# Machマイクロカーネルをベースとした並列OS DenEnの実現

3H－8

高野 陽介, Christopher Howson, 荒木 宏之, 菅原 智義, 小西 弘一, 小長谷 明彦
NEC C&C研究所

## 1 はじめに

分散メモリ型、すなわち、多数のプロセッサを共有メモリを使わずに高速なネットワークで結合する並列コンピュータが内外で開発され商用的にも利用され始めている。**DenEn**はこのような分散メモリ型並列コンピュータをターゲットとして、高度な計算・入出力能力をユーザに効率よく伝達する制御を実装し、並列計算の利用を容易化するプログラミング支援環境を提供することを目的としている並列OSである。本稿では、分散メモリ型の並列コンピュータ**Cenju-3**[2]をターゲットとして実現した**DenEn Version 1.0**の狙いとソフトウェア構成について概説する。

## 2 システムの狙い

DenEnは、並列コンピュータの持つ計算能力をユーザプログラムに直接伝達し、高速なアプリケーションの実現を支援することを狙いとして持つ。この目的を達成するためには、アプリケーションと同一のプロセッサ上で実行されるOSが生み出すオーバヘッドは最低限であることが望ましい。例えば、OSがCPUパワー、メモリ、入出力能力、通信能力などの計算機資源を無駄に消費しないようにすべきである。

DenEnでは、この目的のためにCenju-3の各要素プロセッサ上に**Machマイクロカーネル**[1]を搭載した。Machマイクロカーネルは、CPUとメモリの制御、プロセッサ間通信機能、デバイスドライバなどを提供する小規模なOSである。マイクロカーネルは計算機資源の消費が少ないため、アプリケーションはより多くの計算機資源を専有できる。Machはオブジェクトサイズは1Mバイト足らずの小規模なものとはいえ個々のOS機能はかなり高度なものを持つ。

その一方で、アプリケーションがMachの提供しないOSサービスを必要とする場合にはMachでは次のように解決できる。まず、任意の要素プロセッサ上にそのようなOSサービスをマイクロカーネルの上位のサーバプログラムとして実装することができる。Machの通信機構を使うことにより、別の任意の要素プロセッサ上で動作するアプリケーションはあたかも同一プロセッサ上にそのサービスがあるかのごとくサービスを利用することができる。

DenEnでは、この方法に従い一部のOS機能を特定の要素プロセッサ上にサーバプログラムとして実装している。さらに、Cenju-3に付属するEWS4800ワークステーションではUNIXのフルセット(SystemV 4.2)が動作することを利用し、このUNIX機能も要素プロセッサ上のアプリケーションは必要に応じて仮想的に利用可能になっている。

このように、DenEn上では要素プロセッサの持つ計算機資源を専有して自由度の高いアプリケーションのプログラミングが可能である。

## 3 システム構成

DenEn Version 1.0の構成を図1に示す。Cenju-3はネットワーク結合された要素プロセッサの集合とそれに制御用に付属するホストコンピュータ(EWS4800ワークステーション)で構成される。machiplというDenEnのコンソール端末の機能を実現するプログラムがこのホストコンピュータ上で動作する以外はCenju-3の要素プロセッサ上で動作する。

### 3.1 マイクロカーネル

OSのベースとしてCMU版のMachマイクロカーネルを採用している。特徴は、Machが持つネットワーク透過なプロセッサ間通信機能である**NORMA-IPC**[3]を高速性・安定性の観点からほぼすべて再実装した点である。この通信機能は、通常の通信、システムコール、デバイスアクセスなどをプロセッサ間においてネットワーク透過に扱うことができるため、任意のプロセッサ上で提供されるOSサービスを任意のプロセッサ上で動作するアプリケーションプログラムから利用可能にする。アプリケーションが遠隔プロセッサ上のOSサービスをあたかもローカルのサービスのようにアクセスできるのはこのNORMA-IPCの機能による。

"Overview of Parallel Operating System *DenEn* Based on Mach MicroKernel"
Yosuke TAKANO, Chiristopher HOWSON, Hiroyuki ARAKI, Tomoyoshi SUGAWARA, Koichi KONISHI, Akihiko KONAGAYA
NEC Corporation
4-1-1 Miyazaki,Miyamae,Kanagawa 216, Japan

