

## 大型機における状態表示コマンドについて

東京大学大型計算機センター

石田晴久・野本征子

## §1. はじめに

ハードウェア・モニタあるいはソフトウェア・モニタによるコンピュータの稼働状態の測定および性能評価はあちこちで行なわれているが、それを行うのは、通常、システム関係の技術者や研究者、あるいはシステムの管理責任者に属する人々に限られている。しかもシステムの状態の測定を行うには、何らかの準備が必要であり、また常時測定することはオーバーヘッドを大きくする原因になりうることもあって、そうした測定はいつでもできるようにはなっていないことが多い。

しかし、大学の計算センターの大型コンピュータ・システムのように、TSS 利用も含めて常時多種多様なユーザにサービスを提供しているようなシステムでは、いろいろな意味のシステムの状態が常に誰にでも測定できるようになっていることが望ましい。これはシステムの管理改善や改良や次期システム設計用のデータ取得などをいろいろな角度からいろいろな人によって行う必要があるためでもあり、また混雑のひどいピーク・タイムにレスポンスやターンアラウンドの遅い理由をユーザに納得させる手段としたいためでもある。

さらにシステムの状態が常時誰にでも分かるようになっていけばガラスばりのシステムでは、能率の悪い運転(システム)は容認されなくなるから、メーカーとしても、運用の管理者としても、システムを改善して効率よい運用をせざるをえなくなるから、ユーザ・レベルでの測定が可能になることは、システム全体の改良にも望ましい。

そこで我々は、システムの状態をバッチ処理でもTSS でもユーザ・レベルのコマンドで簡単に知ることができるようにする目的で、一般ユーザで使用可能なくつかの命令を開発することを試みた。対象としたシステムは東京大学大型計算機センターのHITAC 8800/8700システム(CPU 4台, コア4MB, OSはOS7<sup>リ</sup>)<sup>2</sup>、後述のMONITORコマンドの実現には、先に開発したソフトウェア・モニタ<sup>5)</sup>を拡張して使用した。

開発した主なコマンドは次の通りである。

// MONITOR ソフトウェアによる測定結果(主にスナップ・ショットと累積値)を表示させる。

// HOWBUSY システムの混雑状態を表示させる。

// WHO 現在、TSSを使用しているのは誰かを知る。

// RWAITJOB 各リモート・ステーションから投入されたジョブの状態を知る。

なお、開発中のものとして次のコマンドがある。

// QUEUE 入カスタックなどにあるクラス毎のジョブ数などを表示させる。

将来のコンピュータ・システムでは、こうしたコマンドは標準的なユーザ・コマンドとして備えられるべきだと思われる。

## §2. HOWBUSY コマンドとWHO コマンド

//HOWBUSY (計算機の混み具合、ターンアラウンドタイムを見るコマンド)

### NUMBER OF ACTIVE JOBS/TERMINALS ###

```
BATCH      12
TSS        13
R-BATCH    06  AT 11:29:18
### JOB    INPUT TIME (WAITING TIME) ###
A  08-24  11:21  (00:08)
B  08-24  11:23  (00:06)
C  08-24  11:09  (00:20)
### END OF HOWBUSY ###
```

TSS 端末約100台 うち10~25台同時使用  
リモートステーション約20台 うち10台常時オープン

現在実行しているジョブは何日の何時に入力したものであるか、どの位待ちは実行開始かをジョブクラスごとに表示している。  
各ジョブクラス内では先着順に実行する。

次に示すWHO はアメリカ系の OS には標準的に備っているが、われわれの OS にはなかったのが新規に作成した。TID は端末(回線)番号、J.S.N. はジョブ通し番号、MPG は仮想メモリ使用量である。

//WHO (誰が使用しているかを見るコマンド)

```
TSS      9 USERS
BATCH    12 JOBS
R-BATCH  6 STATIONS
```

