Multiple usage of the CD PLUS/UNIX system: performance in practice

By Aloysius C. W. Volkers
Medical Library

Ingeborg A. S. Tjiam
Medical Library

Abraham van Laar
Information and Automation Department (DI&A/COH)

Ans Bleeker
Medical Library

Faculty of Medicine and Health Sciences
Erasmus University Rotterdam
*The Netherlands**

In August 1994, the CD PLUS/Ovid literature retrieval system based on UNIX was activated for the Faculty of Medicine and Health Sciences of Erasmus University in Rotterdam, the Netherlands. There were up to 1,200 potential users. Tests were carried out to determine the extent to which searching for literature was affected by other end users of the system. In the tests, search times and download times were measured in relation to a varying number of continuously active workstations. Results indicated a linear relationship between search times and the number of active workstations. In the "worst case" situation with sixteen active workstations, the time required for record retrieval increased by a factor of sixteen and downloading time by a factor of sixteen over the "best case" of no other active stations. However, because the worst case seldom, if ever, happens in real life, these results are considered acceptable.

INTRODUCTION

In 1986, the first CD-ROM workstation with MEDLINE was introduced in the Medical Library of the Faculty of Medicine and Health Sciences of Erasmus University in Rotterdam, the Netherlands. After a short time, the single system proved to be insufficient and more CD-ROM workstations were added. Meanwhile, microcomputers were introduced in most departments. Subsequently, end users expressed the wish to consult MEDLINE from their workplaces. During 1989, a local area network was built and in 1991 a multi-CD player (capable of playing fourteen compact discs) was installed. In this way, it became possible for faculty members to retrieve literature from the workplace.

There are about 1,200 connections to the network. Not all of these connections are expected to be used for literature searches. Excluding administrators and technicians, about 500 staff members are expected to use the system regularly for searching literature. Many network applications are accessed through menus. The administrative functions included in the menu system enabled the librarians to estimate that a maximum of about ten persons simultaneously were accessing the most recent MEDLINE year on the multi-CD player.

Meanwhile, development of microcomputers continued. With the introduction of increasingly powerful, personal computers (PCs), the literature retrieval programs for use with CD-ROMs also became more

* Communications should be addressed to Dr. A. Bleeker, Medical Library, Faculty of Medicine and Health Sciences, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.

powerful. This situation presents a problem for owners of older, slower microcomputers with less memory than the new models. Although old computers still function satisfactorily for many other tasks, they cannot be used with the new CD-ROM systems.

A recent development in this field is the UNIX host, with databases uploaded from CD-ROMs to a high-capacity hard disk. All kinds of microcomputers can be connected to the UNIX host. With communication programs, such as the National Center for Supercomputing Applications' telnet, the PCs function only as terminals. The main work is done by the central processor of the host; the local PC only needs to run the keyboard, monitor, and connection to the network. A second advantage of the UNIX host system is that much less information is transported via the network than is the case with the central CD-ROM system.

MULTIPLE USAGE STUDY

As after the introduction of the central multi-CD player [1], the medical librarians at Erasmus began to wonder whether one end user doing a literature search would be affected by the presence of other end users of the system. This article describes a study performed to answer this question and interprets the results.

