Free store distribution under random fit allocation: Part 1

C. M. Reeves

Department of Computer Science, University of Keele, Keele, Staffs ST5 5BG

Hitherto, most of our understanding of dynamic storage allocation techniques has come from simulation studies, with Knuth's Fifty Percent Rule as the notable exception. This paper extends the analytic approach to derive equations relating the equilibrium distribution by length of free blocks to the corresponding distribution of reservations under a random-fit allocation strategy. In the special case of single word reservations, it is predicted that the existence of a stable solution requires at least 37% store utilisation, and an analytic expression for the free store profile is obtained.

(Received August 1978)

1. Introduction

The classic introduction to the concepts and methods of dynamic storage allocation is that of Knuth (1968). He states as the objective, 'We want algorithms for reserving and freeing variable size blocks of memory from a larger storage area, where these blocks are to consist of consecutive memory locations'.

Each individual store location experiences alternate reservations and releases so that, at any instant, there is likely to be considerable fragmentation of free storage into blocks of different size. Randell (1969) has defined external fragmentation to be 'the loss in storage utilisation caused by the inability to make use of all available storage after it has been fragmented into a large number of separate blocks', and internal fragmentation to be 'the loss of utilisation caused by rounding up a request for storage, rather than allocating only the exact number of words required'. Internal fragmentation depends only upon the distribution of request sizes (Wolman, 1965; Gelenbe *et al*, 1973) and is not of interest here.

Garbage collection provides one distinctive approach to the control of external fragmentation. At the start of a cycle of operations, reserved blocks are contiguous at one end of the store, leaving a single block of unfragmented free store at the other end. The initial configuration of an empty store is such a situation. During the cycle, fresh reservations are placed progressively through the free block to extend the reserved area. Releases are marked as such but the space they occupy is not released. When the remaining free store is insufficient to meet the current request, garbage collection is invoked: all reserved blocks are relocated to form a fresh contiguous sequence, and a new cycle commences. This method has its devotees and is widely used. Its disadvantage is that the convenience of ignoring releases must be paid for in mounting the periodic rescue operations.

The characteristic feature of truly dynamic methods of storage allocation is that they coalesce contiguous sections of free storage, arising from releases, without delay. In this way they combat the steady diminution of free block sizes produced by external fragmentation. A dynamic equilibrium results, in which requests continue to be met, without the overhead implicit in block relocation. Knuth commented that there were few analytical results in this area, most information coming from simulations. Ten years later the situation is little different.

Two classes of method may be distinguished, corresponding to structured and unstructured external fragmentation.

In unstructured fragmentation, any free block of sufficient size is eligible to satisfy a request, any residue forming a fresh, smaller free block. Similarly, following a release, any contiguous sequence of free blocks is instantly compacted into a single, larger free block. Algorithms in this class differ in their method of selecting the free block appropriate to each request.

Methods involving structured fragmentation are the buddy system (Knowlton, 1965) and its derivatives. Block lengths are not arbitrary but are powers of 2. A request for L locations will always be allocated the smallest possible free block, of size 2^n where $2^n \geqslant L > 2^{n-1}$, thus minimising internal fragmentation. If no such free block is available, a larger one must be subdivided. A block may only be subdivided into two equal halves corresponding to the relation

$$2^n = 2^{n-1} + 2^{n-1} \tag{1.1}$$

Conversely, following a release, two contiguous free blocks may be compacted only if they are 'buddies' and so would reconstitute the block from which they were derived.

The merit claimed for the buddy system is that, by restricting the number of block sizes, the administration of free storage is simplified. Amongst the variations that have been described, aimed at improved flexibility and reduced internal fragmentation, are the weighted buddy system (Shen and Peterson, 1974; 1975) and the Fibonacci buddy system (Hirschberg, 1973) following the original suggestion by Knuth.

In the Fibonacci buddy system, block lengths are Fibonacci numbers $f_1, f_2, f_3, \ldots$ satisfying the relation

$$f_n = f_{n-1} + f_{n-2} \tag{1.2}$$

in place of equation (1.1). More general systems have been proposed, based upon the relation

$$f_n = f_{n-1} + f_{n-k} \tag{1.3}$$

but it is not clear that they would be more efficient (Nielsen, 1977). It may be noted that the choices $k = 1$ and $k = 2$ recover the two simpler systems. Further discussion of the simple buddy system is given by Purdom and Stigler (1970). Practical improvements to the various systems are discussed by Isoda *et al* (1971), Cranston and Thomas (1975) and Hinds (1975).