NO	USER-ID	TID	JOB NAME	J.S.N.	MPG(KW)
1	6041405042	04	KAWABATA	000040	60
2	4743303001	251	C35	000098	60
3	3366193059	200	E310	000103	60
4	6305413001	05	0	000117	100
5	0085960000	253	E823	000127	60
6	5400403061	29	D29	000206	60
7	5799403001	28	D28	000220	60
8	5825403002	03	R3	000278	60
9	0078900000	24	D24	000280	60

RWAITJOB コマンドは当該端末より入力したリモートバッチジョブの実行待ち状態を表示する。IJQ は入力ジョブキューである。

1) ..RWAITJOB

```
####JOB COUNT OF IJQ ####
A=0010, B=0002, C=0001
###JOB COUNT OF WAITING ###
A=0000, B=0001, C=0002
###END OF RWJ ###
```

2) ..RWAITJOB ,B

```
####JOB COUNT OF IJQ ####
B=0001
###JOB COUNT OF WAITING ###
### NONE ###
###END OF RWJ ###
```

### §3. MONITOR コマンド

MONITOR コマンドは、ソフトウェアモニターで常時採取できる情報のうち一般ユーザにも興味のあるものを表示するためのもので、表示される情報は次のようなものにしてある。

- 1) 4台のCPUのアイドル状態。
- 2) 主記憶領域の使用状態。
- 3) 状態別タスク数(ランニング、レディ、ウエイト、ページング、ブロック、ペンディングの各状態について)
- 4) リエントラントプログラム(コンパイラ、実行時ルーチン、リンケージエディタ等)の共用数。
- 5) ページフォールトとスワッピングの状態。
- 6) ドラム(スワップイン、スワップアウト用)の使用状態。
- 7) ドラム(スワップイン、スワップアウト用)にREAD/WRITEしたページ数(チェンジビットの効果)
- 8) ドラムとディスクの使用状態(I/O待ち行列の数)

また、その他にOS7に特別なパッチを入れたときにのみ表示できる特殊な情報として次のものがある。現在これはユーザに開放していないが将来、ユーザにもみられるようにするつもりである。

(1) 各CPUの使用状況

アイドル時間、ユーザタスクのCPU時間、システムタスクのCPU時間、アクションのCPU時間。

(2) アイドルの要因分析(WAIT, PAGING)

(3) アクティブタスク数

(4) 強制インアクティベート回数

次にこれらの情報の意味を少し詳しく述べる。

(1) 4台のCPUごとのアイドル・タイム。それぞれCPUはアイドル状態の時、アイドルタスクをランさせている。そこでそのアイドルタスクをランさせている時間の累積値をアイドル・タイムとしている。

(2) 主記憶、仮想記憶の使用状態。主記憶1024ページ(4MB)のうちシステム常駐部分、システム非常駐部分、ユーザ領域、未使用領域が占める割合、またシステム領域中でどういう部分が主記憶に存在するか。

(3) ドラムの使用状態。ドラム(各1040ページ≒4MB)6台のうち、実際に使用している部分はどの位あるか。

(4) ページングの回数。システム領域およびユーザ領域について、それぞれ次の3つの値を表示する。

(a) 領域を参照してページ・フォールトを発生した回数。

(b) 領域を参照してページ・フォールトを発生し、ドラムをREADした回数。

(c) 領域を参照してページ・フォールトを発生し、当ページに割当てするための主記憶が不足したり、また、当ページはページインかページアウトのI/O動作中で完了を待ったためにページング状態となった回数。

(5) リエントラント・プログラムの共用数。コンパイラ、実行時ルーチン等はほとんどバリエントラントな構造をしており、現時刻にどのようなモジュールがタスク間で共用され同時にランしているかを表示する。

(6) ディスク(入カスタック、出カスタック、WORK用のファイル、システム・ファイル)、ドラム、ラインプリンタ、カード・リーダ、磁気テープ、紙テープ等100台以上の装置のそれぞれについてI/O発行回数と装置使用の待ち行列数を表示する。

(7) スワッピング用ドラムに対するREAD/WRITEのページ数を表示する。主記憶にとり込まれ、次にそれをスワップ・アウトする場合、もしこのページに書き換えが生じておらず(この場合、当ページのC(Change)ビットは0のままになる)、ドラムに全く同じものが存在する場合は実際にはドラムにWRITEされない。このような現象がどの位起きているか、すなわちCビットの効果はどの位あるかが分る。

