G. S. Shedler
D. R. Slutz

Derivation of Miss Ratios for Merged Access Streams

Abstract: An access stream is the sequence of storage accesses made by an executing program; a merged stream results from the multiprogramming of a number of individual access streams. Assuming that LRU (least recently used) miss ratio functions for individual streams are known, we consider the problem of predicting the LRU miss ratio function for merged streams. Each access stream is modeled as a sequence of independent, identically distributed LRU stack distances which evolves in time as a Poisson process and the merged stream is taken to be the superposition of these processes. For an arbitrary number of such streams, a closed form expression for the expected miss ratio function is obtained.

1. Introduction

Performance evaluation of computer systems often involves the study of a formal (mathematical) model of some portion of the system. For meaningful storage system evaluation it is necessary to incorporate the referencing behavior of executing programs into the model. This referencing behavior is conceptualized as an access stream – a sequence of requests for data access. Often a distinction is made between the specification of the workload characteristics and the structure of the system, yielding a "workload model" as a driver of a "storage system model." The apparent difficulty of obtaining an adequate mathematical description of the time-varying characteristics of a storage system workload has led to the practice of using either actual access streams obtained by tracing executing programs or over-simplified workload models. An example of the latter is the *miss ratio function* for an access stream which tabulates the (long-run) fraction of references made to the second level of a two-level memory hierarchy as a function of the capacity of the first level; this function depends implicitly on many other parameters such as block size, replacement algorithm, etc. [1].

Experience with the use of miss ratio functions, particularly in cache-main memory hierarchy design, has shown them to be useful representations of workload characteristics when the referencing activity of only a single stream is considered and when only the number of misses (not their occurrences in time) is important for determining performance.

The relative simplicity of a miss ratio function inspires its use for more complex workloads, such as for a concurrent set of executing programs. Such multistream environments exist in current systems that utilize multiprogramming or multiprocessing but use a single memory system. In this case the stream of accesses to a storage device or subsystem is a composite stream arising from several access streams, and the miss ratio function depends not only on the individual streams but also on the mechanisms that affect the mixing, e.g., the process scheduler and the device timing parameters. The miss ratio function is also affected if the separate streams access common information, but this is not considered here.

This paper is concerned with the prediction of miss ratio functions for multistream environments from the miss ratio functions of the individual streams. The proposed basis for the prediction is the derivation of the miss ratio function for a stochastic merged stream. Each access stream is taken to be a stochastic point process (i.e., a series of events) evolving in time and the superposition of these independent processes provides the mixing or merging mechanism. For the case of $J \geq 2$ streams, each represented as independent, identically distributed (i.i.d.) least recently used (LRU) stack distances (cf., [2]) evolving in time as a Poisson process, a closed form solution for the expected miss ratio function of the merged stream is obtained. (An alternative method for predicting miss ratios for multistream environments has been studied by MacDonald et al. [3].)

- *Notation*

$\mu_i(A_j)$	miss ratio function for stream A_j, evaluated at capacity $i-1$
$N(x:y)$	number of occurrences of symbol x in set or sequence y
$M_k(c)$	fraction of misses in first k references for the merged stream for capacity c

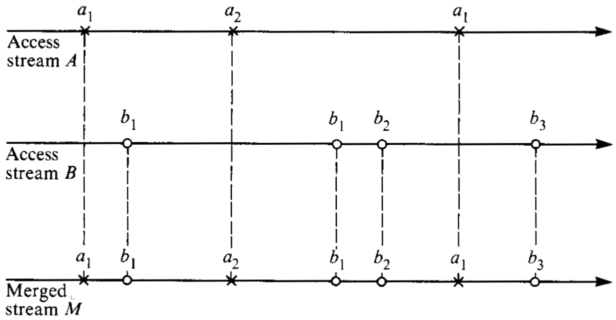

Figure 1 Merging of two access streams by superposition.

$M(c)$	long-run expected miss ratio function for the merged stream
$r(A)$	reference sequence for stream A
$s(A)$	(well formed) stack sequence for stream A
$d(A)$	(well formed) distance sequence for stream A
$r(A) \rightarrow s(A)$	stack sequence $s(A)$ is associated with reference sequence $r(A)$
$r(A) \leftrightarrow s(A)$	reference sequence $r(A)$ and stack sequence $s(A)$ are mutually associated
$r(A) \rightarrow d(A)$	distance sequence $d(A)$ is associated with reference sequence $r(A)$
$r(A) \sim r(A')$	equivalent reference sequences $r(A)$ and $r(A')$
$\overleftarrow{r}(A)$	reverse of sequence $r(A)$
$D^*(A)$	distance process for stream A
$d^*(A)$	deterministic distance sequence or realization of distance process $D^*(A)$
$d'(A)$	twin of distance sequence $d(A)$
$d^{*\prime}(A)$	twin of realization $d^*(A)$
$p^*(M)$	merged distance sequence
$p(M)$	(well formed) merged distance sequence
$r(M)$	merged reference sequence
$s(M)$	(well formed) merged type stack sequence
$d(M)$	M-distance sequence or well formed distance sequence associated with a merged reference sequence
$t(M, c)$	(well formed) merged type stack sequence
$P^*(M)$	merged distance process
$T(M, c)$	merged type stack process
$Z(M, c)$	set of recurrent states for merged type stack process
$\Psi(z)$	set of finite length merged reference sequences that yield type stack z
$\Omega(z)$	set of finite length, well formed, merged distance sequences associated with the reverses of the sequences in $\Psi(z)$
$H(z)$	Markov chain based on $P^*(M)$
$Q(k_1, \cdots, k_J; c)$	long-run probability that a merged type stack for capacity c contains k_j entries from stream A_j, $1 \leq j \leq J$

2. Definition of the multistream environment

An *access stream A* is taken to be a sequence of references occurring in continuous time to data items called blocks. If a reference is viewed as an event, an access stream is a (multivariate) stochastic point process [4]. Thus, there is an increasing sequence of epochs of times-to-events $\{t_k(A)\}$, $t_k(A) < t_{k+1}(A)$, $k = 1, 2, \cdots\}$, and at each epoch $t_k(A)$ a reference $r_k(A)$ is made to one of a finite set of blocks $\mathscr{A} = \{a_1, \cdots, a_m\}$. *It is assumed throughout that the sets of blocks referenced by separate access streams are mutually disjoint.* The sequence $\{r_k(A); k = 1, 2, \cdots\}$ is called a *reference sequence* and is sufficient to determine any miss ratio function for the access stream. (In this paper only the LRU miss ratio function is considered.)