The present paper is concerned with unstructured fragmentation. It is convenient, both for exposition and as a reflection of common practice to facilitate compaction, to suppose that the free blocks are arranged as a list, ordered according to their position in the store. Knuth describes two reservation strategies. In 'best fit', a request is allocated the smallest free block of sufficient length. This requires a full search through the free block list. In 'first fit' on the other hand, the first free block of sufficient size in the list is taken.

The 'first fit' strategy tends to generate a preponderance of short blocks at the head of the free block list and so extends the mean search time if a significant proportion of requests are for longer blocks. Knuth proposed a modified 'first fit' strategy, subsequently known as 'next fit', to avoid this demerit. In this, each new search in the list starts, not at the head but at the point where the previous search terminated. For this purpose the list is regarded as circular. By this means a more even spatial distribution is encouraged. An interesting alternative strategy, avoiding the extremes of 'first fit' and 'best fit' is the 'optimal fit' strategy described by Campbell (1971).

'First fit' is a crude, and 'next fit' a less crude, approximation to an idealisation which we may call 'random fit'. In this, every free block of sufficient size is equally likely to be allocated to a request. Knuth's Fifty Percent Rule, the best known (by default) analytic result in dynamic storage allocation, presupposes a 'random fit' strategy. Indeed, observed deviations from the rule have been attributed to a breakdown in randomness. Shore (1977), in particular, has noted the effect of systematic placement of a reservation at one end or the other of an oversize free block. By varying the end selected, at random, better agreement with the rule is obtained.

Clearly a deeper theoretical understanding of 'random fit' would be a practical advantage. It would extend the role of this strategy as a yardstick against which more efficient practical algorithms may be assessed. It is to this end that this paper is directed.

2. Equilibrium under Random Fit

In this section equations are derived relating the equilibrium distribution by length of free blocks to the corresponding distribution of request sizes. The Fifty Percent Rule is recovered by summing over block size.

2.1 *Definition and notation*

Conditions are sought for statistical equilibrium between reservations and releases in a store of N locations operating under a 'random fit' allocation strategy. In order to avoid consideration of the effects at the two ends of a linear store, the store is assumed to be circular. It is further assumed that each request can immediately be satisfied so that there is no necessity to queue it pending sufficient releases to free more storage.

The distribution of reservations by length is assumed known: b_r is the probability that an arbitrary request is for a block of length r. With maximum request length R, where normally $R \ll N$, it follows that

$$\sum_{r=1}^{R} b_r = 1 \tag{2.1}$$

In equilibrium, for each block size, the average rates of reservation and release are the same and so the size distribution of releases is the same as for requests.

Under the above assumptions of equilibrium in an unsaturated store, the size distribution of reserved blocks in store is determined by considerations outside the present analysis such as the initial loading history. The case treated here is that in which the size distribution in the store matches that for reservations and releases. This is the situation that would arise for example if an initially empty store were connected to the input stream of requests with releases inhibited for a period.

The significance of the assumption of a common size distribution for requests, stored blocks and releases is its compatibility with the simple release mechanism in which every stored block, irrespective of size, is equally likely to be released next. It is this mechanism that is applied in the subsequent analysis. In general this would imply a wholly artificial relationship between the duration of a reservation and its allocated position in the store but under the random fit allocation strategy this mechanism is the natural consequence of the simple assumption that the duration of a reservation, whether determinate or random, does not depend upon its size.

It remains to note, as a final preliminary to the analysis, that the quantities introduced are to be understood as denoting average values under the conditions of statistical equilibrium that have been outlined. At certain points of the discussion however, variations in block counts due to individual reservations and releases are analysed and there the counts have the significance of actual values of random variables close to their average values. It is to be expected that the averaged block counts will not be whole numbers but the discrepancies will be insignificant for a sufficiently large store where the counts will be large. The store utilisation θ is defined as the fraction of store occupied by reservations. If the number of reserved blocks is B, then $b_r B$ of them are of length r and so

$$\sum_{r=1}^{R} r b_r B = \theta N \tag{2.2}$$

The free store may be treated similarly. Let f_r be the number of free blocks of length r, F the total number of free blocks, and define

$$\phi_r = f_r / F \tag{2.3}$$

so that ϕ_r is the probability that a free block chosen at random is of length r. For maximum free block size L, the analogous relations hold.

$$\sum_{r=1}^{L} \phi_r = 1 \tag{2.4}$$

and

$$\sum_{r=1}^{L} r \phi_r F = (1 - \theta) N \tag{2.5}$$

where an upper bound on L is seen to be $(1 - \theta)N$, corresponding to zero fragmentation.