(8) タスクスイッチの累積値。タスクスイッチとはシステムタスク、ユーザタスク、アイドルタスク間の変化のことである。シングルタスクでもタスクT<sub>1</sub> → アイドルタスク → タスクT<sub>2</sub>と変化するとカウンタには2が加えられることになる。

(9) 状態別タスク数。タスクにはランニング、レディ、ウェイト、ページング、ブロック、ペンディングという状態があるが、これらを分析し、タスクの状態を表示する。

(10) TSSレスポンスの分布。入カ、レスポンス、出力、連続出力、思考、連続入カのCPUタイムとETIMEをヒストグラムとして表示する。

(11) アクションのCPU時間とシステムタスクのCPU時間およびユーザタスクのCPU時間を求めCPUバドのように使用されているか効率ほどの位であるかを分析する。アクションのCPU時間とは主にI/Oの割込みなどの外部割込処理に要するCPU時間である。

(12) アイドルになった主な要因がI/Oネックのためか、メモリネックのためかを分析し、より効率のよいシステムにするにはどうすればよいかを考える手掛りとする。

(13) アクティブのタスク数、アクティブのバッチタスク数を求め表示する。

(14) 強制的にインアクティブにした回数を表示する。強制インアクティブとは過度の主記憶装置の競合(スラッシング)およびページングI/Oをなくし、CPUと主記憶を有効に利用するため、いずれかのタスクを強制的にインアクティブ状態に落すことである。

これらの情報はMONITORコマンドによるほか、特定ユーザによって、グラフィックディスプレイ(H8811)にグラフィック形で表示することもできる。またこれらのデータの1日分の変化の様子は、リモートステーション(H10ミニコン、ネットワーク)についているテクトロニクスストレージスコープにも表示できるようになっている。

#### §4. MONITORコマンドによる測定と評価の例

次にMONITORコマンドによる測定値とその評価(解釈)の例をあげる。ここに示すデータの大部分はある測定時の値であり、定常値というわけではなく、システムの状態によって時々刻々変わるものであるが、以下にあげるデータは典型的な測定値の例である。

(1) 本システムではH-8800 2台は主に演算用で使用し、H-8700 2台は主に入出力を含むOSの動作用で使用しているが、アイドルは2台のH-8800の方が常に少なくなるよう制御しているため、5%~15%位であり、2台のH-8700の方は12%~35%のアイドル率である。

各CPUのアクションの時間、システムタスクのCPU時間、ユーザタスクのCPU時間、アイドルの時間の3日間の平均のパーセンテージを表1に示す。

前述したとおり、8800はユーザタスクの演算に、8700はシステムタスクとI/Oの動作に使用するよう割り当てているが、表1

```

***** CPU IDLE TIME *****
                                76/03/11    01:25:46
                                (IPL 76/03/10  08:54:32)
CPU0(8800)                      165 MIN      16.74 %
CPU1(8800)                      287 MIN      27.92 %
CPU2(8700)                      363 MIN      36.66 %
CPU3(8700)                      337 MIN      34.15 %
ACTIVE CPU--- 0 1 2 3
  
```

図1 CPUのアイドル時間

	ACTION CPU TIME	SYSTEM TASK CPU TIME	USER TASK CPU TIME	IDLE CPU TIME
8800 (CPU0)	1.40%	0.71%	92.79%	5.10%
8800 (CPU1)	2.42	0.94	88.49	7.81
8700 (CPU2)	25.33	5.76	40.65	28.25
8700 (CPU3)	27.91	12.70	32.69	26.75

表1 各CPUの使用状況

から、この割り当てがうまく行なわれていることがわかる。アクションのCPU時間とは主にI/Oの割込みなどの外部割込みの処理をする時間である。8800ではユーザタスクが88%~93%のCPU時間を使用しており、8700でも33%~41%を使用している。一方、システムタスクのCPU時間は8700でも5%~13%位であり、システムタスクが過度にCPUを使用していないことが分かる。

