

# сообщения <br> обьвдинвиного <br> ииститута <br> пдерных <br> исследований <br> дубна 

## $6509 / 83$

E11-83-522

Z.K.Loboz, T.I.Zaboy

MEASUREMENTS AND TUNING
OF OPERATING SYSTEM
PERFORMANCE NOS/BE ON CDC-6500
AT JINR, DUBNA

1. TOOLS USED TO MEASURE AND ANALYSE SYSTEM PERFORM $\operatorname{NCE}$

System work was analyzed by program package systim performace analyzer (further referred as SPA) designed and wri ten by Zdzislaw Loboz. This paragraph describes main charasteristics of system architecture and program package used to measure system work.

### 1.1. System Architecture

Detailed system architecture is described in app sopriate manuals. The parts more important from the point of view of our measurements are:

1. Two central processors (CPU*S). Because of thゝir identical speed and management by system monitor thay are considered as one processor.
2. Central memory (CM). Size 128 K words.
3. Ten peripheral processors (PPU*S).
4. Six 100 MB disks. They are connected to the system through one disk channel. Operating system resides on one of them. Three are for permanent files and Input/Output queues. Two are for user's private packs. Local (scratch) files may reside only on system disk or on permanent file disks.
5. Six 30 MB disks. Used exlusively for users private packs. 1.2. SPA

Package consists of two main programs - one for gathering data about system work and one for analysis of this data. There are also other programs to assure proper work of main programs between deadstarts and ease analysis of received data.

### 1.2.1. Data Gathering - SPA (PP)

Data is gathered by a bouncing peripheral proces sor program SPA(PP). It scans central memory in specified time interval and writes information obtained on each scan onto system file. Information contains actual values of about 400 system rariables like, for example, IDLE CPU time, number of jobs ir central memory, channel statuses and so on.
1.2.2. Data Analysis - SPACP

File produced by $\operatorname{SPA}(P P)$ is analysed by fortran program. On input to it we specify time period which we want $t$, analyze, variables to be analyzed over specified time perioi and type
of analysis. For each value we can obtain following characteristics: average, minimum, maximum, standard deviation, histogram distribution, correlation with other values. In addition we have a possibility of combining several values written by SPA (PP) into one value analysed by SPACP. Due to package modularity we can easily analyze behaviour of system variables using other methods. (See App.B, App.C).

## 2. METHODS OF MEASURING AND ANALYSIS

Interval of central memory scan by SPA(PP) was 60 seconds 1440 times per day. It proved to be enough for statistical analysis of received data and system file produced by SPA(PP) was of reasonable size - about 280 PRU per day. After installation of SPA in the system the received data were analysed daily to find main system characteristics and estimate which system variables are most important - we can analyse about 400 of them, but only a small subset is critical to system performance.

These analyses have shown, that system bottlenecks are delays caused by jobs waiting for disks input/output operations and peripheral processors. Additional counters were installed in the system to obtain more detailed information about these parts of the system.

Initial measurements have shown that the system work may be divided into two parts - peak hours and normal hours - and the
 tely: from 13 till 22 hours and from 22 till 8 hours. The remaining period - from 8 till 13 hours is an engineer's time.

## 3. RESULTS

This part describes values received for main system characteristics. Only typical values are given - system was measured for several months and printing all received values here would result in report of unreasonable size.

The word day denotes time interval from 13 to 22 hours - i.e., time which may be described as peak hours - due to the number of logged in Intercom users, jobs processed and so on. The word night denotes time interval from 22 to 8 hours in the morning. The average values below are for April-May 1983. These months were chosen for analysis, because at this time all additional counters were installed in system - and we could analyse system work in detail.