Loosely analogous to the store utilisation θ, a block utilisation ratio k is defined as

$$k = B/F \tag{2.6}$$

In the analysis of the distributions arising, it is convenient to work with their generating functions. The generating function corresponding to a distribution v_r, $r \geqslant 1$, is denoted $\tilde{v}(a)$ and defined by

$$\tilde{v}(a) = \sum_r v_r a^r \tag{2.7}$$

2.2 *Block reservation*

Let Q_{ir} denote the probability that a request for a block of size i is satisfied by using a free block of size r, $r \geqslant i$. In the 'random fit' strategy, every sufficiently long free block is equally likely to be used, so that

$$Q_{ir} = f_r / g_i \tag{2.8}$$

where

$$g_i = \sum_{u=i}^{L} f_u \tag{2.9}$$

More conveniently, in terms of ϕ,

$$Q_{ir} = \phi_r / \gamma_i \tag{2.10}$$

where

$$\gamma_i = \sum_{u=i}^{L} \phi_u \tag{2.11}$$

As the result of such a reservation, the free block counts are modified as follows

$$f_r \leftarrow f_r - 1 \tag{2.12}$$

and, if $r > i$,

$$f_{r-i} \leftarrow f_{r-i} + 1 \tag{2.13}$$

Accordingly, the expected increment to $\tilde{f}$ following a single random reservation is denoted by $\Delta^+ \tilde{f}$ and given by

$$\Delta^+ \tilde{f} = \sum_{i=1}^{R} b_i \left\{ -\frac{\phi_i}{\gamma_i} a^i + \sum_{r>i} \frac{\phi_r}{\gamma_i} (a^{r-i} - a^r) \right\} \tag{2.14}$$

2.3 *Block release*

2.3.1 *The compaction mechanism*

The release of a block of size i produces changes in the free block counts. Three cases arise.

1. *No compaction*

The released block lies between two reserved blocks.

$$f_i \leftarrow f_i + 1 \quad (2.15)$$

2. *Single compaction*

The released block lies between a reserved block and a free block of length r.

$$f_{i+r} \leftarrow f_{i+r} + 1 \quad (2.16)$$

$$f_r \leftarrow f_r - 1 \quad (2.17)$$

3. *Double compaction*

The released block lies between two free blocks of lengths r and s.

$$f_{i+r+s} \leftarrow f_{i+r+s} + 1 \quad (2.18)$$

$$f_r \leftarrow f_r - 1 \quad (2.19)$$

$$f_s \leftarrow f_s - 1 \quad (2.20)$$

where, if $r = s$, the latter two relations combine to give

$$f_r \leftarrow f_r - 2 \quad (2.21)$$

2.3.2 *The spatial distribution of blocks*

As a consequence of the immediate compaction of adjacent free blocks, each free block is bounded on each side by a reserved block whereas sequences of contiguous reserved blocks are permitted. The assumption of a circular store thus leads to

$$B \geqslant F \quad (2.22)$$

Three categories of reserved block may be distinguished according to the number of free neighbours. Thus category 2 consists of blocks bounded on each side by a free block. Blocks of category 1 are the ends, and those of category 0 are the interior members, of a sequence of two or more contiguous reserved blocks.

The numbers of blocks in each category are denoted by n_0, n_1, and n_2 where

$$n_0 + n_1 + n_2 = B \quad (2.23)$$

On the assumption of a circular store, the number of separate sequences of one or more reserved blocks is F. Of these, n_2 are of unit length and $F - n_2$ of length two or more. The latter each have blocks of category 1 as their ends so that

$$n_1 = 2(F - n_2) \quad (2.24)$$

To continue the argument, an estimate of n_2 is required. This is obtained by averaging over a representative set of arrangements of the B reserved blocks and F free blocks, treating each such arrangement as equally likely. The resulting expected value for n_2 is carried forward into the block release equations.

Suppose the reserved sequences are numbered 1 to F. Each necessarily contains one of the B reserved blocks. The remaining $B - F$ reserved blocks may be allocated arbitrarily. The arrangements chosen to reflect these constraints correspond to the terms, before collecting coefficients, in the expansion of

$$x_1 x_2 \ldots x_F (x_1 + x_2 + \ldots + x_F)^{B-F} \quad (2.25)$$

There are F^{B-F} such terms and hence F^{B-F} different arrangements. A term such as

$$T = x_1^{l_1} x_2^{l_2} \ldots x_F^{l_F} \quad (2.26)$$

represents an arrangement for which the numbers of blocks in the F sequences are $l_1, l_2, \ldots, l_F$. The number of single block sequences in this arrangement is n, say, where n is the number of variables x_r in T for which $l_r = 1$. This is given by

$$n = \sum_{r=1}^{F} \left[\frac{\partial T}{\partial x_r} \right]_r \quad (2.27)$$

where $[\ ]_r$ indicates that the enclosed quantity is to be evaluated at $x_r = 0$, $x_s = 1$ for $s \neq r$.