(2) ユーザジョブの少ない4月、5月はOSの領域が広がり、4MB(1024KW)の主記憶

のうち、OSの占める領域(SYSRES, SYSLIB, PUBLIB, テーブル類)が時として50%(2MB)を越えることがあるが、2月頃の混んだ状態ではOSは35%~43%位であった。リモート端末やTSS利用が増えるとともにOS領域は増える方向にある。もちろんシステム制御としてはOS領域を故意に小さく制御することもできるが、それは効率とのかねあいになる。いずれにせよOS領域とユーザ領域の区別はたんだん意味を失いつつある。

(3) タスクの状態は図3の通りである。

WAITINGは主にディスクI/O待ち、BLOCKEDは主にTSS I/O待ちの状態である。RUNNINGは常に3~4タスク存在しCPUがほぼいつも使用されていることがわかる。全体としてはBLOCKEDの状態にあるタスクが30~40%あり、TSSタスクの80%位がBLOCKEDの状態にある。

(4) リエントラント・プログラムの共用数について調べてみると、モジュールを使用しているタスク数はユーザタスクの多重度がバッチ8、TSS 20、計28の時次のようになっている。(図5)

#FCH00 (FORTRAN コンパイラ) が2~5, #ACL00 (FORTRAN 実行時ルーチンおよび基本外部関数等) が4~15, #FCS00 (FORTRAN サービスルーチン, デバッグルーチン) が1~5でFORTRAN 関係の3つのルーチンは、主記憶からスワップアウトされることなく、ほとんど常に主記憶上に存在していることがわかる。また、#GDA01 (リンケージエディタ) は0~2, #GGZ00 (シンボリックライブラリ保守) 0~2である。

(5) OSのページングの設計方針はFINUFO (First In Not Used First Out) というアルゴリズムで行なわれており、ページの状態変化は図6の通り

```

***** CORE USAGE *****
                                76/08/24  11:02:05
SYSTEM AREA
SYSTEM RESIDENT AREA          9.28 %
SYSTEM NON-RESIDENT AREA      30.47 %
USER AREA                      59.47 %
UNUSED AREA                    0.78 %
MEMORY AREA---1024 KW
  
```

図2 コアの使用状況

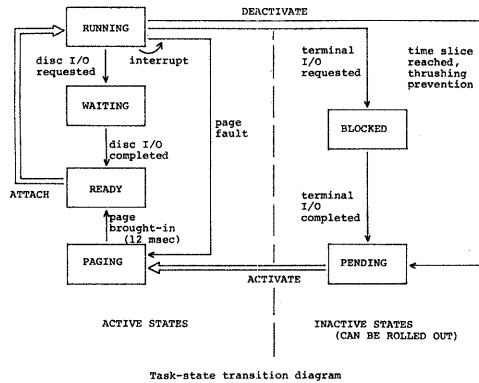


図3 タスクの状態変化図

```

***** STATUS OF TASKS *****
                                76/08/25  11:33:12

```

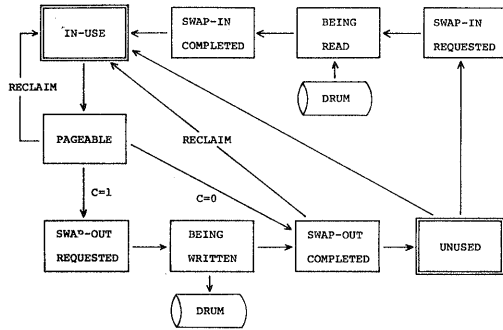
	TOTAL	SYSTEM	R.T.	TSS	BATCH	OTHER
RUNNING	4	0	0	1	3	0
READY	2	0	0	0	2	0
PAGING	0	0	0	0	0	0
WAITING (DISC)	0	0	0	0	0	0
BLOCKED (TSS I/O)	9	0	0	9	0	0
PENDING (MEMORY-WAIT)	0	0	0	0	0	0
-----						
TOTAL	15	0	0	10	5	0

