MONITORING USAGE OF UNIX SOFTWARE

Stephen B Taubman, Federal Reserve Board

ABSTRACT

Fifteen software packages are frequently used by 350 UNIX users at our site. Most of this software is either statistical, such as Gauss and S–plus, or word processing, such as WordPerfect and Interleaf. Each month, usage reports, in the form of graphs and charts, are produced. They show the number of distinct users, number of executions, and other information for each package. This paper will describe how SAS® software is utilized to produce these reports and graphs. These reports are used to help determine the future status of our software environment. BASE SAS and SAS/GRAPH® software are used to produce these reports. A novice SAS user should understand most of this paper.

GOALS

Our UNIX network, which has 350 users, has the following key software packages: SPEAKEASY, S–PLUS, RATS, SAS, MATHEMATICA, MATLAB, TSP, GAUSS, STATA, OCTAVE, EMACS, INTERLEAF, WORDMARC, LATEX, AND WORDPERFECT.

Our goal is to summarize the use of these packages. Specifically, we wish to produce monthly reports which:

1. Display the number of distinct users for each package.

2. Display the number of executions for some packages.

3. Display software usage by time of day and day of the week.

4. Display usage information for the prior three months.

5. Maintain user lists so e–mail can be sent to users of certain software when appropriate.

LOG FILES

A separate log file is maintained for each software product. Every invocation of software adds a line to the appropriate log file. A typical entry in the log file for any software product follows:

user81 Tue Aug 1 17:00:31 EDT 2000 mnane sec1 (Mary Happy)

where:
'user81' is the user's logon_id and Mary Happy is the name of the user,

'Tue Aug 1 17:00:31 EDT' is the date and time of the software execution,

'mname' is the user's machine name, and

'sec1' is the user's section.

REPORT 1

The first report we create is a monthly summary usage chart for each software package.

An example of a monthly summary chart follows:

	Unix Speakeasy users for JULY 2000			
NAME	Frequency	Percent	Cum Freq	Cum%
Mary Happy	49	28.0	49	28.0
I M Content	34	19.4	83	47.4
Ann Smiles	24	13.7	107	61.1
. .	.	. .	. .	.
Phil Good	1	0.6	174	99.4
Sam Helps	1	0.6	175	100.0

This report is created by reading the log file into a SAS dataset and executing the SAS statements

```
proc freq order=freq;
table name;
```

REPORT 2

Next, a single file is produced that summarizes monthly use for all the software packages.

An example of some lines from this file (one line for each user and product) follows:

```
spea    user53      49
spea    user 45     45
spea    user27      11
rats    user45     115
splu    user07       1
tsp     user27     188
tsp     user53      18
```

The three fields above represent software package, user id, and number of executions. This file is created from the individual software summary files described in REPORT 1. Next, we create a single report that summarizes use for all software packages for the three previous months. The three month file is formed from

the three previous monthly files (the prior two monthly files are always saved) using BASE SAS. The format of the three month files is identical to the single month file described above.

REPORT 3

We next create a three month usage report for each software product by both user and section. Using the three month file described in REPORT 2 as input, we use BASE SAS to create two new reports which are then made available on our Intranet. These reports are UNIX software usage by user for the prior three months, and UNIX software usage by section for the prior three months.

A segment of the usage by user file:

TOTAL GAUSS USAGE BY USER
MAY 2000 – JULY 2000

NAME	NUMBER	%OF TOTAL
Fred Pleasant	294	31.85
Mary Happy	166	17.98
Ann Smiles	127	13.76
Sam Helps	47	5.09

A segment of the usage by section file:

TOTAL SAS USAGE BY SECTION
MAY 2000 – JULY 2000

SECTION	NUMBER	% OF TOTAL
section03	12536	22.40
section21	9284	16.59
section14	7374	13.18
section04	5411	9.67

These files contain data for each software product and are the output from the PROC PRINT command. They are then written to HTML files by invoking the SAS out2htm macro. Some of this code follows:

```
%out2htm(capture=on,
    window=output, runmode=b);

    PROC PRINT statements

%out2htm(htmlfile=html_output_file,
        capture=off,
        window=output, runmode=b);
```

GRAPHS

We next produce graphs displaying software usage for the previous few years. To accomplish this, a file must be maintained that records software usage over time. Our file contains usage statistics starting November 1993. This file is updated by adding a new record each month. A segment of this historic usage file follows.

	speakez	rats	matlab	math	splus	tsp	sas
may2000	253 20	3343 38	778 31	567 26	2706 58	243 8	16302 121
jun2000	286 21	5312 44	931 37	497 24	2849 69	192 8	22136 133
jul2000	175 19	3913 32	294 22	271 28	2538 66	460 6	17531 134

Two numbers are listed for each software package. They represent the number of executions and the number of distinct users. For example, in June there were 133 distinct SAS users and 22,136 SAS executions. Then, SAS programs read this file and use SAS/GRAPH software to produce plots displaying usage over time. The key lines of code in these programs are similar to:

```
symbol5 i=join c=blue  ;
proc gplot data=usage;
plot (softw1 softw11 softw3 softw9 ) *month /overlay
haxis='01nov97'd to current by month
vaxis=0 to 4000 by 1000 annotate=legend1;
```

These seven charts are displayed on the Intranet:

NUMBER of MONTHLY EXECUTIONS for SPEAKEZ SPLUS RATS LOTUS MATH

NUMBER of MONTHLY EXECUTIONS for MATLAB TSP GAUSS STATA OCTAVE

NUMBER of DISTINCT USERS for SPEAKEZ SPLUS LOTUS SAS FAT

NUMBER of OF DISTINCT USERS for RATS MATLAB MATHEMATICA TSP GAUSS

NUMBER of DISTINCT USERS for WORD PROCESSING SOFTWARE

UNIX SAS MONTHLY EXECUTIONS

TOTAL SOFTWARE USAGE by HOUR OF THE WEEK

The last graph shows total software usage (the combined executions of all software packages) for each hour of the day and each day of the week over time. Each month, a SAS data set is saved that contains combined usage for each hour of the week. (The time of execution can be obtained from the log file.) A typical graph is displayed at the end of the paper.

MAINTAINING MAILING LISTS

A list of software users is kept for most software packages. These lists are used to send e–mail whenever there is important news pertaining to that software. The mailing lists are updated each month by including new software users. The key lines of code for the software that updates the mailing lists are:

```
 data merdata;
            merge monthusr (in=in1) oldlist( in=in2);
            by name;
            if in1 eq 1 and in2= 0;
run;
proc print;
title ' new users to be added to mailing list';
run;
```

It is also important to periodically review these mailing lists and remove the names of people leaving the network.

OTHER FACTS

1. The above reports are run near the first of each month.

2. Logs from the prior month are removed shortly thereafter.

3. Monthly summaries are retained for three months.

4. Only the latest reports and charts are displayed on the Intranet.

5. We ignore all log records where the user is either the software maintainer or a systems programmer because we want to report only application users.

PURPOSE OF THE SYSTEM

1. Observe trends in software usage.

2. Decide if the number of current software licenses (number of simultaneous users permitted) is adequate.

3. Determine the level of user support necessary for each package.

4. Maintain mailing list so software news can be sent to current users current users

5. Produce ad hoc reports based on monthly summaries or logs.

ACKNOWLEDGMENTS

SPEAKEZ   SPLUS RATS and SAS
NUMBER OF DISTINCT USERS

LEGEND

SPEAKEZ

SPLUS

RATS

SAS

MONTH