The estimate of n_2 now comes from averaging n over all arrangements.

$$n_2 = F^{F-B} \sum_{r=1}^{F} \left[\frac{\partial}{\partial x_r} x_1 x_2 \ldots x_F (x_1 + x_2 + \ldots + x_F)^{B-F} \right]_r \quad (2.28)$$

$$= F \left(\frac{F-1}{F} \right)^{B-F} \quad (2.29)$$

Introducing $k = B/F$ (equation 2.6) now yields

$$n_2 = \frac{B}{k} \left\{ \left(\frac{F-1}{F} \right)^F \right\}^{k-1} \quad (2.30)$$

It is a well known result of classical analysis that

$$\lim_{n \to \infty} \left(\frac{n-1}{n} \right)^n = e^{-1} \quad (2.31)$$

and hence, for large F,

$$n_2 = B(ke^{k-1})^{-1} \quad (2.32)$$

A comment on the above derivation is in order. The method of enumerating and weighting arrangements may well be questioned. Its tractability is a substantial merit and it has plausibility. It would perhaps be foolish to say more, though dishonest to say less.

2.3.3 *The effective compaction rate*

Under a 'random fit' strategy and the assumptions already made, the spatial position of a reserved block has no influence upon its becoming due for release. It follows that the probability that a released block is in category 0, 1 or 2 is p_0, p_1 or p_2 where

$$p_0 = {}^{n_0}/B, \quad p_1 = {}^{n_1}/B, \quad p_2 = {}^{n_2}/B \quad (2.33)$$

so that, using the results above,

$$p_2 = (ke^{k-1})^{-1} \quad (2.34)$$

$$p_1 = 2\left(\frac{1}{k} - p_2 \right) \quad (2.35)$$

$$p_0 = 1 - p_1 - p_2 \quad (2.36)$$

The same assumptions also determine that blocks are distributed randomly in space as regards their length, whether reserved or free. It follows that the probability that the release of a block of length i will yield

(*a*) no compaction is p_0

(*b*) single compaction with a free block of length r is $p_1 \phi_r$, and

(*c*) double compaction with free blocks of length r and s is $p_2 \phi_r \phi_s$.

In the double compaction case it is assumed that the probabilities of the free blocks can be taken as independent. This is valid for any particular r and s if the store is large so that f_r and f_s are large.

The expected increment to $\tilde{f}$ following a single random release is denoted by $\Delta^{-}\tilde{f}$ and given by

$$\Delta^{-}\tilde{f} = p_0 \sum_{i=1}^{R} b_i a^i + p_1 \sum_{i=1}^{R} b_i \sum_{r=1} \phi_r (a^{i+r} - a^r) + p_2 \sum_{i=1}^{R} b_i \sum_{r=1} \phi_r \sum_{s=1} \phi_s (a^{i+r+s} - a^r - a^s) \quad (2.37)$$

$$= p_0 \tilde{b} + p_1(\tilde{b} - 1)\tilde{\phi} + p_2(\tilde{b}\tilde{\phi} - 2)\tilde{\phi} \quad (2.38)$$

which may be further simplified using the expression above for p_1,

$$\Delta^- \tilde{f} = (p_0 + p_1\tilde{\phi} + p_2\tilde{\phi}^2)\tilde{b} - \frac{2}{k}\tilde{\phi} \quad (2.39)$$

2.4 *Equilibrium conditions*

As indicated in Section 2.1, in equilibrium reservations and releases take place at the same mean rate in order to maintain a stable store utilisation. The conditions for equilibrium are thus seen to be

$$\Delta^+\tilde{f} + \Delta^-\tilde{f} = 0 \quad (2.40)$$

2.5 *The Fifty Percent Rule*

The equation above, in generating function form, is to be regarded as an identity in a hence as a shorthand for the equations corresponding to each power of a in a series expansion. By evaluating the generating functions at $a = 1$, these equations are summed, and the result expresses the constancy of the number of free blocks.

i.e.

$$\Delta^+ F + \Delta^- F = 0 \quad (2.41)$$

since

$$\tilde{f}(1) = F \quad (2.42)$$

Substituting $a = 1$, and using $\tilde{b}(1) = \tilde{\phi}(1) = 1$, yields

$$\Delta^+ F = -\sum_{i=1}^{R} \frac{b_i\phi_i}{\gamma_i} \quad (2.43)$$

and

$$\Delta^- F = 1 - \frac{2}{k} \quad (2.44)$$

so that

$$F/B = \tfrac{1}{2}p \quad (2.45)$$

where

$$p = 1 - \sum_{i=1}^{R} \frac{b_i\phi_i}{\gamma_i} \quad (2.46)$$

This is Knuth's Fifty Percent Rule with p being the probability that an arbitrary request will be satisfied by using an oversize free block. It may be noted that this imposes the bound

$$B \geqslant 2F \quad (2.47)$$

i.e.

$$k \geqslant 2 \quad (2.48)$$

which is stronger than that of equation (2.22).

