The Economics of Dumping from Electronic Computers

By D. Kershaw and S. Vajda

We assume that it is known that a job takes a time H on an automatic computer to complete, provided no breakdown of the computer occurs. We also assume that the probability of a breakdown occurring during a small time interval Δt is $\Delta t/A$, so that the probability of its working for at least t hours without a breakdown is $\exp(-t/A)$. With this assumption, A is the average time interval between breakdowns.

The possibility of a breakdown before the job is finished, and the resulting loss of time spent on computing, suggests that it might be advisable to arrange for intermediate results to be punched out ("dumped") at some intervals. The length of such intervals will depend on H, A, the time the punching out takes, and on the frequency of such intermediate output. It will be assumed here that the time of punching out is independent of the stage of the computation reached, and will be denoted by K. It will also be assumed that there is no intermediate input, and that the output of the final result at the end of the production time takes the same time K. The unit of time is irrelevant.

The time of the original input of information, and the possibility of the input equipment breaking down, will be ignored.

As the probability of a successful run of length of at least t is $\exp(-t/A)$, the probability of a breakdown occurring during the time interval $(t, t+dt)$ is

$$\frac{1}{A}\exp(-t/A)dt. \tag{1}$$

After a breakdown we start again, and it is possible that a further breakdown occurs during the same stage. Denoting the expected total time lost during this stage, due to any number of breakdowns, by $X(t)$, we have the relation

$$X(t) = \frac{1}{A}\int_0^t [\tau + X(t)]\exp(-\tau/A)d\tau, \tag{2}$$

so that
$$X(t) = A(e^{t/A} - 1) - t. \tag{3}$$

Here A is the average time interval between breakdowns resulting from any cause whatsoever. If the length of a run is taken to mean the time T during which the computer works on a problem, together with the time K for dumping, the expected time taken to complete the run will be

$$X(T+K) + T + K = A[e^{(T+K)/A} - 1]. \tag{4}$$

Now if a job takes a total computing time of H to complete, we may ask when dumpings should take place in order to minimize the total expected time taken. The following argument shows that dumpings should be made after equal intervals of computing time.

Let dumpings be made at times $t_1, t_2, \ldots, t_{n-1}$, then the expected time taken to complete the job is

$$A\sum_{i=1}^{n}[e^{(t_i - t_{i-1}+K)/A} - 1], \qquad t_0 = 0, \quad t_n = H. \tag{5}$$

Equating the partial derivatives of this expression w.r.t. t_i, $i = 1, \ldots, n-1$, to zero gives the set of equations:

$$e^{(t_i - t_{i-1}+K)/A} = e^{(t_{i+1} - t_i + K)/A}, \qquad i = 1, \ldots, n-1$$

and so
$$t_i - t_{i-1} = t_{i+1} - t_i$$

i.e.
$$t_i = i\frac{H}{n}.$$

It can be shown that these values make (5) a minimum.

Denoting $\frac{H}{n}$ by T, the expected time taken to complete the job is then

$$A\frac{H}{T}[e^{(T+K)/A} - 1] = S, \text{ say.} \tag{6}$$

It is now necessary to find, for a given A, H and K, the time interval T which will minimize this expression such that H/T is an integer. For convenience we introduce the non-dimensional notation:

$$Y = T/A, \; L = K/A$$

making
$$S = \frac{H}{Y}[e^{Y+L} - 1] = S(Y). \tag{7}$$

The turning points of S are given by

$$\frac{dS}{dY} = 0 = \frac{H}{Y^2}[(Y-1)e^{Y+L} + 1]$$

i.e. when
$$1 - Y = e^{-Y-L}. \tag{8}$$

As
$$\frac{d^2S}{dY^2} = -\frac{2H}{Y^3}[(Y-1)e^{Y+L} + 1] + \frac{H}{Y}e^{Y+L} \tag{9}$$

$$\frac{d^2S}{dY^2} = \frac{H}{Y}e^{Y+L} > 0 \text{ at a turning point}$$

and so there is only one solution Y_0 of (8) giving a minimum value of $S = \frac{H}{1 - Y_0}$.

This value of Y makes the number of intervals n to be taken equal to $\frac{H}{AY_0}$. In general this number will not be integral, and the value of n to take will be the one of $\left[\frac{H}{AY_0}\right]$ and $\left[\frac{H}{AY_0}\right]+1$ which makes S smaller ($[x]=$ integral part of x). An example will show how this is done.

Table 1 gives a table of the solutions of (8) for varying $L = K/A$, so if $A = 10$, $K = 1$, $H = 10$, then $L = 0\cdot 1$ and the solution of (8) in this case is $Y = 0\cdot 3832$, leading to $n = 2\cdot 6\ldots$

The number of intervals to take is then either 2 or 3, depending on which gives the smaller value of S.

When $n = 2$, $\frac{H}{T} = 2$ and the time interval $T = 5$ leads to $S = 16\cdot 44$, computed from (6).

Similarly, when $n = 3$, it is found that $S = 16\cdot 27$ and so the minimum expected time taken in this case is obtained when the job is divided into 3 intervals.

Table 1 can be used in this fashion to find the optimum number of intervals for any job subject to the restriction that $0\cdot 005 \leqslant K/A \leqslant 0\cdot 2$. But for a given machine of known reliability and for a fixed dumping time it would be more convenient to have a critical table showing the range of H where a certain number of dumpings is optimum.

The critical values of H are those for which $S(H, n) = S(H, n+1)$,

where $$S(H, n) = An[\exp(H/An + K/A) - 1], \qquad (10)$$

i.e. the expected time taken for a job of length H with n intervals.

The tables were constructed according to the following argument.