The mechanism considered here for mixing several access streams to create a multistream environment is the superposition of the individual (point process) access streams. The theory of the superposition of point processes is discussed by Çinlar [5]. Figure 1 illustrates this type of merging for two access streams A and B. The resulting composite reference sequence is called a *merged reference sequence* and determines the miss ratio function of the multistream environment. Observe that the reference sequences of the individual access streams are primarily determined by the respective executing programs, but the times of reference (and thus the merged reference sequence) are primarily determined by the system environment.

Determination of the LRU miss ratio function for a given reference sequence, either for an individual stream or for a merged stream, is given next. For a finite reference sequence $r_1(A), \cdots, r_L(A)$ of length L, an LRU stack distance of k, $1 \leq k \leq m$, is associated with $r_i(A)$ if and only if $r_i(A) = a_l$, and for the largest $j < i$, where $r_j(A) = a_l$, exactly k distinct blocks are referenced by $r_j(A), \cdots, r_i(A)$. If $r_i(A) = a_l$ is the first reference to block a_l, then the stack distance is undefined. Also, let $Q(k)$, $1 \leq k \leq m$, be the number of associated stack distances equal to k in the finite reference sequence.

Recall that an LRU stack distance equal to k signifies that a miss would occur to a first-level memory if and only if it is of capacity in blocks $c > k$; then the LRU miss ratio function is given by $1 - L^{-1}\Sigma_{k=1}^{c} Q(k)$ for $1 \leq c \leq m - 1$, and by $1 - L^{-1}\Sigma_{k=1}^{m} Q(k)$ for $c \geq m$.

For a stochastic reference sequence generated by i.i.d. LRU stack distances (with distribution $\{\alpha_i\}$ as given in following sections), the long-run expected fraction of misses for first-level capacity c approaches $\Sigma_{i=c+1}^{m}\alpha_i$, for $c = 1, \cdots, m-1$, and approaches 0 for $c \geq m$. Thus, if the miss ratio function for an actual reference sequence is known, probabilities $\{\alpha_i\}$ can be easily determined such that the expected fraction of misses for capacity c for the (i.i.d.) distance-generated stochastic reference sequence approaches the miss ratio function as the number of generated references becomes large.

3. Probabilistic assumptions and statement of main result

The derivation of the expected LRU miss ratio function formed by merging $J \geq 2$ individual access streams $A_1, \cdots, A_J$ is made under the following probabilistic assumptions.

1. Access streams $A_1, A_2, \cdots, A_J$ are independent.
2. Access stream A_j evolves as a Poisson process of rate $\lambda_j > 0$, i.e., $\{t_k(A_j) - t_{k-1}(A_j)\}$ is a sequence of independent and identically distributed random variables having an exponential distribution with rate parameter λ_j; i.e.,

   $$\Pr\{t_k(A_j) - t_{k-1}(A_j) \leq t\} = 1 - e^{-\lambda_j t}, \; t \geq 0.$$
3. The LRU stack distance sequence $\{D_k^*(A_j)\}$ for access stream A_j is a sequence of i.i.d. random variables with distribution $\{\alpha_i(A_j)\}$, i.e., $\Pr\{D_k^*(A_j) = i\} = \alpha_i(A_j)$, $1 \leq i \leq m_j$.

It follows from the Poisson assumptions that the stream identity of a reference in a merged sequence is determined by an independent trials process, i.e., for all i the probability that the ith reference in the merged stream is from stream A_j is $\lambda_j/\Sigma_{k=1}^{J}\lambda_k$. For $j = 1, \cdots, J$ let

$$\mu_i(A_j) = \begin{cases} \sum_{k=i}^{m_j} \alpha_k(A_j), & 1 \leq i \leq m_j, \\ 0, & i > m_j, \end{cases}$$

denote the miss ratio function for the separate streams and let $\delta_j = \lambda_j/\Sigma_{k=1}^{J}\lambda_k$. Also, for x a symbol, c a nonnegative integer, and y a set or a sequence, let $N(x:y)$ denote the number of occurrences of x in y and let $\{y\}^c$ denote all c-tuples over y.

Theorem 3.1

Let $A_1, \cdots, A_J$ be access streams referencing mutually disjoint sets of blocks and satisfying the probabilistic assumptions above. For a first-level capacity c, denote by $M_k(c)$ the fraction of misses in the first k references for the merged stream. Then the long-run expected miss ratio function $M(c) = \lim_{k\to\infty} \mathrm{E}\{M_k(c)\}$ is given by

$$M(c) = \sum_{z \in Z(M,c)} \prod_{j=1}^{c} G(z,j) \sum_{i=1}^{J} \delta_i \mu_{N(A_i:z)+1}(A_i),$$

$$c = 1, \cdots, \sum_{j=1}^{J} m_j,$$

where $Z(M,c) = \{z; z \in \{A_1 \cup A_2 \cup \cdots \cup A_J\}^c, N(A_j:z) \leq m_j, i = 1, \cdots, J\}$ and for each $z = (z_1, \cdots, z_c) \in Z(M,c)$,

$$G(z,j) = \frac{\sum_{i=1}^{J} \Delta_{ij}\delta_i \mu_{N(A_i:z_1,\cdots,z_j)}(A_i)}{\sum_{i=1}^{J} \delta_i \mu_{N(A_i:z_1,\cdots,z_j)+1-\Delta_{ij}}(A_i)}, \; j = 1, \cdots, c,$$

where

$$\Delta_{ij} = \begin{cases} 1, & \text{if } z_j = A_i, \\ 0, & \text{otherwise.} \end{cases}$$

This theorem is the main result of the paper. The computation of $M(c)$ using Theorem 3.1 requires time of the order of J^c. Since direct application is impractical, in Section 7 a second expression for $M(c)$ is given that requires computational time of the order of c^J.

