

## KOCOS のアーキテクチャ (2)

### —「プロセス間通信方式とソフトウェア構成」

徳田英幸 上林憲行 竹山明 石塚朝生 (慶大工学部)  
西垣秀樹 平塚良治 (伊電気工業(株))

#### 1. はじめに

近年、いろいろな分野において数多くのミニコンピュータ・コンプレックスが、提案され、又、稼動しはじめている。<sup>[1]</sup>しかし、それらの多くの場合、計算機間での結合方式に関する研究に重点が置かれ、コンプレックスシステムの環境を生かしたオペレーティングシステムに関する研究は十分な成果をあげていないように思われる。オペレーティングシステムの作成に際しては、できるだけ既存のシステムソフトウェアを効率よく活用して、構成していく例が多く見られ、システム作成当初から、新しい概念で構成されたという例は少ない。この主な理由としては、SUEシステム<sup>[2]</sup>などのようなコンピュータ・コンプレックスを意識したアーキテクチャをもつ計算機システムがあまり開発されておらず、従来からの単体のプロセッサを接続し、既存のソフトウェアに極力手をかけずに実現しようという要求が強かったためであろう。従って、既存のコンピュータ・コンプレックスには、相手プロセッサへそのプロセッサのリモート・ジョブ・エントリ機能を紹介して交信するといった“ルーズ”な結合形態が多く見られた。

汎用性のあるリソースシェアリング、ロードシェアリングを目的とし、さらに並列処理をめざしたミニコンピュータ・コンプレックスにおけるオペレーティングシステムに課せられた課題としては次のものがある。

- [1] システム全体の信頼性の向上
- [2] 各ミニコンピュータの処理能力の限界
- [3] プロセス間通信におけるオーバーヘッドの軽減
- [4] 使いやすさ、ユーザの生産性の向上
- [5] システムの拡張性、可変性の実現
- [6] システムソフトウェアの生産性の向上

筆者らは、これらの[1]~[6]を考慮し、ミニコンピュータコンプレックスの可能性を追求すべく、リソースシェアリング、ロードシェアリング、並列処理を指向する異種ミニコンピュータ・コンプレックス“KOCOS”(Keio-Ok's Complex System)におけるオペレーティングシステム“KOCOS”(Keio-Ok's Complex Operating System)を設計した。本稿では、KOCOSの基本的性格であるプロセス間通信機能、システムソフトウェアの構成について述べる。

#### 2. オペレーティングシステムKOCOSの基本的性格

KOCOSの機能は、単一バス方式による結合形態、接続しているミニコンピュータの処理能力、BIU (Bus Interface Unit)の中枢部に採用したマイクロプロセッサの機能、DBC (Distributed Bus Controller)やBM (Bus Monitor)の機能<sup>[3]</sup>と密接に関係しており、次のような性格をもっている。

[1] システムの管理機構は、“集中分散管理方式”を採用している。これは、システ

ム内に分散して存在するハードウェアリソースやソフトウェアリソースの管理、ジョブのスケジューリングに対しては、SS (System Scheduler) による集中管理方式をとり、各EP (Element Processor) 内で実行されるユーザプロセスやシステムプロセスの実行時の制御に関しては、各EP上に実現されているLOS (Local Operating System) による分散管理方式をとることを示している。

[2] システムモジュールの一部に生じた障害が、システム全体をダウンさせるという事態を生じさせないよう、多レベルの障害対策機構を備え、システム全体の高信頼性を実現している。

[3] 従来のネットワークで採用されているメッセージ転送を主としたルーズなプロセス間通信機能ではなく、プロセス間での直接的な通信を実現している。さらに、プロセスを制御する同期操作も備えているので、コオペレーティブなプロセスの制御<sup>[4]</sup>が、容易に実現でき、並行プログラムを効率よく処理することができる。

## 2.1. プロセス間通信機能 (IPCF: Interprocess Communication Facility)

プロセス間通信機能は、ミニコンピュータコンプレックスに課せられているリソースシェアリング、ロードシェアリング、並列処理環境の実現に関して、もっとも大きな要因であるといわれている。<sup>[5][6][7]</sup>