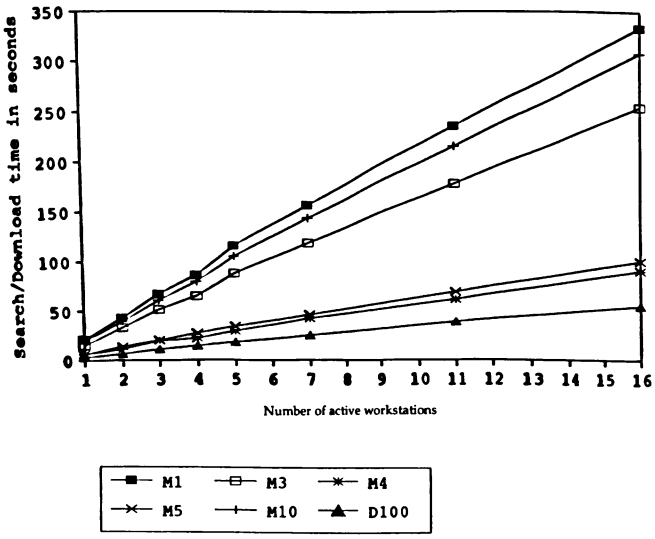
MATERIALS AND METHODS

In the study, test situations were simulated in which one end user searched or downloaded literature references at one workstation, while other end users searched at other workstations. The tests were performed in collaboration with the Information and Automation Department of the Erasmus Faculty of Medicine and Health Sciences.

The central UNIX host consists of a Sun SPARC-Station 10 with a 12-gigabyte hard disk. Sixteen 486 PCs (33 MHz, 8 MB memory) were connected to the network. The Ovid literature retrieval system was installed on the host.

To prepare for the tests, one workstation was used to search the MEDLINE database covering the period January 1990 to May 1994. The time needed to perform standard functions was measured, and this result was used as the baseline for the test. At the same time, users at a number of other workstations, varying in number from zero to fifteen, engaged in concurrent search activities. To ensure consistency in conditions, each user retrieved references using a query that takes more time than available in the measuring period. The request was "BL$ AND CE$". A single active terminal takes about five minutes to complete this search action.

Figure 1
Averaged search time and download time in seconds in relation to number of searching workstations



Standard test functions

The standard test functions were searching with realistic search profiles, and downloading of (sub)sets of 100 references. The search profiles were created and stored in advance to avoid the need to account for the time required to type search terms. Fifteen profiles of different sizes were available to users (Appendix). Profiles marked with "M" are identical to the ones used in the multi-CD-player tests reported in an earlier article [1].

Most single search actions take less than a second. Combining results is a more time consuming process. In all standard test profiles, users had to make many combinations.

Test situations

Test one. One workstation user searched with five standard profiles, while all other users searched for long periods of time (constant test situation). The search time needed to complete five different standard profiles was measured repeatedly. After each measurement one more workstation was added to the test, until all sixteen were operating.

Test two. One workstation downloaded a set of 100 references while other workstations were searching for long periods of time (constant test situation).

Test three. All workstations searched the same database, each using a different realistic profile. For ev-

Table 1
Search and download times with an increasing number of workstations

Search profile	Search time in seconds Total number of active workstations (inclusive of measuring station) 1	2	3	4	5	7	11	16
M4	6	11	19	22	29	42	61	90
M5	6	13	20	27	34	46	69	100
M3	15	33	51	65	88	119	178	255
M10	19	39	60	79	105	144	215	308
M1	21	43	67	86	116	157	236	334
Mean	13.4	27.8	43.4	55.8	74.4	101.6	151.8	217.4
Download	3	6	11	14	18	25	38	54

Table 2
Duration factors with an increasing number of workstations

Search profile	Relative duration factor Total number of active workstations (inclusive of measuring station) 1	2	3	4	5	7	11	16
M4	1.00	1.83	3.17	3.67	4.83	7.00	10.17	15.00
M5	1.00	2.17	3.33	4.50	5.67	7.67	11.50	16.67
M3	1.00	2.20	3.40	4.33	5.87	7.93	11.87	17.00
M10	1.00	2.05	3.16	4.16	5.53	7.58	11.32	16.21
M1	1.00	2.05	3.19	4.10	5.52	7.48	11.24	15.90
Mean	1.00	2.07	3.24	4.16	5.55	7.58	11.33	16.22
Download	1.00	2.00	3.67	4.67	6.00	8.33	12.67	18.00

ery workstation, the time required to finish the profile was measured twice: first during solo operation, and second with all workstations working simultaneously. The latter is not a constant test situation.