In the next three sections the theorem is proved for the case $J = 2$ and the extension to $J > 2$ is outlined. Section 4 introduces precise notions of reference, stack, and distance sequences. This is necessary because finite length sequences are considered during which the LRU stack (initially empty) undergoes a "filling up" phase. Notions of reverse sequences, used in the proof to follow, are also introduced. The extensions of these definitions to merged sequences are given in Section 5. The main proof is contained in Section 6 and it involves viewing the sequence of LRU stacks for the merged reference string as a Markov chain. The basic idea is that a particular stack z will exist at time t if and only if the reverse of the merged reference sequence (actually the merged distance sequence) up to time t has a particular property. This property is related to the limiting distribution of another Markov chain $H(z)$ that is easily calculated.

4. Preliminaries

For notational convenience in treating the case of $J = 2$ streams, the two streams are denoted by A and B, the respective disjoint sets of blocks referenced by $\mathscr{A} = \{a_1, \cdots, a_m\}$ and $\mathscr{B} = \{b_1, \cdots, b_n\}$, and the respective LRU stack distance probability distributions by $\{\alpha_i\}$ and $\{\beta_i\}$. The independent probability of referencing stream A, equal to $\lambda_A/(\lambda_A + \lambda_B)$, where λ_A and λ_B are the Poisson rate parameters, is denoted by δ.

In this section definitions are given (along with consequent properties) of "well formed" LRU stack and stack distance sequences that can be associated with a reference sequence. The well formed stack distance sequence is related to the usual stack distance sequence

[1] and is used subsequently in the derivation of the miss ratio function for merged reference sequences.

For access stream A denote the associated reference sequence by

$$r(A) = \{r_j(A); j \geq 1\}, \text{ where } r_j(A) \in \mathscr{A}.$$

Also, denote an *LRU stack sequence* (or stack sequence) for access stream A by $s(A) = \{s_j(A); j \geq 0\}$, where each $s_j(A)$ is an ordered subset of $\mathscr{A}$. Thus, for $j \geq 0$, $s_j(A)$ has the form $s_j(A) = [s_{j,1}(A), \cdots, s_{j,\gamma_j}(A)]$, where $s_{j,k}(A) \in \mathscr{A}$ and $s_{j,k}(A) \neq s_{j,l}(A)$, for $k \neq l$ and $1 \leq k \leq \gamma_j$. The quantity γ_j, $0 \leq \gamma_j \leq m$, is the *size* of $s_j(A)$. If $\gamma_j = 0$, $s_j(A)$ is the empty set denoted by $\varnothing$. Certain stack sequences are associated with reference sequences as follows.

Definition 4.1 Given a reference sequence $r(A)$, an *associated* stack sequence $s(A) = \{s_j(A); j \geq 0\}$ denoted by $r(A) \rightarrow s(A)$ is defined recursively for $j \geq 0$ as follows:

1. $j = 0$: $s_0(\mathrm{A}) = \varnothing$, $\gamma_0 = 0$.
2. For $j > 0$, given $r_{j+1}(A)$ and $s_j(A)$:
   a. if for some k, $1 \leq k \leq \gamma_j$, $r_{j+1}(A) = s_{j,k}(A)$ then
   $$s_{j+1}(A) = [s_{j,k}(A), s_{j,1}(A), \cdots s_{j,k-1}(A), s_{j,k+1}(A), \cdots, s_{j,\gamma_j}(A)]$$
   and $\gamma_{j+1} = \gamma_j$.
   b. otherwise,
   $$s_{j+1}(A) = [r_{j+1}(A), s_j(A)] \text{ and } \gamma_{j+1} = \gamma_j + 1.$$

Definition 4.2 Given a stack sequence $s(A)$, a reference sequence $r(A)$ is *associated* with $s(A)$, denoted by $s(A) \rightarrow r(A)$, if $r(A) \rightarrow s(A)$. If at least one $r(A)$ is associated with $s(A)$, then $s(A)$ is said to be *well formed.*

The observation in Definition 4.1 that for $r(A) \rightarrow s(A)$, $s_{j,1}(A) = r_j(A)$ for all $j \geq 1$ leads to the following proposition.

Proposition 4.3 Given $r(A)$, there is a unique $s(A)$ such that $r(A) \rightarrow s(A)$. Given $s(A)$ well formed, there is a unique $r(A)$ such that $s(A) \rightarrow r(A)$.

It follows that the set of reference and well formed stack sequences can be partitioned into mutually associated pairs (denoted by $r(A) \leftrightarrow s(A)$). These associations are next extended to include distance sequences.

For access stream A denote an *LRU distance sequence* (or distance sequence) by $d(A) = \{d_j(A); j \geq 1\}$ where $d_j(A) \in \{1, \cdots, m\} \cup \{\bar{1}, \cdots, \bar{m}\}$. Distances of the form $\bar{j}$ correspond to the first reference to a block in a reference sequence.

Definition 4.4 Given a reference sequence $r(A)$ and a well formed stack sequence $s(A)$ where $r(A) \leftrightarrow s(A)$, an *associated distance sequence* $d(A) = \{d_j(A); j \geq 1\}$, denoted by $r(A) \rightarrow d(A)$ or $s(A) \rightarrow d(A)$, is defined recursively for $j \geq 1$ as follows:

1. If for some k, $1 \leq k \leq \gamma_{j-1}$, $r_j(A) = s_{j-1,k}(A)$, then $d_j(A) = k$.
2. If for all k, $1 \leq k \leq \gamma_{j-1}$, $r_j(A) \neq s_{j-1,k}(A)$, then $d_j(A) = \overline{\gamma_{j-1} + 1}$.