プロセス間でのメッセージ転送に際し、メッセージの組み立て、分解や送受信時のコントロール情報の解析といったわずらわしい作業がシステムオーバヘッドとして影響してくると、システムのスループットが低下するだけでなく、多量データを高速転送する必要のあるリアルタイム処理などは、著しく制約されてしまう。

従って、ミニコンピュータコンプレックスにおけるプロセス間通信機能としては、次のような性格を備えておくべきであろう。

- [1] プロセス間通信の際のオーバヘッドを極力のぞき、各プロセッサへの負荷を軽減する。
- [2] ユーザが、ある入出力装置に対して読み書きすると同様に、プロセス間での交信が直接的にでき、かつ、スムーズに複数のプロセッサ内のプロセスによるコオペレーションが行える。
- [3] メッセージ転送の際のエラーに対する自動再試行機構や、フォールバック機能を備えた高信頼性を保証する。
- [4] メッセージ転送に関し、ブロック長やバス使用優先レベルなどの物理的転送パラメタもできるだけシステムのトラフィックにつりあうよう動的にセットでき、システム内で調和のとれたプロセス間通信ができる。
- [5] ブロードキャスト転送や、通信をしている各プロセスが所有しているバッファ量の違いを吸収できる部分転送機能や、転送完了情報の遅延機能を備えている。

KOCOSでは、以上の点を考慮し、論理的な結合を密とするため、送信プロセスの番地空間から受信プロセスへの番地空間へ向けてのDMA (Direct Memory

Access) 機能を生かした直接的な通信方法を採用した。さらに、メッセージ転送の際の種々の仕事や複雑な割込み処理などから、プロセッサを解放し、BIUの中枢部に高度な知的管理機能を付加したので、オーバーヘッドをなくし、LOSへの負荷を軽減しただけでなく、成長可能なIPC Fを設計することができた。

このようなIPC Fの階層構造は図-1.のように表わすことができる。

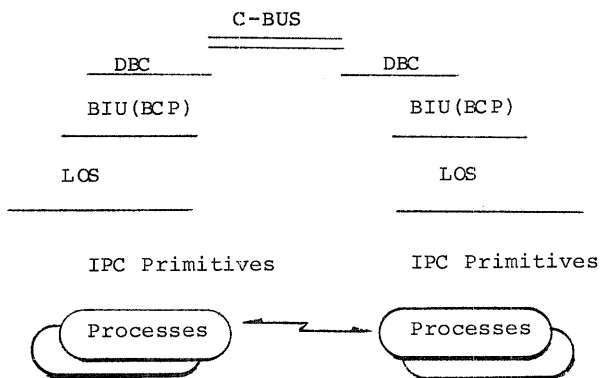


図-1. IPC Fの階層構造

### 2.1.1. プロセス間通信基本命令 (IPC Primitives)

プロセス間通信は、プロセス間通信基本命令を実行することにより行うことができる。この基本命令は、すべてシステムコールという形式で実現されており、図-2.に示されるように分類することができる。