RESULTS

Test situation 1: all workstations searching

The first test evaluated whether search time is influenced by the total number of workstations searching simultaneously. The results are shown in Table 1. Table 2 shows the same data, but the results are normalized. The values are divided by the search time required when only one workstation was active.

Figure 1 shows the relationship between the number of active workstations and the search times. There was a linear increase in the duration of search time as the number of searching workstations increased from one to sixteen. It is clear that the search time increased by a factor equal to the number of active workstations.

Test situation 2: downloading with one workstation while other workstations are searching

Test 2 evaluated the influence of the number of busy workstations on the time needed for downloading with one workstation. The results of Test 2 are presented in the last rows of Tables 1 and 2 as well as in Figure 1. There was a linear increase in the time required to download a standard set of records as the number of busy workstations increased. The downloading process slowed by a factor of sixteen when a total of sixteen workstations were busy.

Test situation 3: all workstations using real profiles

In situation 3, fourteen workstations were started simultaneously. Each workstation searched for a different real profile. The resulting search times are shown in Table 3. A relative duration factor was calculated by dividing the search times for the simultaneously used profiles and the search times for the separately used profiles.

For the shortest search profile the relative duration factor was fifteen. This is about the same as in the situation where fifteen other workstations were searching for "BL$ AND CE$" (Table 2, the last column). The relative duration factors of more time-consuming search profiles were inversely proportional to the search times required for solo usage. The explanation for this is fairly simple. Profiles of short duration can be completed while the situation is static (i.e., the number of users does not change). For profiles of longer duration, the situation may change. With elapse of time, more and more workstations complete their queries, and the remaining search processes accelerate. This means that profiles of longer duration finish more quickly than might be expected based on the total number of workstations operating.

CONCLUSIONS

The results of the present study differ from those obtained in the similar tests using the old multiple CD-ROM player, CD Net [2]. To run several strategies, a lone user needs less time with UNIX than with CD Net. But with a number of concurrent users, the UNIX system is slower than the CD Net system. One explanation is that the two systems are fundamentally different. The CD-ROM system reads information much slower than does the UNIX hard disk system. Also, the CD Net system data are transported to the local microcomputer, which in turn, has to do its job for only one user, while in the UNIX system the central processor does the work of all users concurrently. Thus, the UNIX system employs time-sharing.

While the results of this study clearly show that search and download times increase with the number of simultaneous users, the differences probably will

Table 3
Search times with all workstations simultaneously searching on the same database

Search profile	Search time in seconds: Separate 1	Search time in seconds: Simultaneous 2	Relative duration factor 3
S3	3	45	15.0
M2	5	67	13.4
M16	6	98	16.3
M4	7	92	13.1
S2	12	159	13.3
M3	15	161	10.7
M15	19	206	10.8
M9	21	224	10.7
M1	23	210	9.1
S1	26	210	8.1
M13	29	241	8.3
M8	37	258	7.0
S4	53	300	5.7
M12	64	370	5.8

not have a noticeable effect. In practice, searching references is not a continuous process; it includes typing the keywords, reading the displayed results, printing results, thinking about the next action, and doing nothing at all. These activities do not burden the central processor to a noticeable extent. Furthermore, most retrieval actions occur within fractions of a second. A minor slowdown cannot be considered a problem. Combining sets takes considerable time but is only sporadically necessary.

In sum, an end user may notice some increase in the time required for searching and downloading as a consequence of multiple usage of the CD PLUS/UNIX system. But this increase should not be dramatic during normal usage.

ACKNOWLEDGMENTS

The authors wish to thank the Erasmus Computer Department for installing the hardware and software, especially the CD PLUS/UNIX system in the Novell network, and Mrs. Laraine Visser-Isles for correcting the manuscript.

REFERENCES

1. VOLKERS ACW, TJIAM IAS, VAN LAAR A, VAN EIJK HG, et al. Multiple usage of CD-ROMs using Meridian Data's CD Net: performance in practice. CD-ROM Professional 1992; 5(6):91–6.
2. IBID.

