のために、PEACH チップ作成前は、シミュレーションでの性能予測と FPGA によるプロトタイプでのソフトウェア作成支援を実施した。PEACH チップ作成後は PCI Express 波形品質の確認を通じて、ボード上で発生する信号品質問題や、クロック分配の問題を解決し、また、ドライバソフトウェア作成上に発生する問題点の解決を通信システムアーキテクチャグループと実施した。

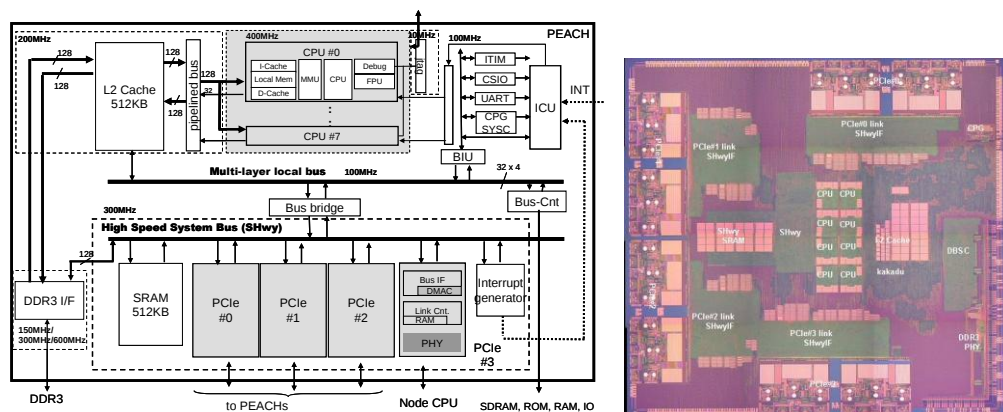


図 2 PEACH ブロック図とチップ図

表 1 チップ諸元

CPU	32-bit Processor (400MHz) x 8 SMP L1-cache:8kB(I)+8kB(D), LM:8kB, MMU, FPU
Memory	L2-cache: 512kB Internal SRAM: 32kB, 512kB
DRAM I/F	DDR3 I/F x 1 SDRAM I/F x 1
PCIe I/F	PCI Express standard Rev.2.0 Transfer speed: 5.0GT/s, 2.5GT/s per lane 4 lanes (20Gbps) x 4 ports Maximum payload size:1024bytes Upconfiguration function Automatic retransmission function Root port / Endpoint selectable
Interrupt Generator	Transfer address, size information register x 3 Automatic transfer mode
Bus	Packet router Multi-layer bus (4-layer) Pipelined bus

§4. 成果発表等

(4-1) 原著論文発表

●論文詳細情報

1. 並列組込み向け高信頼共有メモリ機構および省電力実時間並列実行制御機構(電力制御・高信頼並列グループ)

[1] Toshihiro Hanawa, Takayuki Banzai, Hitoshi Koizumi, Ryo Kanbayashi, Takayuki Imada, and Mitsuhisa Sato, "Large-Scale Software Testing Environment Using Cloud Computing Technology for Dependable Parallel and Distributed Systems," the 2nd International Workshop on Software Testing in the Cloud (STITC2010), co-located with the 3rd IEEE International Conference on Software Testing, Verification, and Validation (ICST 2010), pp. 428-433, Apr. 2010. (DOI: 10.1109/ICSTW.2010.59)

[2] Takayuki Banzai, Hitoshi Koizumi, Ryo Kanbayashi, Takayuki Imada, Toshihiro Hanawa, and Mitsuhisa Sato, "D-Cloud: Design of a Software Testing Environment for Reliable Distributed Systems Using Cloud Computing Technology", the 2nd International Symposium on Cloud Computing (Cloud 2010) in conjunction with the 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing (CCGrid 2010), pp. 631-636, May 2010. (DOI: 10.1109/CCGRID.2010.72)

[3] Toshihiro Hanawa, Hitoshi Koizumi, Takayuki Banzai, Mitsuhisa Sato, and Shin'ichi Miura, "Customizing Virtual Machine with Fault Injector by Integrating with SpecC Device Model for a software testing environment D-Cloud," the 16th IEEE Pacific Rim International Symposium on Dependable Computing (PRDC'10), pp. 47-54, Dec. 2010. (DOI: 10.1109/PRDC.2010.37)

2. 並列システム内高信頼高性能通信機構(通信システムアーキテクチャ・グループ)

[4] 米元大我, 埜敏博, 三浦信一, 朴泰祐, 佐藤三久:トラフィック量に適応する非対称マルチリンク Ethernet トランキング, 情報処理学会論文誌コンピューティングシステム, Vol.3, No.1, pp.25-37, 2010.

[5] Toshihiro Hanawa, Taisuke Boku, Shin'ichi Miura, Mitsuhisa Sato, and

Kazutami Arimoto, "Power-aware, Dependable, and High-Performance Communication Link Using PCI Express: PEARL," IEEE International Conference on Cluster Computing (Cluster2010), poster, 4 pages, Sep. 2010.

[6] Toshihiro Hanawa, Taisuke Boku, Shin'ichi Miura, Mitsuhisa Sato, and Kazutami Arimoto, "PEARL: Power-aware, Dependable, and High-Performance Communication Link Using PCI Express", IEEE/ACM International Conference on Green Computing and Communitations (GreenCom2010), pp. 284-291, Dec. 2010. (DOI: 10.1109/GreenCom-CPSCoM.2010.115)

3. 低電力高速インターコネクトと省電力高密度並列ハードウェアプラットフォーム(高速ネットワークグループ)

[7] S. Otani, H. Kondo, I. Nonomura, A. Ikeya, M. Uemura, Y. Hayakawa, T. Oshita, S. Kaneko, K. Asahina, K. Arimoto, S. Miura, T. Hanawa, T. Boku, M. Sato, "An 80Gb/s Dependable Communication SoC with PCI Express I/F and 8 CPUs", International Solid-State Circuits Conference (ISSCC2011), pp. 266-268, Feb. 2011. (DOI: 10.1109/ISSCC.2011.5746312)

(4-2) 知財出願

① 平成22年度特許出願件数(国内 0 件)

② CREST 研究期間累積件数(国内 0 件)