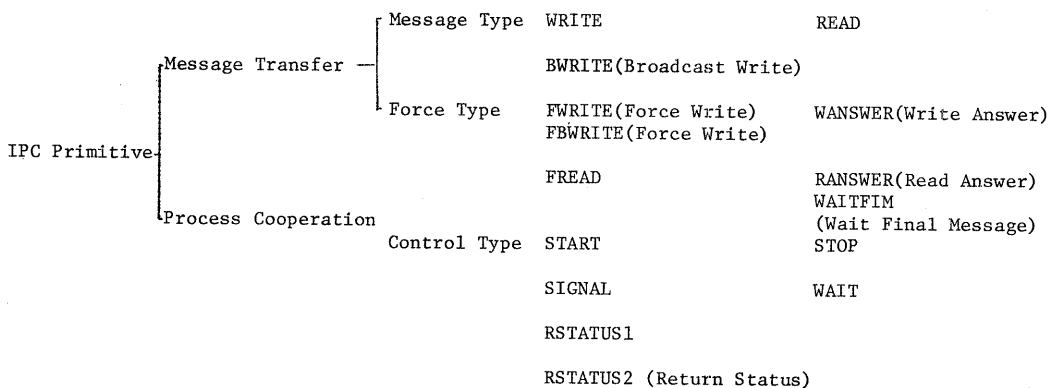


図-2. プロセス間通信基本命令

メッセージ転送を行なう基本命令は、相手プロセスとのマッチングがメッセージ単位に、BIUのランデブーテーブル(後述)上で行なわれロジカルリンクが確立するメッセージタイプ(M-タイプ)と、相手プロセスに対し、強制的にメッセージ転送の旨を伝え、ロジカルリンクを確立するフォースタイプ(F-タイプ)の2種類がある。プロセスの相互制御を行なう基本命令は、コントロールタイプ(C-タイプ)といい、LOSの提供しているダイナミックなプロセスの制御機能を活用し、複数EP内のユーザプロセスやシステムプロセス間でのコ・オペレーションを実現している。

### 2.1.2. プロセス間通信方式

プロセス間通信方式については、プロセス間通信基本命令の3つのタイプに代表される次のような例をあげることができる。

- (1) M-タイプ: あるプロセスαが、他EP上のプロセスβに対し、aa番地

以降の語をWRITEする場合。

(2) F-タイプ: あるプロセス $\alpha$ が、他EP上のプロセス $\beta$ に対し、 $aa$ 番地以降の語をFWRITEする場合。

(3) C-タイプ: あるプロセス $\alpha$ が、他EP上のプロセス $\beta$ をwake upする場合。

これら(1)~(3)に対する交信手順は、それぞれ図-3、図-4、図-5のように表わすことができる。ここでは、図-3のM-タイプにおけるWRITE-READについて図中の番号によって詳細について述べる。

1. プロセス $\alpha$ がシステムコールWRITEを実行する。

2. これを受けつけて、L O S 1のIPC handler (Interprocess Communication Handler) が、BCW (BIU Control Word; チャンネルコマンドを機能拡張したものに相当する。)の語数と、バス使用要求の優先順位の情報を知らせるためにBIU1に対してListen I/Oと呼ぶ入出力命令を出す。次に、BCWのアドレス $xx$ を知らせ、BIU1に対して起動をかけるStart I/O命令を実行する。この後L O S 1は、プロセス $\alpha$ をwaiting状態にし、他プロセスにプロセッサを割り当てるためのスケジューリングを行う。

3. BIU1は、BCWをEP1のメインメモリから読み出し、その解読を行う。解読後、ファンクション部がWRITEであるので、ランデブーテーブル (RT; プロセス間でのメッセージ転送に関するコントロール情報のマッチングをとるためのデータポートをいう。)上で、プロセス $\alpha$ - $\beta$ 間の転送のためのパラメタをセットする。(但し、既に、プロセス $\beta$ からのREAD要求が到着していればステップ5へ移る。)

4. プロセス $\beta$ からのREAD要求が到着するまで、RT上で、待っている。

1. プロセス $\beta$ がシステムコールREADを実行する。

2. L O S 2は、これを受けつけ、Listen I/O、Start I/O命令を実行する。

3. BIU2は、Start I/Oで知らされたBCWのアドレス $yy$ より読み込み、ファンクション部を解読後、READであるので、 $yy$ 番地以降の語をIPC M (Interprocess Communication Message; メッセージ転送のためのコントロール情報をいう。)として、BIU1へ転送する。

4. BIU1は、このIPC Mを受信し、解読後、RT上でプロセス $\alpha$ とプロセス $\beta$ とのマッチングが成立する。

5. BIU1のRT上でのマッチング成立後、BIU1は、DMAを起動してプロセス $\alpha$ の $aa$ 番地以降の語をプロセス $\beta$ の $bb$ 番地以降の語に向けて転送を開始する。転送完了後、BIU1のRT上のパラメタは解消され、転送終了をしめすFIM (Final Message) をプロセス $\alpha$ 、 $\beta$ へ向けて転送する。

以上のようなプロセス間通信方式により、各L O Sは、BIUに対してListen I/O、Start I/O命令を出すだけで、後のメッセージ転送に関しては、一切関与していない。転送に際して、RTは必ず送信側BIUにセットされており、実際の転送管理を送信側BIUが行っている。このように、プロセス間通信の際の種々の仕事をBIUへ負わせたので、以下のような特徴をもったIPC Fを容易に実現することができた。