3. Unit size requests: a special case

The case in which all requests are for a single store location is of interest both in regard to the results obtained for the free store profile and to the methods for obtaining them.

3.1 *Equilibrium conditions*

The restriction to unit size requests is represented by

$$\tilde{b} = a \quad (3.1)$$

i.e.

$$R = 1, b_1 = 1 \text{ and } b_i = 0 \text{ for } i > 1 \quad (3.2)$$

Substituting into equations (2.14) and (2.39), noting that

$$\gamma_1 = 1 \quad (3.3)$$

and setting

$$\frac{2}{k} = x \quad (3.4)$$

yields

$$\Delta^+\tilde{f} = (a^{-1} - 1)\tilde{\phi} - \phi_1 \quad (3.5)$$

and

$$\Delta^-\tilde{f} = (p_0 + p_1\tilde{\phi} + p_2\tilde{\phi}^2)a - x\tilde{\phi} \quad (3.6)$$

The Fifty Percent Rule here determines that

$$x = p \quad (3.7)$$

where

$$p = 1 - \phi_1 \quad (3.8)$$

so that

$$0 \leqslant x \leqslant 1 \quad (3.9)$$

and hence

$$\infty \geqslant k \geqslant 2 \quad (3.10)$$

Using these relations to eliminate ϕ_1, and equations (2.35) and (2.36) to eliminate p_0 and p_1, the equilibrium conditions (2.40) are obtained in the form

$$a^2p_2(\tilde{\phi} - 1)^2 + (1 - a)(1 - ax)(\tilde{\phi} - 1) + (1 - a)^2 = 0 \quad (3.11)$$

Further simplification results from the substitution

$$\tilde{\phi} - 1 = -(1 - a)\tilde{\chi} \quad (3.12)$$

where

$$\tilde{\chi} = \sum_{r=0}^{\infty} \chi_r a^r \quad (3.13)$$

to yield

$$a^2p_2\tilde{\chi}^2 - (1 - ax)\tilde{\chi} + 1 = 0 \quad (3.14)$$

and

$$\phi_r = \chi_{r-1} - \chi_r, \quad r \geqslant 1 \quad (3.15)$$

In these derivations some care is needed at $a = 1$ since, for example, equation (3.11) collapses at $a = 1$. However $\tilde{\phi}$ and $\tilde{\chi}$ are continuous and so the difficulty is removed.

3.2 *Existence of solutions*

Combining equations (2.2) and (2.5) in the case of unit size requests results in

$$k = \frac{\theta}{1 - \theta}\bar{r} \quad (3.16)$$

or, equivalently,

$$\theta = \frac{k}{k + \bar{r}} \quad (3.17)$$

where

$$\bar{r} = \sum_{r=1}^{L} r\phi_r = \tilde{\phi}'(1) = \tilde{\chi}(1) \quad (3.18)$$

Now the condition for the existence of a real solution $\tilde{\chi}$ of equation (3.14) is

$$(1 - ax)^2 - 4a^2p_2 \geqslant 0 \quad (3.19)$$

For given x, $0 \leqslant x \leqslant 1$, this condition is satisfied for sufficiently small a. However, equation (3.18) shows that a solution is required for $a = 1$ and it is therefore necessary that

$$(1 - x)^2 - 4p_2 \geqslant 0 \quad (3.20)$$

This condition, together with equations (2.34) and (3.4), determine that solutions exist within the range

$$0{\cdot}3737 \leqslant \theta \leqslant 1 \quad (3.21)$$

with corresponding ranges for the other major parameters as follows

$$3{\cdot}1936 \leqslant k \leqslant \infty \quad (3.22)$$

$$0{\cdot}6263 \geqslant x \geqslant 0 \quad (3.23)$$

$$0{\cdot}0349 \geqslant p_2 \geqslant 0 \quad (3.24)$$

$$5{\cdot}3513 \geqslant \bar{r} \geqslant 1 \quad (3.25)$$

It would appear that no stable solution exists for less than 37% store utilisation.