図4 タスクの状態 (R.T.はリアルタイムタスク)

```

***** SHARING OF REENTRANT PROGRAMS *****
                                76/08/24  11:03:03
#FCL00  FORTRAN RUN-TIME ROUTINE      12
#FCS00  FORTRAN SERVICE ROUTINE       5
#FCH00  FORTRAN COMPILER               3
#GDA01  LINKAGE EDITOR                 0
#GGZ00  SYMBOLIC LIBRARY MAINT.       2
  
```

図5 リエントラント・プログラムの共用数  
右の数字はリエントラント・プログラムを現在使用しているタスク数を示す。0は使用していない事を示し、1は1つのタスクで使用していることを示す。



Status movement of each main-memory page

図6 ページの状態変化図

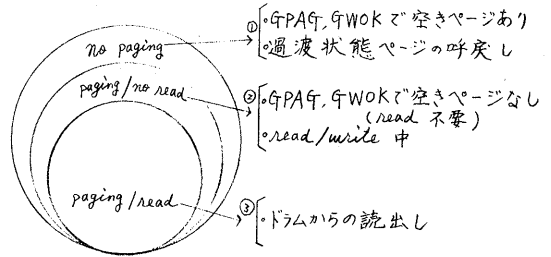


図7 ページングとスワップイン

である。ページング状態というのは、ドラムからの読出しを待っている状態、GPAQ(get-page)あるいはGWOK(get-work-area) マクロ命令がでて(作業領域を最初に割当てるとき)割当てべきコア領域がなく、どれかのページを追い出している状態、およびいまドラムへのread/writeを行っている最中の状態のいずれかを表わす。ページフォルトが起きた時のページの状態を図7に示す。ページフォルトが起きると次のような3つの状態となる。①ページフォルトが起きてもページング状態にならない。②ページフォルトが起きてページング状態になったがドラムからのreadは行なわれない。③ページフォルトが起きてページング状態となり、ドラムからのreadを行う。この3つの割合を示すと表2のようである。ページの呼出し(reclamation)ができるのは、ページがドラムへ追い出される直前の直後で、特にinput readerやoutput writerタスクなどで、他のタスクが同じページでページフォルトを起しているときなどに多い。ページの状態の比率(表2)を見るとシステム領域ではページング状態になりドラムから読み出す割合いが90.5%もある。ユーザ領域はページング状態になりドラムから読み出す割合いが51.0%で、空きページがあったり、ページの呼出しでページング状態にならなかつた割合いが47.2%もある。これはシステム領域をあまり増やさないようにコントロールしているためである。主記憶にとり込まれ、次にスワップアウトされる時、Cビット(change bit)を調べ、そのページが書き換えられていたら、スワップアウトされ、再びドラムに書き出されるが、Cビットが0の時は、そのページは書き換えられていないということでドラムに書き出すことはしない。この効果が高いことが表3よりわかる。

	システム領域	ユーザ領域
NO. PAGING	3.9%	47.2%
PAGING / NO READ	5.5	1.8
PAGING / READ	90.6	51.0

表2 ページの状態の比率

	回数	総ページ数	1回に処理 するページ数
READ	929799	945582	1.02
WRITE	211687 (22.8%)	239888 (25.4%)	1.10

表3 READ/WRITEの回数とページ数  
%はREADに対するパーセンテージ

\*\*\*\*\* PAGING INFORMATION \*\*\*\*\*

76/08/24 11:06:44  
 PAGE FAULT OF SYSTEM AREA 82326  
 PAGE FAULT S.A., READ DRUM 76707  
 PAGE FAULT S.A., STATE OF PAGING 78374  
 PAGE FAULT OF USER AREA 70134  
 PAGE FAULT U.A., READ DRUM 42350  
 PAGE FAULT U.A., STATE OF PAGING 43542