Definition 4.5 A distance sequence $d(A)$ is said to be a *well formed distance sequence* if the following conditions are satisfied:

1. Denote by $K = \{k_l; l \geq 1\}$ where $k_l < k_{l+1}$ the sequence of indices of $d(A)$ such that $d_{k_l}(A) = \bar{j}$ for some j. Then $d_{k_l}(A) = \bar{l}$ for $l = 1, 2, \cdots$.
2. For all $k \geq 1$, if $d_k(A) = j$, then there exists $l < k$ such that $d_l(A) = \bar{j}$.

Table 1 illustrates these definitions for a particular reference sequence $r(A)$ over $\mathscr{A} = \{a_1, \cdots, a_5\}$.

Proposition 4.6 Given $r(A)$ (or $s(A)$ where $r(A) \leftrightarrow s(A)$), there is a unique $d(A)$ such that $r(A) \rightarrow d(A)$ and $d(A)$ is well formed. Given $d(A)$ well formed, there exists a reference sequence $r(A)$ (not necessarily unique) such that $r(A) \rightarrow d(A)$. This association is denoted $d(A) \rightarrow r(A)$.

The first statement in Proposition 4.6 follows from Definition 4.4 and the second statement can be derived by an inductive proof in which an associated reference sequence is constructed. To see that the association $d(A) \rightarrow r(A)$ is not unique, observe in Table 1 that if $r_6(A)$ and $r_{10}(A)$ are both changed from a_4 to a_5, the associated distance sequence is unchanged. Thus, the unique mutual association between a reference sequence and a well formed stack sequence does not directly extend to well formed distance sequences. There is, however, a unique mutual association between well formed distance sequences and classes of "equivalent" reference sequences in which only the namings of blocks differ.

Definition 4.7 Let $r(A)$ and $r(A')$ be two reference sequences defined over sets of blocks $\mathscr{A}$ and $\mathscr{A}'$, respectively. Let $\mathscr{B}$ be the minimal subset of $\mathscr{A}$ such that for all $j \geq 1$, $r_j(A) = a_i$ implies that $a_i \in \mathscr{B}$ and let $\mathscr{B}' \subseteq \mathscr{A}'$ be similarly defined for $r(A')$. Then reference sequences $r(A)$ and $r(A')$ are said to be *equivalent* (denoted $r(A) \sim r(A')$) if and only if there exists a one-to-one mapping $g \in \mathscr{B} \times \mathscr{B}'$ such that $r_j(A) = g(r_j(A'))$ for all $j \geq 1$.

This notion of equivalence is extended to stack sequences and it is observed that for two well formed stack sequences $s(A)$ and $s(A')$, the stack sequences are equivalent (denoted $s(A) \sim s(A')$) if and only if $r(A) \sim r(A')$, where $s(A) \leftrightarrow r(A)$ and $s(A') \leftrightarrow r(A')$.

Table 1 Examples of well formed sequences.

Sequence		j: 0	1	2	3	4	5	6	7	8	9	10
Reference	$r_j(A)$		a_1	a_2	a_1	a_3	a_1	a_4	a_2	a_2	a_1	a_4
Well formed stack	$s_j(A)$	$\varnothing$	a_1	a_2 a_1	a_1 a_2	a_3 a_1 a_2	a_1 a_3 a_2	a_4 a_1 a_3 a_2	a_2 a_4 a_1 a_3	a_2 a_4 a_1 a_3	a_1 a_2 a_4 a_3	a_4 a_1 a_2 a_3
Size	γ_j	0	1	2	2	3	3	4	4	4	4	4
Well formed distance	$d_j(A)$		$\bar{1}$	$\bar{2}$	2	$\bar{3}$	2	$\bar{4}$	4	1	3	3

Proposition 4.8 Consider two reference sequences $r(A)$ and $r(A')$ and their associated well formed distance sequences, $r(A) \rightarrow d(A)$ and $r(A') \rightarrow d(A')$. Then $d(A) = d(A')$ if and only if $r(A) \sim r(A')$.

To see this, suppose first that $r(A) \sim r(A')$. It is easily shown that $d(A) = d(A')$ by induction on the index of the reference sequence. Conversely, suppose that $r(A)$ and $r(A')$ are not equivalent. For k as large as possible, choose $r(A'')$ defined over the set of block $\mathscr{A}$ and equivalent to $r(A')$ such that $r_j(A'') = r_j(A)$, $1 \leq j \leq k$. It can easily be shown that $d_{k+1}(A'') \neq d_{k+1}(A)$ and thus $d(A') = d(A'') \neq d(A)$.

From Propositions 4.6 and 4.8, it follows that the product space of equivalent classes of reference sequences and well formed stack and distance sequences can be partitioned into triples of mutually associated sequences. In the sequel, equivalent reference sequences are distinguished only when necessary; thus, $r(A)$ is often referred to as "the" reference sequence associated with $d(A)$.

This discussion provides a framework within which the miss ratio function for merged stochastic reference sequences generated by (stochastically) independent LRU distances can be determined. Recall that for stream A and nonnegative integer m, the LRU distance process is a sequence $D^*(A) = \{D_j^*(A); j \geq 1\}$ of i.i.d. nonnegative integer random variables in the range $\{1, \cdots, m\}$, where $\Pr\{D_j^*(A) = i\} = \alpha_i$, $1 \leq i \leq m$. Subscripted upper case letters such as D_j are used to indicate random variables. Subscripted lower case letters such as d_j refer either to realizations of a random variable or to elements of a deterministic sequence. A realization of the stochastic process $D^*(A)$ is interpreted as a distance sequence of the previous discussion.

Definition 4.9 Given a realization $\{d_j^*(A); j \geq 1\}$ of the distance process $D^*(A)$, an *associated* unique well formed distance sequence $d(A) = \{d_j(A); j \geq 1\}$ is defined recursively as follows:

1. For $j = 1$, $d_1(A) = \bar{1}$.
2. For $j \geq 2$, if $d_j^*(A) = k$ and there exists $l < j$ such that $d_l(A) = \bar{k}$, then $d_j(A) = k$.
3. Otherwise, if i is the maximum integer such that for some $l < j$, $d_l(A) = \bar{i}$, then $d_j(A) = \overline{i+1}$.