3.3 *Approximation for high utilisation*

As k increases, equations (2.34) and (3.4) indicate that both p_2 and x decrease, the former much more rapidly. At small x

values, p_2 may therefore be neglected, and equation (3.14) yields

$$\tilde{\chi} = (1 - ax)^{-1} = \sum_{r=0}^{\infty} a^r x^r \qquad (3.26)$$

leading to

$$\phi_r = (1 - x)x^{r-1} \qquad (3.27)$$

Thus the proportion of free blocks with given length decreases with length in simple geometrical progression. The mean free block length in this approximation is given by equation (3.16) as

$$\bar{r} = \tilde{\chi}(1) = (1 - x)^{-1} \qquad (3.28)$$

In this approximation the store utilisation is given by equation (3.17) as

$$\theta = 1 - \frac{x}{2 - x} \qquad (3.29)$$

3.4 *The complete solution*

In this section the solution $\tilde{\chi}$ of equation (3.14) is derived quite straightforwardly as a series expansion in powers of a. The expression for $\tilde{\chi}$ is

$$\tilde{\chi} = \frac{1 - ax}{2a^2 p_2}\left\{1 - \left[1 - \frac{4a^2 p_2}{(1 - ax)^2}\right]^{\frac{1}{2}}\right\} \qquad (3.30)$$

where the negative square root is chosen to ensure that $\tilde{X}$ is bounded as $p_2 \to 0$ and so recovers the approximation of the previous section.

Expanding first in powers of p_2 gives

$$\tilde{\chi} = -\sum_{n=0}^{\infty} \frac{1 - ax}{2a^2 p_2}\left(\frac{-4a^2 p_2}{(1 - ax)^2}\right)^{n+1} \times \frac{\frac{1}{2} \cdot -\frac{1}{2} \cdot -\frac{3}{2} \cdot \dots \cdot \frac{1 - 2n}{2}}{1 \, . \, 2 \, . \, 3 \dots\dots (n + 1)} \qquad (3.31)$$

$$= \sum_{n=0}^{\infty} \frac{a^{2n} p_2^n}{(1 - ax)^{2n+1}} \, \frac{(2n)!}{n!(n + 1)!} \qquad (3.32)$$

and then, in powers of a,

$$\tilde{\chi} = \sum_{n=0}^{\infty} a^{2n} p_2^n \frac{(2n)!}{n!(n + 1)!} \sum_{u=0}^{\infty} a^u x^u \frac{(2n + u)!}{(2n)!u!} \qquad (3.33)$$

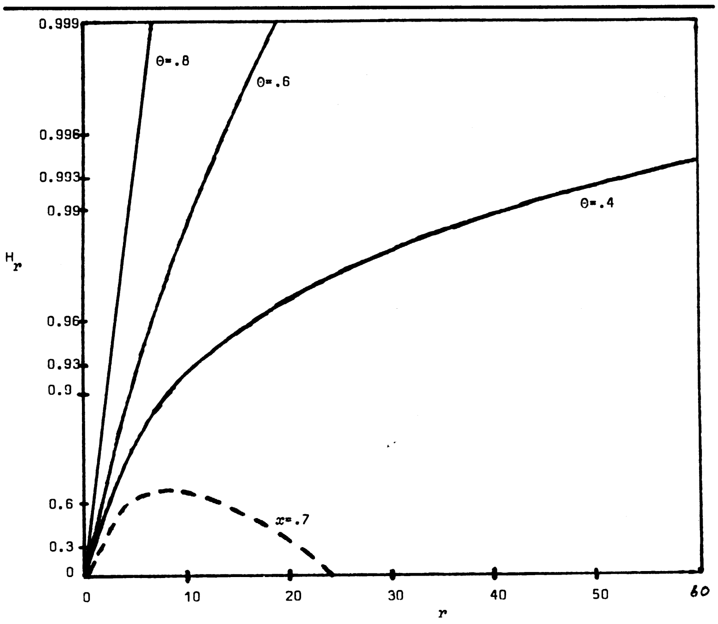

Fig. 1 Variation of H_r with r at different utilisations

Table 1 Parameter sets for plotting H_r curves

θ	x	k	p_2	$\bar{r}$
0·9	0·1818	11·0000	0·000004	1·2222
0·8	0·3328	6·0100	0·0011	1·5025
0·7	0·4549	4·3964	0·0076	1·8842
0·6	0·5450	3·6695	0·0189	2·4463
0·5	0·6015	3·3251	0·0294	3·3251
0·4	0·6252	3·1989	0·0347	4·7984
0·3737	0·6263	3·1936	0·0349	5·3513
–	0·7	2·8571	0·0546	–

$$= \sum_{n=0}^{\infty} \sum_{u=0}^{\infty} \frac{(2n + u)!}{n!(n + 1)!u!} a^{2n+u} p_2^n x^u \qquad (3.34)$$

Hence, putting $2n + u = r$,

$$\chi_r = \sum_{n=0}^{[\frac{r}{2}]} \frac{r!}{n!(n + 1)!(r - 2n)!} p_2^n x^{r-2n} \qquad (3.35)$$

where $\left[\frac{r}{2}\right]$ denotes the largest integer not exceeding $\frac{r}{2}$.