[1] プロセスの物理的位置に関係なく自由に、直接的にプロセス間通信が行える。すなわち、RT上でIPC Mのマッチング成立後、送信プロセスの番地空間から受信プロセスの番地空間へDMA機能を利用してL O Sに全く無関係に転送を行う。従って、L O Sは、メッセージ転送のための特別なバッファを

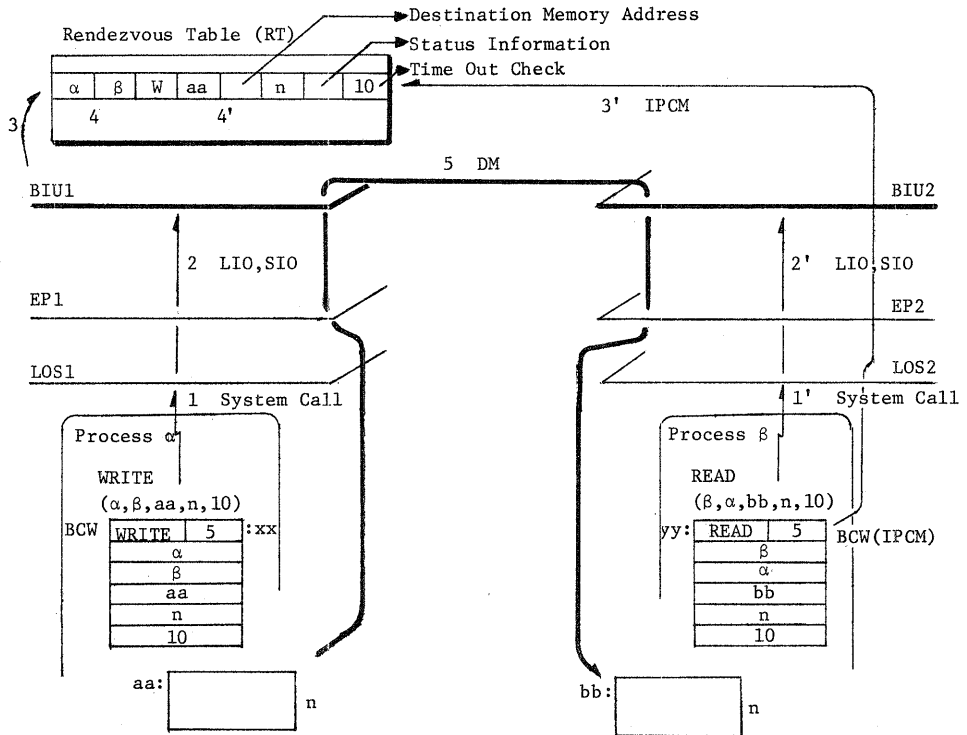


図-3. M-タイプ プロセス間通信例 (WRITE-READ)

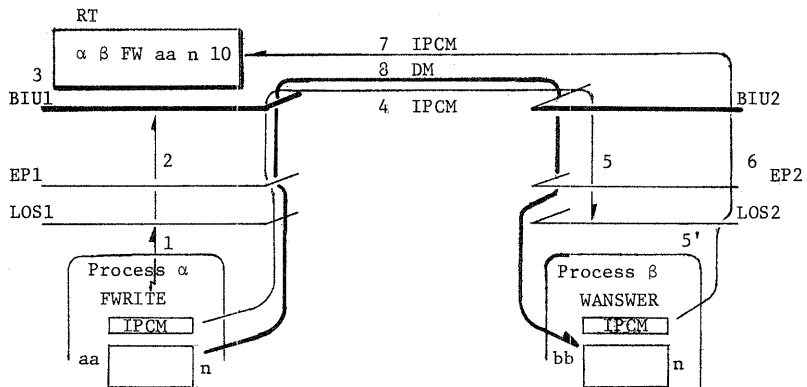


図-4. F-タイプ プロセス間通信例 (FWRITE-WANSWER)