Received November 1994; accepted April 1995

APPENDIX

Search strategies

M1: ((HEART$.TI,AB,SH. OR CARDIAC$.TI,AB, SH.) AND TRANSPLANT$.TI,AB,SH. AND (REGIMEN$.TI,AB,SH. OR COMPLIANCE$. TI,AB,SH.))
M2: (((EXP NERVOUS SYSTEM/ OR EXP NERVOUS SYSTEM DISEASES/) AND EXP NEOPLASMS/) OR EXP NERVOUS SYSTEM NEOPLASMS/)
M3: ((INFLUENZA$.TI,AB,SH. AND (ETIOL$.TI, AB,SH. OR CAUS$.TI,AB,SH.)) OR INFLUENZA/ET)
M4: (EUTHAN$.TI,AB,SH. AND (STATIST$.TI,AB, SH. OR EPIDEMIOL$.TI,AB,SH) OR EUTHANASIA/SN OR EUTHANASIA/EP)
M5: (MYOCARDIAL INFARCTION.TI,AB,SH. AND DUT.LG. AND HUMAN.SH.)
M8: ((MELANOM$.TI,AB,SH. OR NEVUS$.TI,AB, SH. OR NAEVUS$.TI,AB,SH.) AND (THERAP$.TI,AB,SH. OR TREAT$.TI,AB,SH.))
M9: (((SPORT$.TI,AB,SH. OR ATHLET$.TI,AB,SH.) AND INJUR$.TI,AB,SH. OR ATHLETIC INJURIES$.TI,AB, SH. AND DIAGN$.TI,AB,SH.) AND (LOWER$.TI,AB, SH. AND EXTREM$. TI,AB,SH. OR LEG.TI,AB,SH. OR LEGS.TI,AB, SH.))
M10: ((ATHEROSCLER$.TI,AB,SH. OR ARTERIOSCLER$.TI,AB,SH.) AND FAT$.TI,AB,SH. AND (CONSUMPT$.TI,AB,SH. OR FOOD$.TI,AB, SH. OR FEED$.TI,AB,SH. OR MEAL$.TI,AB,SH. OR INTAKE$.TI,AB,SH.))
M12: ((BLOOD$.TI,AB,SH. OR PLASM$.TI,AB,SH. OR SERUM$.TI,AB,SH.) AND PROTEIN$.TI, AB,SH. AND (ANALY$.TI,AB,SH. OR DETERM$.TI,AB,SH.))
M13: ((STOMACH$.TI,AB,SH. OR GASTRIC$.TI, AB,SH. OR BOWEL$.TI,AB,SH.) AND HORMON$.TI,AB,SH. AND RECEPTOR$.TI,AB, SH.)
M15: (((CEREBRAL$.TI,AB,SH. OR BRAIN$.TI,AB, SH.) AND (HEMODYNAM$.TI,AB,SH. OR HAEMODYNAM$.TI,AB,SH. OR BLOOD$.TI, AB,SH. AND (FLOW$.TI,AB,SH. OR VELOCIT$.TI,AB,SH.)) OR CBF$.TI,AB,SH.) AND (XENON$.TI,AB,SH. OR CLEAR$.TI,AB,SH.))
M16: ((VISUAL$.TI,AB,SH. AND ACUIT$.TI,AB,SH. OR VISUS$.TI,AB,SH. OR CONTRAST$.TI, AB,SH. AND SENSITIV$.TI,AB,SH.) AND AMBLYOP$.TI,AB,SH.)
S1: EXP NEOPLASMS/DH,DT,NU,PC,RH,RT, TH,SU,TR
S2: (BLOOD AND CELL).TI,AB,SH.
S3: (LIVER AND CIRRHOSIS).TI,AB,SH.
S4: ((LIVER OR HEPA$) AND CIRRHO$ AND (THERAP$ OR TREAT$)).TI,AB,SH.
S5: EXP*VITAMIN D DEFICIENCY/ AND RH.XS. AND EXP AGED/