Further thoughts on economics of SUMEX.

Since we will be giving out tokens with one hand, and collecting with
the other, there are ibvious limitations in analogizing the liberty
of consumer preference on SUMEX and on a free market. It will then
help in building an economic model to clarify just what we are trying
to optimize. I will make a fisrt attempt at this, and then suggest
that we may be able to dispense with a token economy (and its
non-trivial costs of administration, negotiation etc.) in favor of a
centrally administered patterming of priorities.

Certainly we have no objective PROFIT function to maximize; though in
the long run SUMEX should be operated AS IF it maximized the income
that it could extract from users at their own valuation of marginal
services, i.e., what the traffic could bear == what the service is
worth to them in aggregate.

However, Health has been professionalized, and HEALTH-RESEARCH
nationalized long since, and we can get into serious trouble by
inappropriate mixes of central planning and free market arrangements
The MAXI-PROFIT notion isan abstraction that can give us some
(limited) guidance to planning.

Some postulated principles of operation include:

The SUMEX-AIM community comprises a limited set of investigators who
are to be encouraged and supported in the pursuit of 1) explicit
research programs and 2) related but less well-defined explorations.

We have a constrained budget for capital investment that is the
principal limitation to the overall volume of service that can be
delivered.

3. Each user will be judged to have some service-value function which
falls asymptotically to zero (that is plotting the utility of the
next increment of service against total volume consumed). At zero
but not necessarily to first order social utility may or may not
correspond to how a user would spend his own dollar budget in a
convertyible currency.

4. Given the constraints of a fixed-size machine, and the managerial
ones of a finite community of users we should, roughly, optimize the
integral of the product of services , utility, i.e., allocate the
next increment of available service to the user judged to consume it
at huighest utility. In practice this means we must assure each user
an opportunity to get his central quantum of work done; and we must
take account of the side-costs of delivering services are uneven high
loadings etc.

5. We have to be able to justify any overt inefficiency, idle time
etc. We also have to be accountable for various aspects of FAIRNESS,
especially in re the charter of a 40-40-20 slicing.

6. There aee three evident measures of service: how-much (cpu); when
(time-of-day or demand availability); and how-fast (throughput rate)
athat will bear differently oin different users at different times.
   These consideratuions suggest the following approach to allocation.

We will not have tokens at all (except perhaps in re connect time,
especially for remote users a/c the relevant costs.) Instead a 3- or
4-tier priority system.

PRIORITY 1. USE during SCHEDULED time of day. Each USER will be
allotted an hour per day of connect time (or some multiple thereof)
which he can
'reserve' for a week in advamnce by voluntary posting. This does not
interfere with overlapping schedules by others, nor with voluntray
side-agreements to avoid overlap. The point is to have some
framework in which users can PLAN to have the most efficient access
possible SUMEX management can also play a persuasive role in such
postings.

      Users in priority-1 status will compete for the first 60% of
machine cycles (divided 30:30 AIM:SUMEX) regardless of other quotas.

PRIORITY 2. (SYS lives here chronically). Users compete for 80% of
the RESIDUE of the machine, (i.e. 24% plus spillover from priority 1.
Users will work in this level until they have used up their daily
quota of CPU time.In addition work involving routine EXEC and text
and file-handling should be upgraded from level 3 into this queue.
PRIORITY 3 gets the rest, either at par, or after some further
adjustments.

      This general scheme of course admits of many further tuning
steps. Under level 3, for example, we should consider unloading jobs
that do nt get enough CPU attention to warrant keeping them in the
queue. In the same vein we should have some provision for autologout
of inactive connect lines that merely burn up communications costs,
and for handling inactive forks... but we have to analyze what
penalty these drags (and their solutions) impose.

      The present proposal is of course substantially what Rainer has
discussed and partly implemented, except for the overlay of
Priority-1. I am not averse to some gradualism in shifting between
the levels, but a user should have enough predictability about how he
is being handled by the system that he can plan his work, rather than
just sit totally passively hoping for the best possible.   In
allocations of level-1 svhedules, SUMEX-management can of course play
a more or less active role in structuring the traffic if the
cirstances require (like a traffic cop relates to signal lights and
stco signs); and I can eventually foresee some game-like algorithms
to nelp organize these schedules.   The assignment of %-ages among
different levels is of course a further management option.

The problem with the token economy alternative is just how to allocate the chips to start with and the time we will spend negotiating grievances whenever the currency is re-flated and the pricing system altered. If we did go that route, we might want to think of a continuous auction to set current prices -- which is a kind of parody of the difficulties.

But there is some room for chips in the present scheme: 1. The user's choice in posting level-1 connect time-advance-reservation. 2. Spending CPU quota to stay in this priority 3. Overall connect-time limitations.

The effucient exploitation of the resource is certainly going to require some form of quasi-batch-background level of operation to spend the non-prime-time cycles. I need to know more about connect time costs to judge the related issue of DETach/REDIRect activity.

Behind all this discussion is a mmodel of user activity that I will be trying to make more explicvit and perhaps to simulate. (I have wanted for some time to start some work on applying AI to treaty negotiation, and the mechanizedf induction of schjmes like this one might be a reasonable challenge.)

     FURTHER COMMENT   in re Priority 1   (11/29/74)

     The basic logic of this arrangement is to find the optimum comoromise between the level of structuring that will enable users to plan the most efficient use of their own time, and the flexibility that enables ad hoc response to the exigencies of their task. Management has available the options of varying the relative allocation of priority one cycles down or up from 60%, and also of taking a more active role in arranging for the staggering of such scheduled use.
     It may be asked, cuite reasonably, whether there will not be an automatic regulation of usage in the light of the diurnal cycle of responsiveness of the system -- i.e., whethers users will not simply adjust their own schedules to what they observe in loading. This may well happen; but I foresee that there will still be disappointments arising from unpredicted interference. We must also take account that many of our users will have special constraints --e.g. interaction with knowledge-consultants,patients, demonstrations, etc., that will be greatly hindered by lack of some scheduling structure. On the other hand, we wish to avoid an excessively competitive framework -- a rush to the starting line -- that will distract from the actual substance of working on the system.

     Another possible option will be to allow a highly limited opportunity for trading in some other 'assets' in favor of a URGENT priority level that can override the current schedules.

     Only the major project directors need to inform themselves of the details

of these scheduling arrangements which are rather complex, and may be
subject to change from time to time during our shakedown period. They
should be able to communicate to their collaborators simple
guidelines about when to work and what priority to request.

Another advantage of this system is the possibility it offers of
altruistic cooperation, viz., the voluntray self-assighment of a
reduced priority level for tasks that do not have an urgent need for
prompt completion. Likewise, managewill be able to enable a wider
range of projects, for the most efficient utilization of the machine,
if some of these can be automatically kept from interfering with
high-priority users at busy times.

Some earlier comments on economic model

ONE OF THE DANGERS OF SETTING UP TOKENS IS THAT IT WILL PROMPT
QUESTIONS, WHY FUNNY MONEY INSTEAD OF "REAL" DOLLARS?

WE SHOULD BE PREPARED WITH THE FOLLOWING ARGUMENT: TO JUSTIFY TOKENS
VS. REAL MONEY, WE SHOULD STRESS THAT THIS IS AN EXPERIMENTAL SCHEME,
UNLIKE REAL MONEY, WE CAN START OVER AGAIN AND MAKE MANY OTHER KINDS
OF ADJUSTMENTS IN THE COURSE OF PERFECTING THE PRICING ALGORITHMS
ETC..

MANY A COMPANY HAS GONE BROKE OVER ERRORS IN ITS PRICING
ALGORITHMS. JOSH