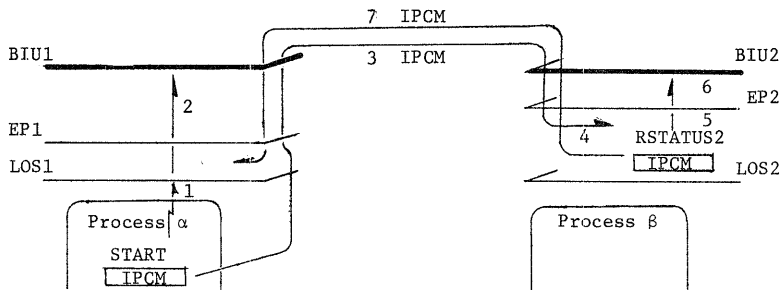


図-5. C-タイプ プロセス間通信例 (START-RSTATUS2)

持つ必要がなく負荷が軽減された。

- [2] M-タイプ, F-タイプの2種類の転送方式が可能であり, 部分メッセージ転送, 離散型データ転送, one to any転送の高度な機能を備えている。
- [3] LOSの管理下にある複数プロセスによる同時送信, 同時受信が可能であり, 又, 送信時には, メッセージの優先順位に基づいて送信可能である。
- [4] IPCMを基礎とした制御情報を介しての交信なので, 結合形態に拘束されずに, 機能拡張が容易に行える。
- [5] Cタイプのプロセス間基本命令により, コンアレックスシステムの環境を生かしたコオペレーティングプロセスが実現できる。
- [6] LOSのサポートを受けず, IPCFとして転送時のエラーに対するバックアップ機能を備え, 高信頼性を実現している。

## 2.2 システムソフトウェア

### 2.2.1 システム管理形態——SSとLOSのロジカルな関係

KOCOSのようなコンピュータコンプレックスでは, ジョブの制御, リソースの管理を集中管理にするか分散管理にするかは重要な問題である。KOCOSではこの問題に対して, 次のように考えて設計した。すなわち, これらの機能を各EPに分散すると, システム全体の信頼性, 負荷の分散化という面で望ましい。しかし, 各EPの処理能力にかなりの制約があるため, かえって負荷が増加するだけでなく, システムの拡張に不都合が生じてしまう。従って, これらを守るため1台のEPにジョブの制御, リソースの管理を行わせる集中管理方式(この制御プログラムを実行するEPをシステムスケジューラ(SS)という。)を採用した。SSがダウンしたときには, 他EPが新しいSSとして機能するようなバックアップ体制を備えることにより, 信頼性の問題を解決している。一方, 各々のEP内で実行されるプロセスの管理に関しては, それぞれ独立に行う分散管理方式(このような各EPの制御プログラムをローカルオペレーティングシステム(LOS)という。)を採用した。このような集中分散管理方式を行っているKOCOSの制御環境の論理的な関係は, 図-6のような高さの低い円錐形で表わすことができる。円錐の頂点にはSSが位置し, 底面の円周上には, 各LOSが分散

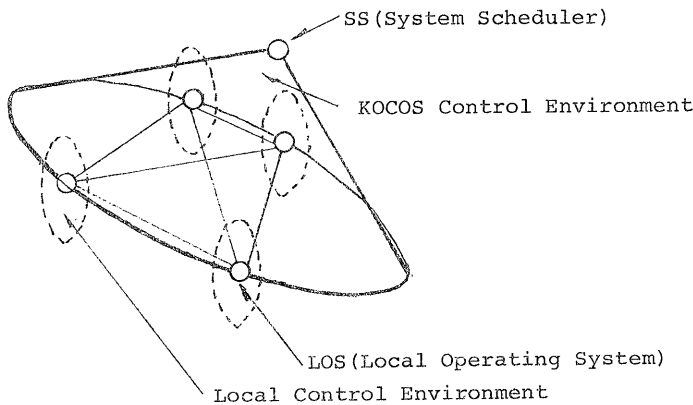


図-6. KOCOS 制御環境の論理的関係図

して存在している。各LOSの管理下にあるプロセス間のコミュニケーションは, SSを介さず, 自由に行うことができ, SSとも自由に行うことができることを示めている。KOCOSの各LOSはいくつかのユーザプロセス, システムプロセス, およびI/Oデバイスなどのローカルリソースと共に1つのローカル制御環境を形成している。又, SSおよびLOSで