図8 ページングの情報

\*\*\*\*\* NUMBER OF DRUM I/O PAGES \*\*\*\*\*

76/08/24 11:04:42  
 NUMBER OF WRITE PAGES 31478  
 NUMBER OF READ PAGES 113413  
 NUMBER OF I/O OPERATIONS 142043

図9 ドラム READ/WRITEしたページ数

(6) I/O発行回数については、少しの差はあるがカードリーダー、ラインプリンタはほぼ平等に使用されていることがわかる。ディスクのI/O発行回数は1回のI/Oで4回とカウントされている。①WRITE DATA ②WRITE EOF ③READ LABEL ④WRITE LABELの4回である。又、アカウンタファイル(K02)は1ジョブについて8回アクセスしている。(図10)

(7) I/O装置への待ち行列はほぼ0か1でラインプリンタ、カードリーダー、ドラム、ディスクの一部に並ぶ程度であり、OSによる装置割当てがうまくいっていることがわかる。ワーク用のファイルはアクセスの回数が多く、パックを増した方がよいと思われる。この点はパックの大容量化(100MBから200MBへ)に伴い深刻になるので、OSで目下検討中である。

(8) 仮想記憶として使用しているドラムの実際の使用率は15~35%であり、6台のドラムがほぼ均等に使用され、スケジューリングがうまく行なわれていることがわかる。(図11)

しかし、実際には使用していないが60~70%近くの領域の確保は行なわれており、来年2月の混む時期にメモリページ(MPG)オーバーフローが出ないか心配されるので、よりよい割当て法を検討中である。

(9) アイドルの要因としてI/Oネックとメモリネックが考えられる。数値の上から見ると図12のようにI/O待ちでCPUが不必要となった場合の方が2倍位になっている。I/O待ちの方には本来にJOBがなくアイドルした場合も含まれているので数値の上で実際より多くなっていると考えられる。メモリ待ち(PAGING)については、現在主記憶4MBで未使用領域は0.5%~6%と少なく、CPUのアイドルは8800で6.6%、8700で27%あることから現状では主記憶をもう少し増し、タスクの多重度を増すとCPUのアイドルは当然もっと減ると考えられ、次の増設で主記憶をさらに4MB増し8MBにすること

\*\*\*\*\* I/O COUNTS AND QUEUE \*\*\*\*\*

	COUNTS	QUEUE
D00	55264	0
D01	58000	0
D02	57010	1
D03	58693	1
D04	59438	1
D05	54773	0
K02	67574	0
K40	17835	0
K41	19780	0
K42	19566	0
K46	250	0
K47	16	0
K50	55921	1
K51	65043	0
K52	9	0
K53	2643	0
K54	18341	0

図10 I/O発行回数と待行列

K54 ---- IJQ (入カジョブファイル), K53 ---- ライブラリー  
K40, K41, K42 ---- OJQ (出カジョブファイル)  
K50, K51 ---- W+7

\*\*\*\*\* DRUM USAGE \*\*\*\*\*

	76/08/24	11:03:39
DRUM 0	33.08 %	
DRUM 1	32.21 %	
DRUM 2	31.15 %	
DRUM 3	30.87 %	
DRUM 4	30.29 %	
DRUM 5	24.23 %	

図11 ドラムの使用状態

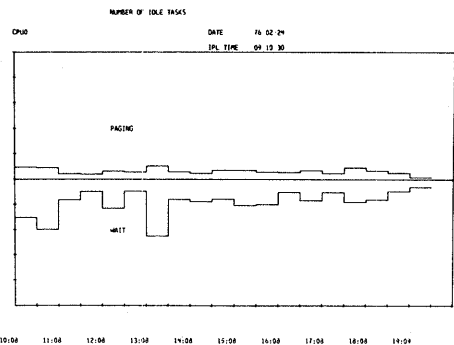


図12 アイドルの要因分析(旧)  
横が時間、縦が回数である。

にしている。

(10) 主記憶装置の競合, およびページングI/Oをなくし, CPUと主記憶を有効に利用するためにOS7では強制的にあるタスクをインアクティブにすることがある, その結果を図13に示す。1分間に20回位, インアクティブ状態に強制的に移行されており, この結果からも主記憶はより大きい方が当然よりよいCPU利用が出来ると思われる。