As an example, $\{d_j^*(A)\} = \{3, 3, 2, 5, 2, 4, 4, 1, 3, 3\}$ yields the distance sequence shown in Table 1. It can now be seen that the use of $d_j(A) = \bar{k}$ represents adding a kth block to the stack. Recall that the memory is assumed to be initially empty. When a distance in the realization larger than the current stack size is observed, it is viewed as a reference to a new (i.e., not previously referenced) block. This convention is made for convenience and does not affect the (long-run) miss ratios.

Some final definitions involving "twin" distance sequences complete the preliminaries. These twin sequences are the key to describing the structure of the merged reference sequences. For any finite length sequence $d = (d_i, \cdots, d_k)$ the *reverse* of d is $(d_k, d_{k-1}, \cdots, d_1)$ and is denoted by $\overleftarrow{d}$.

Definition 4.10 Suppose that $r(A)$ is a finite length reference sequence and that $d(A)$ is the associated well formed distance sequence, $r(A) \rightarrow d(A)$. Then, if $\overleftarrow{r(A)} \rightarrow d'(A)$, the well formed distance sequence $d'(A)$ is called the *twin* of $d(A)$.

The notion of twin sequences is extended to realizations of the stochastic distance process as follows.

Table 2 Examples of twin sequences.

Sequence		j 1	2	3	4	5	6	7	8	9	10	11	12
Distance	$d_j^*(A)$	5	1	3	1	4	2	6	1	3	1	2	4
Well formed distance	$d_j(A)$	$\bar{1}$	1	$\bar{2}$	1	$\bar{3}$	2	$\bar{4}$	1	3	1	2	4
Reference	$r_j(A)$	a_1	a_1	a_2	a_2	a_3	a_2	a_4	a_4	a_3	a_3	a_4	a_1
Reverse of $r(A)$	$\overleftarrow{r_j}(A)$	a_1	a_4	a_3	a_3	a_4	a_4	a_2	a_3	a_2	a_2	a_1	a_1
Twin of $d(A)$	$d_j'(A)$	$\bar{1}$	$\bar{2}$	$\bar{3}$	1	2	1	$\bar{4}$	3	2	1	4	1
Twin of $d^*(A)$	$d_j^{*\prime}(A)$	5	3	4	1	2	1	6	3	2	1	4	1

Definition 4.11 Let $d^*(A) = \{d_j^*(A);\ 1 \le j \le l\}$ be a finite length realization. The *twin* of $d^*\ (A)$, denoted $d^{*\prime}\ (A)$, is defined by the following procedure.

1. Determine $d(A) = \{d_j(A);\ 1 \le j \le l\}$, the well formed distance sequence for $d^*(A)$. Let $\{i_1,\ i_2, \cdots\}$ with $i_l < i_{l+1}$ be the set of indices such that for some k, $d_j(A) = \bar{k}$ and let $\{i_1',\ i_2', \cdots\}$ be similarly defined for the twin $d'(A)$.
2. $d^{*\prime}(A) = \{d_j^{*\prime}(A);\ 1 \le j \le l\}$, the twin of $d^*(A)$, is given by

$$d_j^{*\prime}(A) = \begin{cases} d_j'(A),\ j \neq i_1',\ i_2', \cdots; \\ d_{i_l}^*(A),\ j = i_l',\ l = 1, 2, \cdots. \end{cases}$$

Table 2 illustrates these definitions. The first two lines show a realization of length 12 and its well formed distance sequence $d(A)$. The third line shows one of the reference sequences $r(A)$, where $r(A) \rightarrow d(A)$, and the next line shows the reverse reference sequence $\overleftarrow{r}(A)$. The fifth line shows the twin $d'(A)$ of $d(A)$, i.e., $\overleftarrow{r}(A) \rightarrow d'(A)$, and the last line shows the twin $d^{*\prime}(A)$ of $d^*(A)$. Observe that the index set $\{i_1,\ i_2, \cdots\}$ is $\{1,\ 3,\ 5,\ 7\}$ and $\{d_j^*(A); j = 1,\ 3,\ 5,\ 7\}$ is $\{5,\ 3,\ 4,\ 6\}$. Also, $\{i_1',\ i_2', \cdots\}$ is $\{1,\ 2,\ 3,\ 7\}$ and consequently $\{d_j^{*\prime}(A); j = 1,\ 2,\ 3,\ 7\}$ is $\{5,\ 3,\ 4,\ 6\}$.

This example illustrates some properties of twin distance sequences which are easily shown to hold in general.

Proposition 4.12 Let $d^*(A)$ be a finite realization of a distance process and let $d(A)$ be the corresponding well formed distance sequence. Then

1. the twin $d'(A)$ of $d(A)$ is unique and the twin of $d'(A)$ is $d(A)$. Furthermore, $d'(A)$ is a permutation of $d(A)$.
2. the twin $d^{*\prime}(A)$ of $d^*(A)$ is unique and the twin of $d^{*\prime}(A)$ is $d^*(A)$. Furthermore, $d^{*\prime}(A)$ is a permutation of $d^*(A)$.

5. Merged sequences

In this section the structure of merged reference sequences is investigated. Two access streams A and B are considered initially where the respective sets of blocks accessed, $\mathscr{A} = \{a_1, \cdots, a_m\}$ and $\mathscr{B} = \{b_1, \cdots, b_n\}$, are assumed disjoint ($\mathscr{A} \cap \mathscr{B} = \emptyset$). Recall that an independent LRU stack distance process is associated with each access stream and that the merged stream is obtained by the superposition of the individual streams.

The definitions and concepts of the previous section are extended to merged stack and distance sequences and well formed merged stack and distance sequences in order to describe the structure of merged reference sequences. Additionally, a particular merged reference sequence (corresponding to a realization) is also viewed as a single access stream over the set of blocks $\mathscr{A} \cup \mathscr{B}$ for the purpose of calculating miss ratios.