形成されている円錐は、KOCOSの制御環境を表わしている。

### 2.2.2. システムソフトウェアの構成

KOCOSでは既存のソフトウェアの有効的活用を考慮しながら、さらにプロセスレベルでの並行処理を考慮したコンプレックス・オリエンテッドなシステムの開発を目指している。又、システムソフトウェアの構成に関しては、高度なモジュール性を持たせ、システムの可変性、拡張性に対処している。SSやLOSの開発および製作において、構造的アプローチ<sup>[8][9]</sup>を更実践し、異種E Pとの互換性を考慮して、システム作成用言語<sup>[10][11]</sup>による設計をし、ポータビリティ、トランスピリティを高めると同時に信頼性の向上をめざしている。次に、図-7によりSSの階層構造を示めし、図-8によりLOSの階層構造を示めす。LOSの各レベルの機能は、SSの各レベルとほとんど同一の機能なので、ここではSSについて概説する。さらに、図-9にプロセスステートダイアグラムを示めす。

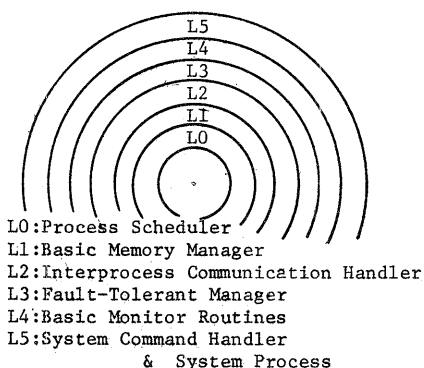


図-7. SSの階層構造

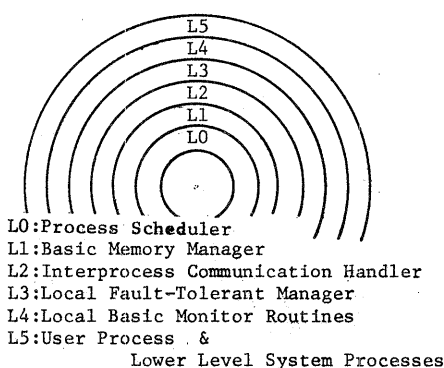


図-8. LOSの階層構造

#### ・SSの各レベルの概略機能

レベル0：プロセススケジューラ

リソース管理、ジョブ管理などを行うSS内のシステムプロセスのスケジューリングを行う。スケジューリングアルゴリズムは、各プロセスのもつプライオリティをもとにしたマルチプライオリティ方式である。障害対策用のプロセスには最も高いプライオリティが与えられている。

レベル1：基本メモリ管理

SSが使用するワークスペースのアロケーションや、フリーになったストレージスペースの管理を行う。

レベル2：IPCハンドラ

SS内のプロセス間通信の実際の管理を行っている。ローカルプロセスとリモートプロセス間でのメッセージ転送だけでなく、プロセスのコオペレーションに関係するプリミティブをも管理している。

レベル3：障害対策ルーチン

システム内のE PやBIUに生じた障害の早期発見、識別、システム再構成、再開を行う。

レベル4：基本モニタルーチンズ

このレベルでは、SS下のハードウェアハンドリングルーチン、リソースやジョブを管理するルーチンなどSSの主要な機能が行われる。

レベル5：システムコマンドハンドラ

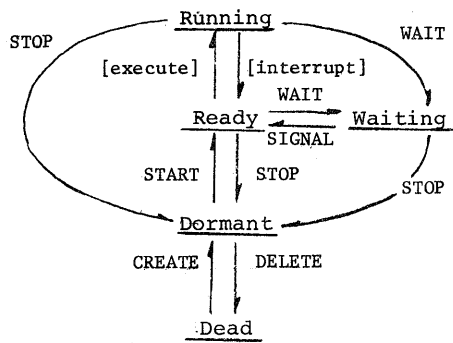


図-9 プロセス・ステート・ダイアグラム

や転送語数のカウントなどの前処理の後、SSへ転送される。SSのシステムコマンド・ハンドラはそれを解読し、要求された手続きを行い、その結果を該当ローカル・コンソールへ知らせる。次に、基本的なシステムコマンドを表-1に示す。