### 3.1. IDLE CPU Time

For measured period of time in analyzed time intervals (days and nights) IDLE CPU time varied from 0.1 to 0.5 . Average value 2
for working days - Monday co Friday - was 0.34. During night hours IDLE CPU time was on average 0.23. On week-ends during the day IDLE CPU time is between 0.20 and 0.25 , during the night it is sometimes less than 0.10. The average LDLE CPU time varies significantly from month to month, for example, in November 1982 it was about 0.40 . Variations in IDLE CPU time measured by SPA are similar to those measured by dayfile analyser. The differences between months are about $10 \%$. This differences are mainly due to job mix prevailing on a given month - CPU bound jobs, short I/O bound jobs and so on.

### 3.2. Peripheral Processors

IDLE PPU time - for all PPU'S - ranges from 2.5 to 4.5 seconds for one physical second. It is smaller during the day 2.5 to 3.0 - when a lot of $1 / 0$ operations are performed. Another variable which also describes the use of peripheral processors is average number of free PP. For working days it ranges from 1.2 to 1.9 during day hours and from 2.3 to 3.1 during the night. From analysis of this variable we can see that maximal number of unsatisfied PP calls (in other words queve to PP) is 10 on average day. In all, during the day, queue to PP exists for 6 to 16 percent of day time. The same percentage for nights ranges from 0 to $2 \%$.

## 3:3. 1)isk Calls

In NOS/BE a program to perform disk Input/Output operation calls the system. If disk channel is busy servicing other call the next call must wait. The number of disk calls per second ranges from 18 to 21 during the day and from 16 to 19 during the night - for all 100 MB disks. For all 30 MB disks number of calls per second varies from 0.5 to 1.5 - small disks are used for private packs only.

System variable, which proved to be important to system performance is a length of disk queue. Disk calls which cannot be satisfied immediately are waiting in disk queue - and programs issuing them cannot use CPU. Average disk queue length varies from 2.5 to 2.9 during the day and from 1.6 to 1.8 during the night.

Although there are 12 disks in system disk queue to each disk is not the same. There is no queue to 30 MB disks. There is practically no queue to two 100 MB used for private packs. For average queue length during the day -2.7 in all - 1.5 comes from system disk, 0.5 from permanent file master disk, 0.3 from each of two remaining permanent files disks.

### 3.4. PP Program Ca1ls

Each programs in NOS/BE to perform Input/Output operation on any device must call PP program. Number of calls per second for PP programs varies from 35 to 55.

There are about 500 PF programs in the system and they usually reside on disk; most frequently used ones reside in central memory. System, however, gives us no tools to determine exactly which programs are frequently called - and to determine how many calls per second are made for disk resident PP programs. After installation of appropriate counters it was found that number of calls per second during the day varies from 5.0 to 6.5 - i.e., about $30 \%$ of all disk calls are calls for disk resident PP programs. Calls for disk resident PP programs comprised on average 10 to $12 \%$ of all calls for PP programs. (See App.A).

## 4. IDLE CPU TIME - REASONS

Analysis of correlations between system variables has shown, that IDLE CPU time is mainly due to disk queue. Disk queue is also responsible for increasing PPU activity - because PP programs are waiting for information from disk. Analysis of plot "average disk queue length "versus" average IDLE CPU time" for 44 experimental points has shown, that we may describe IDLE CPU time as a linear function of disk queue length:

LDLE CPU $=0.07+0.11 *$ disk queue.
The correlation coefficient between the above function and results of experiments is 0.90 .

In actual system version the dependence between disk queue length and number of disk calls per second (for average valucs) is

## disk queue $=0.28 * \exp (0.12 *$ disk calls) .

Correlation coefficient between the function and results of experiment is 0.97 .

Dependencies described above are statistical. For disk queue length 2.5 measured IDLE CPU time varied between 0.25 and 0.35 because IDLE CPU time is produced by many factors - from untypical job mix to wrong operator intervention. What's more we have to deal with working system? At any time a new job may be read into the system or end because of error, at any time a user may LOG any terminal IN or OUT. Therefore such variations should be expected - but big correlation coefficients show, that there are certain statistical regularities in system work.

Although from the above functional. dependencies we can compute, that for 19 disk calls $p \in e^{\text {second }}$ tDLE CPU ime stould
be 0.35 rhis is only an average (expected) value of IDLE CPU time.