Definition 5.1 For access streams A and B

1. a *merged distance sequence* $p^*(M)$ is a sequence of pairs $p^*(M) = \{u_j(M),\ d_j^*(u_j(M));\ j \ge 1\}$, where for $j \ge 1$,
   a. $u_j(M) \in \{A,\ B\}$ denotes the *type* of the jth distance;
   b. $d_j^*(u_j(M)) \in \begin{cases} \{1, \cdots, m\} \text{ if } u_j(M) = A; \\ \{1, \cdots, n\} \text{ if } u_j(M) = B. \end{cases}$
2. a *well formed merged distance sequence* $p(M)$ is a sequence of pairs $p(M) = \{u_j(M),\ d_j(u_j(M)); j \ge 1\}$, where for $j \ge 1$, $u_j(M) \in \{A,\ B\}$ and
   a. the sequence $\{d_j(u_j(M));\ u_j(M) = A\}$ is a well formed distance sequence for stream A;
   b. the sequence $\{d_j(u_j(M));\ u_j(M) = B\}$ is a well formed distance sequence for stream B.

A merged distance sequence is thus a bivariate distance sequence with $u_j(M)$ denoting the type (or identity)

Table 3 Examples of merged stack and M-distance sequences.

Well formed distance	$d(A)$	$\bar{1}$		$\bar{2}$	2			$\bar{3}$	$\bar{4}$			3	2
	$d(B)$		$\bar{1}$			1	$\bar{2}$			2	$\bar{3}$		
	$p(M)$	$A\bar{1}$	$B\bar{1}$	$A\bar{2}$	$A2$	$B1$	$B\bar{2}$	$A\bar{3}$	$A\bar{4}$	$B2$	$B\bar{3}$	$A3$	$B2$
Reference	$r(A)$	a_1		a_2	a_1			a_3	a_4			a_1	
	$r(B)$		b_1			b_1	b_2			b_1	b_3		b_1
	$r(M)$	a_1	b_1	a_2	a_1	b_1	b_2	a_3	a_4	b_1	b_3	a_1	b_1
Merged stack	$s(M)$	a_1	b_1 a_1	a_2 b_1 a_1	a_1 a_2 b_1	b_1 a_1 a_2	b_2 b_1 a_1 a_2	a_3 b_2 b_1 a_1 a_2	a_4 a_3 b_2 b_1 a_1 a_2	b_1 a_4 a_3 b_2 a_1 a_2	b_3 b_1 a_4 a_3 b_2 a_1 a_2	a_1 b_3 b_1 a_4 a_3 b_2 a_2	b_1 a_1 b_3 a_4 a_3 b_2 a_2
M-distance	$d(M)$	$\bar{1}$	$\bar{2}$	$\bar{3}$	3	3	$\bar{4}$	$\bar{5}$	$\bar{6}$	4	$\bar{7}$	6	3

of each distance. For notational convenience the pair $[u_j(M), d_j^*(u_j(M)]$ is often written as a concatenation and denoted by $p_j^*(M)$ as, for example, in $p_j^*(M) = A5$. A similar notation is used for well formed merged distance sequences, e.g., $p_j(M) = B\bar{4}$.

A *merged reference sequence* (for access streams A and B) is denoted by $r(M) = \{r_j(M); j \geq 1\}$, where $r_j(M) \in \{\mathscr{A} \cup \mathscr{B}\}, j \geq 1$. For merged reference sequences the type of each reference is not included explicitly since it can be deduced from $r_j(M)$. A *well formed merged stack sequence* $s(M) = \{s_j(M); j \geq 0\}$, with each $s_j(M)$ an ordered subset of $\mathscr{A} \cup \mathscr{B}$, is defined by viewing $r(M)$ as a reference sequence for a single access stream and requiring $r(M) \rightarrow s(M)$. Finally, the well formed distance sequence $d(M) = \{d_j(M); j \geq 1\}$ associated with $s(M)$ is called the *M-distance sequence.* Observe that each distance in the M-distance sequence is an element of the set $\{\bar{1}, \bar{2}, \cdots, \overline{m+n}, 1, 2, \cdots, m+n\}$ and that the M-distance sequence determines the miss ratio function for the merged reference sequence.

Definition 5.2 Let $p(M)$ be a well formed merged distance sequence, let $d(A) = \{d_j(u_j(M)); u_j(M) = A\}$, let $d(B) = \{d_j(u_j(M)); u_j(M) = B\}$, and let $r(A)$ and $r(B)$ be reference sequences where $r(A) \rightarrow d(A)$ and $r(B) \rightarrow d(B)$. A merged reference sequence $r(M)$, *associated* with $p(M)$ and denoted by $p(M) \rightarrow r(M)$, is obtained by replacing the subsequences $d(A)$ and $d(B)$ in the sequence $\{d_j(u_j(M)); j \geq 1\}$ by $r(A)$ and $r(B)$, respectively.

Table 3 illustrates these definitions. The following definition and proposition provide a mechanism for associating a merged stack sequence $s(M)$ directly with a well formed merged distance sequence $p(M)$.

Definition 5.3 For access streams A and B suppose that $p(M) = \{u_j(M), d_j(u_j(M)); j \geq 1\}$ is a well formed merged distance sequence. A merged stack sequence $s(M) = \{s_j(M); j \geq 0\}$ *associated* with $p(M)$ is defined recursively as follows:

1. For $j = 0$, $s_0(M) = \varnothing$ and $\gamma_0 = 0$.
2. For $j > 0$, suppose $s_j(M)$ and γ_j have been determined;
   a. if $d_{j+1}(u_{j+1}(M)) = k$ and l is the kth smallest index such that $s_{j,l}(M)$ is of type $u_{j+1}(M)$, then

$$s_{j+1}(M) = [s_{j,l}(M), s_{j,1}(M), \cdots, s_{j,l-1}(M), s_{j,l+1}(M), \cdots, s_{j,\gamma_j}(M)]$$

   and $\gamma_{j+1} = \gamma_j$.

   b. if $d_{j+1}(u_{j+1}(M)) = \bar{k}$, then

$$s_{j+1}(M) = \begin{cases} [a_k, s_j(M)], \text{ if } u_{j+1}(M) = A, \\ [b_k, s_j(M)], \text{ if } u_{j+1}(M) = B, \end{cases}$$

   and $\gamma_{j+1} = \gamma_j + 1$.