ユーザがキーインしたシステムコマンドが、どのようなIPC Primitiveを利用してSSのシステムコマンド・ハンドラへ転送されていくか、簡単なモデルを用いてその処理フローを図-10の番号にそって解説する。

1. ローカルコマンド・ハンドラ (\*LCH; 各LOSに存在する)が、キーボード・リダ (\*KBR)をwake-upし、FREADを実行する。
2. \*KBRは、システムコマンドをKey Boardから実際に読み込むための、前処理を行い、1ラインスカ完了まで自分自身をブロックする。
3. キーボードからコマンドがキーインされ、\*KBRのバッファにバッファリ

表-1. 基本的システムコマンドの一覧表

| コマンド名  | オプション   | パラメータ            | 機能                                    |
|--------|---------|------------------|---------------------------------------|
| \$CON  | A       | SMN              | ローカルな資源をKOCOSのシステム資源として登録する           |
| \$DCON | A       | SMN              | KOCOSの管理下にあるシステム資源をローカルにする            |
| \$JOB  |         | JOB-id, pro-n, L | ユーザ・ジョブを宣言する                          |
| \$FIN  |         |                  | ジョブ・ステップの終了を宣言する                      |
| \$END  |         |                  | ジョブの終了を宣言する。                          |
| \$RES  |         | SMN {:LFN}       | ユーザ・ジョブで必要とするシステム・リソースを確保する           |
| \$REL  | A       | SMN {:LFN}       | 確保したリソースを解放する                         |
| \$LOAD | L, H    | SFN>SMN          | システムに登録されているファイルを指定のシステム・モジュールへロードする。 |
| \$BRK  |         |                  | ユーザ・プロセスを停止し、システムコマンド・インバットモードにする。    |
| \$RUN  |         |                  | ロードされているプロセスの実行を指示する                  |
| \$DISP | A, U, R | MN               | システム・リソースの状態をユーザへ知らせる。                |
| \$MSG  | A       | Message, SMN     | ローカル・コンソールへメッセージを転送する。                |

コマンドの一般形

\$<コマンド名> <オプション> <パラメータ1>, ... , <パラメータM>

A: All    L: Local    H: Here    U: User    R: Resource

SMN: System Module Name    IFN: Logical File Name    L': Limit Parameters

SFN: Standard File Name    MN: Module Name



"Local Command Handler"

repeat

SIGNAL(\*LCH,\*KBR);

1: FREAD(\*LCH,\*KBR,aa,n,cl);

    syntax check and set parameter;

5: FWRITE(\*LCH,\*SCH,aa,n,cl);

7: READ(\*LCH,\*SCH,bb,m,cl);

    print out to local cosole;

forever

"System Command Handler  
Receiving process \*SCH"

repeat

WANSWER(\*SCH,\*,dd,n,cl);

6: semantics check and  
    set parameter

WRITE(\*SCH,\*,dd,m,cl);

forever

"Key Board Reader"

repeat

2: WAIT(\*KBR,\*KBR);

    key board release and  
    interrupt enable;

    clear(k);

    P(ucb.sem);

4: RANSWER(\*KBR,\*,buf,n,cl);

    set process name;

    WAITFIM(\*KBR,\*);

forever

var buf: buffer area for \*KBR;  
    k : buffer pointer  
    c : clock limit

"interrupt from key board"  
{get 1 character from Areg}

    buf(k):= Areg;

    k:= succ(k);

    if buf = CR  
        then release and RTI;

3: V(ucb.sem);

    RTI;{return from interrupt}

図10. システムコマンド内部処理フロー

- グされる。読み込んだ文字がCR(キャリッジリターン)であれば、ucbのセマフォsemに対しVオペレーションを行い、\*KBRをwake upする。
4. wake-upされた\*KBRは、\*LCHへ向けて、RANSWERを実行し、転送完了までブロックする。
  5. \*LCHは、SS内のシステムコマンドハンドラの受信プロセス(\*SCH)に向けて、FWRITEを実行する。
  6. \*SCHはdd番地以降にシステムコマンドを受信し、そのコマンドで要求された処理を行い、その結果を\*LCHへ向けてWRITEする。
  7. \*LCHは\*SCHの返信を受けとり、ローカルコンソールへプリントアウトする。