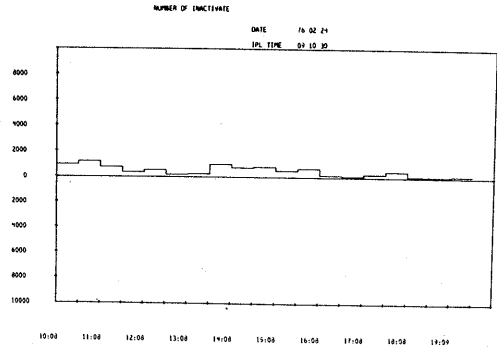


図13 強制インアクティブの回数(1日) 30分ごとの回数を表示している

(11) ブロックマルチプレクサチャンネルのビジー率は平均5%~8%である。2つにわけると

H09 ---- 4台のディスクが繋がれている(ビジー率5.1%)

H39 ---- 8台のディスク " (ビジー率8.5%)

であり, 8台までは10%以下のビジー率で処理出来ていることを示している。ドラムに接続されているチャンネルのビジー率はドラム4台の時, 平均16%ぐらい混んでおり, チャンネルを増す必要が示されている。

(12) TSSレスポンスタイムは次の6つについてのCPU時間と経過時間をヒストグラムとレタ測定している。図示すると図14の通りである。

- ① 入力始めから入力終わりまで (INPUT)
- ② 入力終わりから出力始めまで (RESPONSE)
- ③ 出力始めから出力終わりまで (OUTPUT)
- ④ 出力終わりから出力開始まで (OUTPUT INTERVAL)
- ⑤ 出力終わりから入力開始まで (THINK)
- ⑥ 入力終わりから入力開始まで (INPUT INTERVAL)

入力を始めてから入力を終わるまではETIME 10秒以内で終わっている人が一番多いが, 30秒以上かかっている人もかなりいる(図15)。入力終わりから出力始めまでのいわゆるレスポンス時間は計算機が長い計算をしていることも多いので, CTIME, ETIME とともに図16に見られるようにばらつきがあり, CTIME 28ミリ秒, ETIME 28秒以上の人もかなり多い。出力始めから出力終わりまではETIME 5秒以内でほとんど終わっている。

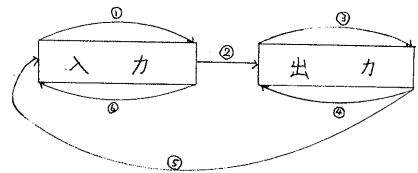


図14 TSSレスポンスタイムの測定時間

出力終わりから出力開始までは, この間に計算機が演算を行なっている場合もあり, CTIME, ETIME とともにばらつきがあるが, ほとんどの場合, ETIME 4秒, CTIME 4ミリ秒以内であった。出力終わりから入力開始までは計算機がプロンプティングを出すまでの時間とユーザが考えている時間が含まれている。また始めに述べた入力始めから入力終わりまでにも, 入力しながら考えるというTHINKの時間が含まれていると考えられるが, ユーザのタイピングの遅さがかなり反映されていると思われる。





## 参考文献

- 1) 石田: 東大超大型コンピュータシステム, 情報処理学会誌, Vol.15, No.7, pp.534-541(1974).
- 2) J.H.Saltzer: The instrumentation of Multics. Comm, ACM, Vol. 13, No.8, pp.495-500(1970).
- 3) J.M.Grochow: Real-time graphic display of time-sharing system operating characteristics, Proc. of AFIPS FJCC, Vol.35, pp.374-386(1969).
- 4) T.Masuda: Optimization of program organization by cluster analysis, Proc. of IFIP Congress 74, Vol.2, pp.261-265(1974).
- 5) Ishida, Nomoto, Ozawa: Graphic Monitoring of the Performance of a Large 4-CPU Multiprocessor System, 2nd UJCC Proceedings pp. 271-275(1975).
- 6) 石田, 野木: 超大型機 のふるまいのグラフィック・モニタリング  
情報処理学会システム性能評価研究会資料 75-9 (1975.3.14)