From this functions we can see, for example, that by reducing number of disk calls per second from 19 to 15 we can reduce IDIE CPU time from 0.35 to 0.25 - i.e., by $10 \%$.

## 5. EXPERIMENTS

After finding regularities and main bottlenecks in systera work some experiments were made to assess possibilities of improving operating system efficiency. The objective was to curtail LDLE CPU tine without worsening any other system characteristic. 1DLE CPU time, as mentioned earlier, depends on many factors - so to estimate improvements in average IDLE CPU time after changing - for example number of disk calls we would have need at least several days of measurements. There was not enough time available for such measurements, therefore we were forced to use functional dependencies described earlier and look rather on disk queue than on IDLE CPU time - mainly because disk queue length and distribution is much more stable than IDLE CPU time.

### 5.1. Calls for Disk-Resident PP Programs

The total size of Central Memory Resident PP programs was 34000 B words. The number of calls per second for Disk-Resident re programs was on avelage j.j. Aitiougit hiio varıion $30 \%$ wín all disk calls - it was about $12 \%$ of all calls for PP prograns. The CM-Resident subset of PP programs was determined by CDC recomendations and otservation of system work.

Changing the system to count the number of times each individual PP program is called enabled us to improve this situation and find optimal subset of CM Resident PP programs for each given CM Resident size- and predict a number of disk calls $\overline{\text { for }}$ each subset.

Two experimental subsets of CM Resident PP programs were tested. The first one to minimize number of calls for Disk Resident PP programs with maximum size of CM Resident 35000 B words. The second one to minimize size of CM Resident with the same number of disk calls per second as for "standard" resident-5.5. The results were as follows

| SIZE | DISK PP <br> CALLS | DISK <br> QUEUE | SYSTEM <br> DISK | PERM. | A | BILE |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | DISKS

For Resident minimizing the number of disk accesses the reduction in number of disk calls was 4 per second. The total queue in result was shortened by O.8. As PP programs reside on system disk the queue length for permanent file disks was not affected but the queue length to system disk was halved. Additional consequence of using such a resident was decrease in PPU activity by about $10 \%$ to $20 \%$. This was due to the reduction of "busy waiting" state of PPU - when PP program waits for information from disk occupying PP - but doing nothing in fact.

From the results of experiment with a resident minimizing central memory occupied by PP programs we can see, that keeping the number of disk calls per second (and disk queue) on the same level as before the experiment we can obtain 14000 B words of central memory more. The number of disk calls during the nigth is smaller - so the gain during the night hours may be even 20000B words - and outside Intercom hours even 24000 B words.

### 5.2. Changing Distribution of Disk Queues

From the previous experiment we have seen, that it is possible to reduce length of disk queue by reducing a number of disk calls. There are, however, other possibilities - changing distribution of disk calls between disks. The total number of disk calls per second is about 19 - but the queue lengths for each individual disk differ greatly. This suggest that the numhar of ralla for parh disk is not identiral. The svstem does not count disk calls for each disk - so we must use a simple model to estimate what gains can be expected from changing distribution of disk calls.

### 5.2.1. Model of Disk Calls Distribution

According to queue theory the average queue length for a given system is given by

## QUEUE $=$ RHO **2 $/(1-$ RHO).

Where $R H O=R Q S / S P E E D)$ RQS is a number of requests per time unit, speed is disk speed measured by a number of requests that may be satisfied in time unit.

This function is highly unlinear and for systems with RHO greater than 0.5 we can expect a significant reduction in queue length for small changes in RHO. It means, that for sustems working at their limits even small reductions in number of request per second should result in significant reductions of queue length. The above is, of course, true also for reduction of total number of disk calls described in experiment with PP
programs. Let's see how it should work for different distribitions of calls for individual disks.