### 3. 今後の計画

KOCOSは、プロトタイプの開発だけで終了するのではなく、長期的なプロジェクトである。プロトタイプKOCOSを基礎にして、将来的にはユーザの便宜をはかるために以下に列挙するものを開発し、KOCOSを総合システムに発展させる予定である。

[1] 異種間の共有ファイルシステムの開発

KOCOSの形態に適合した共有ファイルシステムの開発

[2] 並列処理

共通言語の開発を含めて並列処理の環境を整備する。

[3] リングサブシステム

種々の入出力機器をリング状に結合し、一括してC-Busに結合する。

[4] デバイス指向のBIU

プロセッサ用のBIUだけでなく、種々の入出力機器を直接C-Busに結合するBIUを開発する。

[5] 大型計算機とのリンク

本学情報科学研究所に設置されているUNIVAC-1106とKOCOSを通信回線を介して結合し、ユーザにより強力なサービスを提供する。

### 4. おわりに

SS, LOSの製作に際し, システム作成用言語を用いて, ソフトウェアの生産性, 信頼性の向上をはかっている。しかし, それぞれのミニコンピュータでの割り込み機能の違いなどの物理的環境の細部における相異は, 完全に吸収することができず今後の課題といえよう。又, システム全体の信頼性を高めるためにも各EPにはメモリアプロテクション機能, インターナルクロックを備える必要があるといえる。しかし, KOCOSのアプローチにおける特徴は, 既存のミニコンピュータに変更を加えずに論理的な結合度を密に複合化できる。又, BIUの中核部にマイクロプロセッサを採用し, プロセス間通信の際のオーバヘッドを軽減しただけでなく, 高度なメッセージ転送とプロセスの制御が可能な強かなIPCFにあるといえる。

現在, BIU, DBCは, 本年9月末の完成予定であり, LOS, SSに関しては, 基本設計が終了し, 製作中である。

### 謝辞

KOCOSの開発設計にあたり御協力, 御指導いただいた沖電気工業(株) 杉浦宜紀氏, 松下温氏, さらに本学 相磯秀夫教授, 同 土居巖久講師, 清水洋彦氏(現日本航空(株))に深謝します。又, 論文の作成にあたり御協力いただいた本学 阿多靖広君, 石井和喜君, 瀧塚博志君, 西尾忠幸君, 福島知善君, 松尾繁樹君に深謝します。

### 参考文献

- [1] 元岡達 : コンピュータコンプレックスの展望, 情報処理, vol.15, no.7, pp.525~533, 1974.
- [2] Lochkeed Electronics Company : SUE Computer Handbook, Los Angeles, 1972.
- [3] 上林, 徳田, 竹山, 石塚, 西垣, 平塚 : KOCOSのアーキテクチャ(1) - ソフトウェアとシステム構成, 情報処理アーキテクチャ研究会資料, 1975年7月.
- [4] Dijkstra, E. W. : Cooperating Sequential Processes, Programming Languages, Academic Press, pp.43-111, 1968.
- [5] Walden, D. C. : A System for Interprocess Communication in a Resource Sharing Computer Network, Comm. ACM, vol.15, no.4, pp.221-230, Apr. 1972.
- [6] Akkoyunlu, E., Bernstein, A., Schantz, R. : Interprocess Communication Facilities for Network Operating Systems, Computer, vol.7, no.6, pp.46-55, Jun. 1974.
- [7] Akkoyunlu, E., Bernstein, A., Schantz, r. : An Operating System for Network Environment, Proc. Symposium on Computer-Communications Networks and Teletraffic, Polytechnic Press, New York, pp.529-538, 1972.
- [8] Dijkstra, E. W. : The Structure of the "THE" Multiprogramming System, Comm. ACM, vol.11, no.5, pp.341-346, May 1968.
- [9] Hansen, P. B. : The Nucleus of a Multiprogramming System, Comm. ACM, vol.13, no.4, pp.238-241, Apr. 1970.
- [10] Wirth, N. : The Programming Language PASCAL, Acta Informatica, vol.1, no.1, pp.35-63, 1971.
- [11] Wulf, W. A., et al., : BLISS Reference Manual, Computer Science Department Report, Carnegie-Mellon Univ., Pittsburgh, 1971.