Proposition 5.4 Let $p(M)$ be a well formed merged distance sequence, let $r(M)$ be a merged reference sequence where $p(M) \rightarrow r(M)$, and let $s(M)$ be a merged stack sequence associated with $p(M)$ as given by Definition 5.3. Then $s(M)$ is well formed, $\hat{r}(M) \sim r(M)$ where $s(M) \rightarrow \hat{r}(M)$, and $d(M)$ is unique where $r(M) \rightarrow d(M)$.

6. Proof of main result

Consider now the merged distance sequence for two streams under the assumptions of Section 3. It is easily shown from the association of merged distance sequences with well formed merged distance sequences and from Definition 5.3 above that the well formed merged stack sequence $s(M)$ is a Markov chain (see, e.g., Parzen [6]), and the long-run miss ratio function can in principle be obtained from this chain. However, to obtain the miss

Table 4 Examples of merged type stack sequences.

Sequence		j: 0	1	2	3	4	5	6	7	8	9	10
Merged distance	$d_j{}^*(M)$		$A4$	$A4$	$B2$	$B1$	$B3$	$A4$	$B2$	$B4$	$B1$	$B2$
Merged reference	$r_j(M)$		a_1	a_2	b_1	b_1	b_2	a_3	b_1	b_3	b_3	b_1
Merged type	$t_j(M, 5)$	$\varnothing$	A	A	B	B	B	A	B	B	B	B
		$\varnothing$	$\varnothing$	A	A	A	B	B	A	B	B	B
		$\varnothing$	$\varnothing$	$\varnothing$	A	A	A	B	B	A	A	A
		$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	A	A	A	B	B	B
		$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	A	A	A	A	A

ratio function in closed form it is more convenient to consider another related Markov chain of merged "type" stacks.

Informally, a well formed merged type stack sequence $t(M, c) = \{t_j(M, c); j \geq 0\}$ differs from $s(M)$ in that

1. A type stack only indicates the type of each stack entry. For example, if $s_j(M) = (a_1, b_2, a_3, a_2, b_1)$ then $t_j(M, 5)$ is (A, B, A, A, B).
2. For a nonnegative integer c, the type stack $t_j(M, c)$ contains just c entries corresponding to the first c elements of $s_j(M)$ if $\gamma_j \geq c$, or to $s_j(M)$ followed by $c - \gamma_j$ "empty" entries denoted by $\varnothing$ if $\gamma_j < c$.

In the example above, $t_j(M, 3) = (A, B, A)$ and $t_j(M, 7) = (A, B, A, A, B, \varnothing, \varnothing)$. The miss ratio calculation for capacity c is obtained from $t(M, c)$ for $1 \leq c \leq m + n$. A calculation with type stacks is preferable because the number of possible type stacks is much smaller than the number of stacks.

Let $\{x_1, \cdots, x_n\}^c$ denote the set of all c-tuples over the symbols $x_1, \cdots, x_n$ and let $(x_1, \cdots, x_n)^*$ denote the set of all sequences of length zero or more over these symbols.

Definition 6.1 Let c be a nonnegative integer and let $p^*(M) = \{u_j(M), d_j^*(u_j(M)); j \geq 1\}$ be a merged distance sequence. The associated *well formed merged type stack sequence* $t(M, c) = \{t_j(M, c); j \geq 0\}$, where $t_j(M, c) = [t_{j,1}(M, c), \cdots, t_{j,c}(M, c)] \in \{A \cup B \cup \varnothing\}^c$, is defined recursively:

1. For $j = 0$, $t_0(M, c) = \{\varnothing\}^c$.
2. For $j \geq 0$, suppose that $t_j(M, c)$ has been determined;
   a. if $d_{j+1}^*[u_{j+1}(M)] > N[u_{j+1}(M): t_j(M, c)]$, then

$$t_{j+1}(M, c) = [u_{j+1}(M), t_{j,1}(M, c), \cdots, t_{j,c-1}(M, c)];$$

   b. otherwise, let l be the index such that

$$N[u_{j+1}(M): t_{j,1}(M, c), \cdots, t_{j,l}(M, c)] = d_{j+1}^*(u_{j+1}(M)).$$

Then,

$$t_{j+1}(M, c) = [u_{j+1}(M), t_{j,1}(M, c), \cdots, t_{j,l-1}(M, c), t_{j,l+1}(M, c), \cdots, t_{j,c}(M, c)].$$

Table 4 illustrates Definition 6.1 for the case $c = 5$.

• *Markov chain of merged type stacks*

From the probabilistic assumptions of Section 3, the stochastic process $P^*(M) = \{U_j(M), D_j^*(U_j(M)); j \geq 1\}$, called a merged distance process, is a sequence of pairs of i.i.d. random variables, where

$$\Pr\{U_j(M) = A, D_j^*(U_j(M)) = k\} = \delta\alpha_k, \quad 1 \leq k \leq m,$$

$$\Pr\{U_j(M) = B, D_j^*(U_j(M)) = k\} = (1 - \delta)\beta_k, \quad 1 \leq k \leq n, \tag{1}$$

and $\delta = \lambda_A / (\lambda_A + \lambda_B)$. Note that a merged distance sequence $p^*(M)$ is a realization of the process $P^*(M)$ and consider the stochastic process for merged type stacks

$$T(M, c) = \{T_j(M, c); j \geq 0\}, c \geq 1,$$

where $T_j(M, c) \in \{A \cup B \cup \varnothing\}^c$. A realization of this process, denoted $t(M, c)$, is derived using Definition 6.1 from a merged distance sequence $p^*(M)$, a realization of $P^*(M)$. Definition 6.1, together with the i.i.d. property of $P^*(M)$, leads to the following proposition.