Put $$\phi_n(H) = S(H, n+1) - S(H, n) \qquad (11)$$

then

$$\phi_n'(H) = e^{K/A}\left(e^{\frac{H}{A(n+1)}} - e^{\frac{H}{An}}\right) \leqslant 0 \text{ for } H \geqslant 0, \qquad (12)$$

and as $$\phi_n(0) = A(e^{K/A} - 1) > 0$$

and $$\phi_n[n(n+1)A] = A[e^{n+\frac{K}{A}}(n+1-en) - 1] < 0$$

there is a zero of $\phi_n(H)$ in $(0, \infty)$, H_n say. Hence for a job of length H_n the expected time taken is the same for n and $n+1$ intervals. It will now be shown that for any H in the interval $(H_{n-1}H_n)$ the optimum number of outputs is n.

The number of intervals to be taken is the one of $\left[\frac{H}{AY_0}\right]$ and $\left[\frac{H}{AY_0}\right]+1$ which gives the smaller value of $S(H, n)$.

As $$H_{n-1} < H < H_n$$

then $$\frac{H_{n-1}}{AY_0} < \frac{H}{AY_0} < \frac{H_n}{AY_0}.$$

Now $\frac{H_n}{AY_0}$ gives rise to n and $n+1$ and $\frac{H_{n-1}}{AY_0}$ gives rise to $n-1$ and n, so $\frac{H}{AY_0}$ is one of $(n-1)$, n, $(n+1)$.

It has been seen that $\phi_n(H)$ decreases from the positive value $\phi_n(0) = A(e^{K/A} - 1)$ in $(0, H_n)$, so that in $(0, H_n)$ we have $$\phi_n(H) > 0, \text{ i.e. } S(H, n+1) > S(H, n). \qquad (13)$$

In the range (H_{n-1}, ∞), $\phi_{n-1}(H)$ decreases from the zero value $\phi_{n-1}(H_{n-1})$, and so in (H_{n-1}, ∞)

$$\phi_{n-1}(H) < 0,$$
$$\text{i.e. } S(H, n) < S(H, n-1). \qquad (14)$$

Table 1

$1 - Y = \exp(-Y - L)$

L	Y
0·005	0·0967
0·010	0·1348
0·015	0·1634
0·020	0·1869
0·025	0·2073
0·030	0·2254
0·035	0·2418
0·040	0·2568
0·045	0·2708
0·050	0·2838
0·055	0·2961
0·060	0·3076
0·065	0·3186
0·070	0·3290
0·075	0·3390
0·080	0·3485
0·085	0·3577
0·090	0·3665
0·095	0·3750
0·100	0·3832
0·105	0·3911
0·110	0·3988
0·115	0·4062
0·120	0·4134
0·125	0·4204
0·130	0·4272
0·135	0·4338
0·140	0·4402
0·145	0·4465
0·150	0·4526
0·155	0·4586
0·160	0·4644
0·165	0·4701
0·170	0·4757
0·175	0·4812
0·180	0·4865
0·185	0·4917
0·190	0·4968
0·195	0·5018
0·200	0·5068

Table 2a

$S(n) = An\,(\exp\,(H/An + K/A) - 1)$

$A = 10$

$K = 1{\cdot}0000$

H	n	$S(n)$	T
	1		
5·377		8·922	8·922
	2		
9·361		15·297	18·183
	3		
13·256		21·577	31·605
	4		
17·123		27·826	51·244
	5		
20·977		34·062	80·040
	6		
24·824		40·291	122·284
	7		
28·667		46·515	184·270
	8		
32·507		52·736	275·226
	9		
36·345		58·956	408·692
	10		
40·183		65·174	604·532

Table 2b

$S(n) = An(\exp\,(H/An + K/A) - 1)$

$A = 3$

$K = 0{\cdot}0167$

H	n	$S(n)$	T
	1		
0·431		0·483	0·483
	2		
0·747		0·833	0·870
	3		
1·057		1·178	1·291
	4		
1·365		1·520	1·754
	5		
1·672		1·862	2·266
	6		
1·978		2·203	2·832
	7		
2·284		2·543	3·459
	8		
2·590		2·883	4·152
	9		
2·895		3·224	4·919
	10		
3·201		3·564	5·769

The result is then that in (H_{n-1}, H_n), the number of intervals to take is n.

Tables 2*a* and 2*b* are critical tables computed for $A = 10$, $K = 1$, and $A = 3$, $K = 0{\cdot}0167$ (which may be interpreted as $A = 3$ hours and $K = 1$ minute) respectively.

The example can be solved quickly by Table 2*a*. As $H = 10$ lies between 9·361 and 13·256, the number of intervals is read off as $n = 3$.

The second and third columns in Tables 2*a*, 2*b* indicate the expected times taken, with and without dumping, at the critical value of H respectively. If the time taken for dumping is very small compared with the average time between breakdowns, e.g. using magnetic tape or drum storage, then Table 2*b* shows, as would be expected, that dumpings should be made relatively more frequently.

Dr. I. J. Good has drawn our attention to the paper "On Optimum Utilization of Faulty Channels," by Arthur Wouk, *J. Soc. Industr. Appl. Math.*, Vol. 9, p. 311 (1961), where, amongst others, a problem similar to ours is studied. Using our concepts, his problem is the same as ours, with the modification that a breakdown is not immediately realized, so that the run is completed and the print-out indicates that a repetition is required. It is interesting to note that in the case of an exponential distribution the equation which solves that problem is equivalent to our equation 8.

Miss E. Fitzpatrick programmed the computation of the tables, and the headings were printed using a subroutine described in the reference.

We are also grateful to Dr. E. M. Wilson and the referee for useful suggestions.

Reference

KITZ, BERYL (1959). "Two New Facilities for the Pegasus Autocode Scheme," Admiralty Research Laboratory Report No. A.R.L./R2/Maths. 2.9.