Detailed examination suggests evaluation of χ_r by a form of nested multiplication.

Case 1: r odd. Put $r = 2s + 1$, $s \geqslant 0$.

$$\chi_r = x^r + \frac{r(r - 1)}{1!2!} p_2 x^{r-2} + \frac{r(r - 1)(r - 2)(r - 3)}{2!\,3!} p_2^2 x^{r-4} + \dots + \frac{r(r - 1)\dots 3\,.\,2}{s!\,(s + 1)!} p_2^s x \qquad (3.36)$$

$$= x^r\left(1 + \frac{r(r - 1)}{1.2}\frac{p_2}{x^2}\left(\dots\left(1 + \frac{5.4}{(s-1)s}\frac{p_2}{x^2} \times \left(1 + \frac{3.2}{s(s+1)}\frac{p_2}{x^2}\right)\right)\dots\right)\right) \qquad (3.37)$$

Case 2: r even. Put $r = 2s$, $s \geqslant 0$.

$$\chi_r = x^r + \frac{r(r - 1)}{1!\,2!} p_2 x^{r-2} + \frac{r(r - 1)(r - 2)(r - 3)}{2!\,3!} p_2^2 x^{r-4} + \dots + \frac{r(r - 1)\dots 2.1}{s!\,(s + 1)!} p_2^s \qquad (3.38)$$

$$= x^r\left(1 + \frac{r(r - 1)}{1.2}\frac{p_2}{x^2}\left(\dots\left(1 + \frac{4.3}{(s - 1)s}\frac{p_2}{x^2} \times \left(1 + \frac{2.1}{s(s + 1)}\frac{p_2}{x^2}\right)\right)\dots\right)\right) \qquad (3.39)$$

Possibly the clearest representation of the free store profile is in terms of the cumulative distribution function

$$H_r = \sum_{u=1}^{r} \phi_u \qquad (3.40)$$

$$= \chi_0 - \chi_r = 1 - \chi_r \qquad (3.41)$$

Here H_r is the proportion of free blocks whose length is r or less. **Fig. 1** shows the variation of H_r with r for a range of θ values. For each θ, the corresponding value of x, and hence k and p_2, was obtained by interpolation in equation (3.17) using the expression

$$\bar{r} = \frac{1 - x}{2p_2}\left\{1 - \left[1 - \frac{4p_2}{(1 - x)^2}\right]^{\frac{1}{2}}\right\} \qquad (3.42)$$

derived from equations (3.18, 3.30). A logarithmic scale is used, with the vertical axis corresponding to $-\log(1 - H_r)$. Relevant parameter values are listed in **Table 1**. The linearity of the curve for $\theta = 0{\cdot}8$ illustrates the high utilisation approximation of Section 3.3. At lower utilisations the curves wilt progressively. The broken curve shows the result of applying equations (3.37, 3.39) outside their range of validity with $x = 0{\cdot}7$—i.e. above the critical value 0·6263 of equation (3.23). The existence of a turning point and subsequent negative gradient is behaviour clearly incompatible with that of a respectable cumulative distribution function.

4. Discussion

On the personal level, quite the most satisfying part of the work reported here was the Eureka experience of stumbling, after some weeks of ineffectual groping, upon the realisation that generating functions were the key to the problem. Not only did this technique permit a compact and elegant formulation of the equations, but, in the case of unit length requests, their solution then required merely schoolboy algebra. Although the introduction of this technique into the study of storage allocation may possibly be the most important contribution of this paper, the prediction of a stability threshold at 37% store utilisation is undoubtedly the most intriguing. Any speculation now, in advance of further investigation, is likely to be a source of embarrassment in the future but some comment is clearly called for.

The theoretical model has two distinct components. The major part is summed up in equations (2.14, 2.39) which lead to equilibrium conditions in terms of the parameters p_2 and k. The second is somewhat separate and is contained in Section 2.3.2. This establishes a relationship between p_2 and k so that the free store profiles $\tilde{\phi}$ form a one-parameter family. It is a simple matter to transform between k, the block utilisation ratio, as parameter and θ, the store utilisation. As was noted at the time, the argument of Section 2.3.2, though logically sound, is somewhat unsatisfying. The difficulty is a familiar one in that a slightly different but equally plausible interpretation of randomness might yield a different relationship. Hopefully the present information-theoretic argument will be wholly replaced by a specific analysis based upon the problem situation. This will in all probability lead to a different equation (2.34) and hence either to a revised value for the stability threshold or, if equation (3.19) is then satisfied throughout $0 \leqslant \theta \leqslant 1$, to a disappearance of the phenomenon entirely.