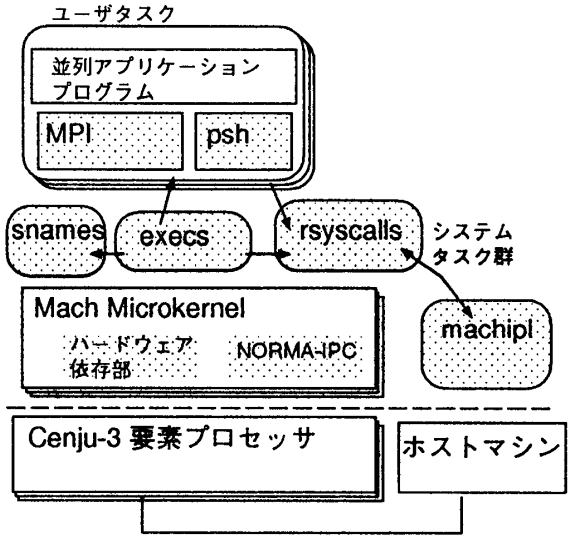


図 1: DenEn Version 1.0 の構成

### 3.2 OS サービス

アプリケーションプログラムは要素プロセッサの持つ計算機資源を専有でき、さらに以下で述べる OS サービスを遠隔利用できる。OS サービスは、Cenju-3 の1つの要素プロセッサ上で動作する以下の3つの OS サーバにより実現されている。

- snames サーバ: サーバの名前の解決を担当する。
- execs サーバ: アプリケーションの要求に応じて、任意の要素プロセッサ上にアプリケーション (タスク) の生成・消滅を実施する。
- rsyscalls サーバ: Cenju-3 に付属する EWS4800 と通信し、EWS4800 上の UNIX のシステムコールを要素プロセッサ上から利用可能にする。

DenEn 上のアプリケーションはこれらの OS サーバが提供する機能、および、rsyscalls により仮想的に提供される UNIX の機能を利用することができる。このほか、

- Mach がもともと持つマルチスレッドなどの機能
- 要素プロセッサに接続されたディスク装置へのデバイスドライバレベルからのアクセス

なども利用可能である。

### 3.3 プロセッサ間通信機構

Cenju-3 ではプロセッサ間に共有メモリがないため、どのようなプロセッサ間通信が提供されるかがアプリケーションの実現性・有効性を大きく左右する。

このため、DenEn は通信の方法に選択肢を用意する。提供するのは

- NORMA-IPC
- MPI/DE

の2種類の通信方法である。NORMA-IPC は前述のようにマイクロカーネル内で実装され、システムコールやデバイスアクセスの伝達などに使用することができる。このため、アプリケーションが OS サービスを利用する際、あるいは、OS サービスを実現するために OS サーバ間で通信を行なう際の利用に適している。他の研究機関等で開発される Mach 用の OS サーバはこの NORMA-IPC を使って実装されているため、その移植性を保証する意味でも NORMA-IPC は重要である。

MPI/DE は、科学技術計算を行なう並列アプリケーション向けの通信インターフェイスとして広まりつつある **MPI**(Message Passing Interface)[4] を DenEn 上で実装したものであり、多彩な通信機能を提供する。実装面では、通信用のハードウェアを OS を経由せずにアプリケーションで使用することができるユーザレベル通信の機能を一部導入して NORMA-IPC よりも高速な通信を達成している。

このように、NORMA-IPC はプロセッサにわたるサーバ・クライアント的な通信に用い、MPI/DE は並列アプリケーションの内部でプロセッサ間の通信に用いるように使い分けられる。

## 4 おわりに

並列 OS DenEn の狙いと構成について概説した。現在、次版に向け、マルチユーザ利用環境の整備、プロセッサ間通信の高速化を進めている。また、今後の課題は、現在ホストコンピュータに依存している UNIX 機能を要素プロセッサ上で効率よく実現すること、および Cenju-3 の外部のコンピュータとの連携機能を強化することである。

## 参考文献

[1] David B. Golub, Randall Dean, Alessandro Forin and Richard Rashid. "UNIX as an Application Program". *USENIX Summer Symposium*, Jun, 1990.

[2] 広瀬他. "並列コンピュータ Cenju-3 のアーキテクチャ". 情報処理学会研究報告, ARC-107-16, pp.121, Jul, 1994.

[3] Joseph S. Barrera III. "A Fast Mach Network IPC Implementation". *USENIX Mach Symposium*, pp. 1-18, Nov. 1991.

[4] The MPI Forum. "MPI: A Message Passing Interface". *Proceedings of Super computing'93*, Nov. 1993.