Suppose, that total number of disk calls per second is 19 and speed of each disk is 14. With distribution of calls between four disks like $(5,5,5,4)$, the total queue length will be 0.71 ( $0.2,0.2,0.2,0.1$ ). If we change the distribution to ( 10,3 , 3,3 ) the total queue length will rise to 1.96 ( $1.8,0.06,0.06$, $0.06)$. Now, suppose that the system will try to change a placement of only one call - from disk 4 to disk 1 - without changing total number of calls. For distribution (11,3,3,2) we have total queue length $3.06(2.9,0.06,0.06,0.02)$ - reassigning one disk request has changed queue length by $30 \%$.

The above is only a simplified model. In our system disk speed is not constant - it changes with number of requests per time unit. Nevertheless it is obvious, that changing distribution of disk calls for individual disks we can change length of disk queue.

### 5.2.2. Second Disk Channel

In this installation there are two kinds of disks $1001 B$ and 30 MB . Total number of request per second for 30 MB disks is on the average 0.5 to 1.5 . This is very small in comparison with 100MB disks with their 19 calls per second - but small disks are used for user private disks, while on 100 MB disks we have


To reduce number of calls for 100 MB disk and distribute disk calls between greater number of disks one 30 MB disk was added to the system and defined as scratch. The total number of calls remained on the same level - 19 , but not 5 of them were for additional disk. The distribution of queue lengths to 100 MB disks was changed from previous ( $1.5,0.5,0.3,0.3$ ) to ( 0.8 , $0.2,0.1,0.0$ ) - total queue to 100 MB disks was reduces from 2.7 to 1.1. But the request for 30 MB disk created - non-existing before - queue to small disk with length 1.9. So, the total queue to all disks was 3.0. Small disks are slower and cannot handle such a number of disk calls per second as l.OOMB disks.

In the next stage of the experiment two small disks were added to the system for scratch files. Total number of disks requests remained on the same level,but now 6 of them were for two additional 30M3 disks. Distribution of queues for 100 mR disks was ( $0.7,0.2,0.1,0.1$ ) - total queve to 100 mb disks was 1.1. Queue length for two additional disks was 1.2,- so total queue length for all disks was 2.3.

The above result: show, that the muber oi requests transferred to each added $301 B$ disk shouldrí be greater than 2 per second - to avoid mume to small disks. To produce such a num-
ber of calls for each small disk added to system we should change relative weights of $100 M B$ and $30 M B$ channels.

Taking into account difference between number of disk calls per second and IDLE CPU time we can expect decrease in IDLE CPU time by $5 \%$ for one small disk added to the system and by $10 \%$ for two small disks added to the system - if the number of requests to these disks will be on the level 2 per second for one disk.

### 5.2.3. Uniform Distribution of Calls for 4 loomb Disks

In the present system configuration we have 6 l00MB disks, but two of them are used exclusively for user's private packs. From the remaining 4 - one contains system files; the other 3, permanent files. Local (scratch) files may reside on any of these four disks. There are big differences in queue lengths to these disks - the distribution is (1.7, 0.5, 0.3, 0.3). The longest queue is to system disk - but this is due not only to the calls for disk resident PP programs or other system programs. The problem is with local files. When the user program (or Intercom) needs a local file, system finds a disk with a biggest amount of free space. And there is more free space on system disk than on any other permanent file disk - most local files are assigned to system disk. So, from 19 calls per second to all l00MB disks there are at least 10 to system disk - 5 for disk resident PP program and about 5 for local files.

Tine simpiest melinu u culeck if il is reaily su is lu calaiug a big dunmy file on system disk and, by this, force the system to use other disks for local files more frequently. After such an experiment distribution of queue lengths changed from (1.7, $0.5,0.3,0.3$ ) to ( $1.1,0.5,0.4,0.2$ ) - and total queue length was reduced from 2.7 to 2.2 .

This result means, that by manipulations with an assignment the local files and number of calls for disk resident PP programs we can significantly reduce queue length.

## 6. CONCLUSIONS

The results of measurements and data received after experiments show, that it is possible to increase CPU efficiency by up to $15 \%$. It is also possible to increase available memory by up to 20000B words.