The availability of a closed analytical solution for unit sized requests simplifies further investigation of the sensitivity of $\tilde{\phi}$ to the relationship between p_2 and k. Results from simulations are required for comparison and it is hoped to report conclusions in due course. Of particular interest will be empirical results from the critical region of 37% store utilisation.

Also left for future investigation is the solution of the equilibrium equations for general request distributions. This represents a non-trivial extension on which there appears to be little published data from simulations apart from hints that free store patterns are relatively insensitive to the request distribution.

In the shadow of all this unfinished business, some may consider the present publication premature. Others will agree that these preliminary findings are of interest and may stimulate parallel activity.

5. Acknowledgements

Most of the ideas that have been described were thought through during a recent period of convalescence. I gratefully acknowledge both the expertise of the doctors whose good offices culminated in securing me this pseudo-sabbatical, and the willingness of my departmental colleagues to undertake my normal duties in the same period. In the preparation of the final manuscript I have been greatly helped by the insight and advice of the referee, and I offer him my sincere thanks.

References

CAMPBELL, J. A. (1971). A note on an optimal-fit method for dynamic allocation of storage, *The Computer Journal*, Vol. 14, pp. 7-9.
CRANSTON, B. and THOMAS, R. (1975). A simplified recombination scheme for the Fibonacci buddy system, *CACM*, Vol. 18, pp. 331-332.
GELENBE, E., BOEKHORST, J. C. A. and KESSELS, J. L. W. (1973). Minimising wasted space in partitioned segmentation, *CACM*, Vol. 16, pp. 343-349.
HINDS, J. A. (1975). An algorithm for locating adjacent storage blocks in the buddy system, *CACM*, Vol. 18, pp. 221-222.
HIRSCHBERG, D. S. (1973). A class of dynamic memory allocation algorithms, *CACM*, Vol. 16, pp. 615-618.
ISODA, S., GOTO, E. and KIMURA, I. (1971). An efficient bit-table technique for dynamic storage allocation of 2^n-word blocks, *CACM*, Vol. 14, pp. 589-592.
KNOWLTON, K. C. (1965). A fast storage allocator, *CACM*, Vol. 8, pp. 623-625.
KNUTH, D. E. (1968). *The art of computer programming*, Vol. 1: Fundamental algorithms, (1st edn. Addison-Wesley) §2.5 Dynamic storage allocation, pp. 435-455.
NIELSEN, N. R. (1977). Dynamic memory allocation in computer simulation, *CACM*, Vol. 20, pp. 864-873.
PURDOM, P. W. and STIGLER, S. M. (1970). Statistical properties of the buddy system, *JACM*, Vol. 17, pp. 683-697.
RANDELL, B. (1969). A note on storage fragmentation and program segmentation, *CACM*, Vol. 12, pp. 365-369, 372.
SHEN, K. K. and PETERSON, J. L. (1974, 1975). A weighted buddy method for dynamic storage allocation, *CACM*, Vol. 17, pp. 558-562 and Vol. 18, p. 202.
SHORE, J. E. (1977). Anomalous behaviour of the fifty-percent rule in dynamic memory allocation, *CACM*, Vol. 20, pp. 812-820.
WOLMAN, E. (1965). A fixed optimum cell-size for records of various lengths, *JACM*, Vol. 12, pp. 53-70.

Book review

The Process of Question Answering, by Wendy G. Lehnert, 1978; 278 pages. (*Wiley*, £11·75)

This book seeks to answer the sixty-four dollar question: What is a question and what is an answer? Although a laudable attempt is made to present a systematic and logical flowchart for this common and complex human activity, the book suffers considerably from that tiresome American scholarly affliction, a combination of indigestible jargon and fourth-form humour. An example of the first runs as follows: 'The processes which assign the proper conceptual category to a question must be sensitive to the context in which that question is asked. Questions cannot be correctly understood by processes that do not consider contextual factors.' In other words, a question like 'Are you a little queer?' depends on its context if it is to generate the correct response. An example of the humour is one subsection heading which goes: 'Smart heuristics know when to quit'. Another subsection heading, which falls between these stools, unintentionally bursts into verse: 'You can't always expect to find / Exactly what you had in mind'. No wonder one of the programs exploited in the production of this book is called QUALM.

R. W. LAST (Hull)