Increase in CPU utilization may be obtained by reducing the number of disk calls and/or even distribution of calls between available disks for scratch or frequently used system files. Additional memory may be obtained by optimizing contents of central memory resident.

The above may be obtained in a number of ways - the most efficient ones require certain changes in the system. Some of them do not - but they may be inconvenient to the computer center because of user demands - like using three $30 M B$ disks for scratch files.

To decide which way will be chosen is not an analyst work. In our opinion a system analyst should only provide a list of possibilities and expected results. Choosing between such a possibilities involves also a policy of computer center towards its users and should be solved on managerial level.

As a typical exatuple of such decision lets consider to CDC6500 connection of additional terminals. If new terminals will be used as frequently as the old ones, we can expect increase of IDLE CPU time by up to $15 \%$ - due to the additional disk calls. But if we need these terminals more than $15 \%$ of processor time, there is no reason not to connect them. The same holds for other similar trade-offs: Intercom response time versus IDLE CPU time, and of these two versus number of terminals connected and so on.

The results of measurements proved, that for a given situation we cars optimize usage of computer resources - but they cannot give answer to a question what situation we should choose.

## ACKNOWLEDGEMENTS

 ski and Dr. V.P.Shirikov for suggesting possibility of realyzing our task and also thank them for support and encouragement. And we are very grateful to Dr. J.I.Shelonzsev for his help which succeeded us to do this work in such a short time.

## REFERENCES

1. NOS/BEI System Progranmer*'s Reference Manual (60494100), Sunnyvale, California, USA.
2. Ferrari D. Computer Systems Performance Evaluation. Prentice-Hall, Englewood Cliffs, New Jersey, 1978.
3. WenderMerr T. GSS-A Package for Performance Analysis. RUG, Proceedings, Ecodu 24/VIM27, Montreux, Switzerland, September 26-29, 1977.
4. Вентцель Е.С. Исследование операций. "Сов.радио", М., 1972.
APPENDIX A






















 Received by Publishing Department

Лобоз З.К., Забой Т.И.
Измерения и оптимизация олерационной системи на сDC-6500, ОИЯИ, Аубна
Исследования работы операционной системы показали, что основными источниками простоя центральных процессоров являются дисковме запросы и наснцение периферийных процессоров. Наибольший вред дает насьңение дискового канала, Т.К. это придодит к насыщенио и периферийных процессоров. На CDC-6500 бнли проведены раяличние эксперименты, чтобы найти методы улучшения коэффициента использования центральных процессоро без ухудшения других характеристик системы. Результаты зкспериментов показали, что козффициент использования центральных процессоров может быть улучшен на $15 \%$ в дневное өремя за счет более равномерного распределения запросов между системными дисками, усиления использования других дисков и контролирования числа дисковых запросов от системы. Судестаует возмошность вмсвобождения дополнительно от 14000 в до 24000 в слов центральной памяти за счет выбора оптималь ного набора системных программ резидента центральной памяти.

Работа выполнена в Лаборатории вычислительной техники и автоматизации оияи.

Сообмение объедиманного икститута пдерных исследований. Дуб̈нब 1983
Loboz Z.K., Zaboy T.I. E11-83-522

Measurements and Tuning of Operating System Performance NOS/BE on CDC-6500 at JINR Dubna

Investigation of operating system work has shown, that main sources of central processor Idie Time are disk requests and peripheral processors saturation. Of these two disks channel saturation is more damaging because it produces PPU saturation. To find a method for improving CPU utilization factor without worsening other system characterlstics several experiments were made. The results of experiments show that CPU utilization coefficient
wer were made. The results of experiments show that CPU utilization coefficien
may be improved by up to $15 \%$ during daytime by evently distributing disk requests between available disks, more active usage of other disks and controlling a number of disk calls for system programs. There exists aiso a. possibility of obtaining additional 14000 B to 24000 B words of central memory by selection of appropriate set of central memory resident system programs.

The investigation has been performed at the Laboratory of Computing Techniques and Automation, JINR.