Proposition 6.2 The stochastic process of merged type stacks $T(M, c)$ is a Markov chain. Furthermore,

1. If $\delta = 1$ and $c \leq m$, the chain has a single recurrent

state $\{A\}^c$. If $\delta = 0$ and $c \le n$, the chain has a single recurrent state $\{B\}^c$;

2. If $0 < \delta < 1$, $c \le m + n$, $\alpha_m > 0$, and $\beta_n > 0$, the chain has a single irreducible closed set of recurrent states

$$Z(M, c) = \{z : z \in \{A \cup B\}^c, N(A:z) \le m,$$
$$N(B:z) \le n\}.$$

To prove 2 in Proposition 6.2 suppose that $c = m + n$ and $t_j(M, m+n) = (X_1, \cdots, X_{m+n}) \in Z(M, m+n)$. It is first shown that this state communicates with $(X_1, \cdots, X_{l-1}, X_{l+1}, X_l, X_{l+2}, \cdots, X_{m+n})$ for any l, $1 \le l \le m+n-1$, where $X_l \ne X_{l+1}$. Since $0 < \delta < 1$, $\alpha_m > 0$, and $\beta_n > 0$, state $(X_1, \cdots, X_{m+n})$ communicates with $(X_{m+n}, X_1, \cdots, X_{m+n-1})$ and thus with $(X_{l+2}, \cdots, X_{m+n}, X_1, \cdots, X_l, X_{l+1})$. Now, since $X_l \ne X_{l+1}$, this latter state communicates with $(X_l, X_{l+2}, \cdots, X_{m+n}, X_1, \cdots, X_{l-1}, X_{l+1})$ and thus with state $(X_1, \cdots, X_{l-1}, X_{l+1}, X_l, X_{l+2}, \cdots, X_{m+n})$. Now, since $t_j(M, m+n)$ communicates with any state obtained by permuting two successive elements of $t_j(M, m+n)$, it communicates with all states in $Z(M, m+n)$. It is easily shown that $Z(M, m+n)$ contains all the recurrent states, completing the proof for $c = m + n$. This result can then be used to establish 2 for $c < m + n$. It is assumed subsequently that the conditions 2 of Proposition 6.2 are met. There is no loss of generality since $\delta = 0$ or $\delta = 1$ corresponds to only a single access stream and $\alpha_m = 0$ or $\beta_n = 0$ can be interpreted as a different access stream referencing a smaller set of blocks. Note that the set $Z(M, c)$ contains 2^c states for $1 \le c \le \min(m, n)$ and $\sum_{i=i_0}^{\min(m,n)} \binom{c}{i}$ states for $\min(m, n) < c \le m + n$, where $i_0 = \max(0, (c - \max(m, n)))$.

The long run probability, $\lim_{j\to\infty} \Pr\{T_j(M, c) = z\}$, where $z \in Z(M, c)$, that a particular merged type stack occurs is determined by considering the sets of merged reference and distance sequences of length j that yield the event $\{T_j(M, c) = z\}$ for a given j. For ease of exposition and to avoid additional notation, the development involves a particular example, viz., $z = (B, B, A, B, A) \in Z(M, 5)$.

For $z \in Z(M, c)$ let $\Psi(z)$ denote the set of finite length merged reference sequences that yield the type stack z. Noting that an LRU stack at index j is a list of the blocks referenced in the first j references and ordered by their last reference, we claim that all the reference sequences in $\Psi(z)$ for $z = (B, B, A, B, A)$ are equivalent to elements of the set

$$(a_1, \cdots, a_m, b_1, \cdots, b_n)^* a_2 (a_1, b_1, b_2, b_3)^*$$
$$b_3 (a_1, b_1, b_2)^* a_1 (b_1, b_2)^* b_2 (b_1)^* b_1. \qquad (2)$$

It is difficult to characterize the well formed merged distance sequences associated with elements of $\Psi(z)$. Instead, let $\Omega(z)$ denote the set of finite length well formed merged distance sequences such that an associated merged reference sequence is the *reverse* of an element of $\Psi(z)$. For the example it is claimed that

$$\Omega(B, B, A, B, A,) = B\bar{1}(B1)^* B\bar{2}(B1, B2)^*$$
$$A\bar{1}(A1, B1, B2)^* B\bar{3}(A1, B1, B2, B3)^* A2\chi, \qquad (3)$$

where χ is any finite string over $(A\bar{3}, \cdots, A\bar{m}, B\bar{4}, \cdots, B\bar{n}, A1, \cdots, Am, B1, \cdots, Bn)$ which yields a well formed merged distance sequence according to Definition 5.1. To see this, consider the reverse of an element of (2). The "first" b_1 is associated with $B\bar{1}$ in (3). Successive occurrences of b_1 [denoted by $(b_1)^*$ in (2)] are associated with $(B1)^*$ in (3) until b_2 occurs, which is associated with $B\bar{2}$; etc.

The notion of twin distance sequences can be extended to merged distance sequences (not necessarily well formed) by defining the *twin* of a merged distance sequence as that (unique) sequence obtained by replacing the subsequence of distances for access stream A by its twin and the subsequence of distances for access stream B by its twin.

Lemma 6.3 For $z = (z_1, \cdots, z_c) \in Z(M, c)$ the event $\{T_j(M, c) = z\}$ occurs if and only if the twin of the well formed merged distance sequence $p_1(M), \cdots, p_j(M)$ is an element of $\Omega(z)$. The set $\Omega(z)$ can be expressed as $\chi_1 \cdots \chi_c$ where

1. For $1 \le k \le c - 1$, let $x_1 = N(z_k : z_1, \cdots, z_k)$, $x_2 = N(A : z_1, \cdots, z_k)$, and $x_3 = N(B : z_1, \cdots, z_k)$. The set χ_k is given by $\chi_k = z_k \bar{x}_1 (A1, \cdots, Ax_2, B1, \cdots, Bx_3)^*$.
2. For $k = c$ and $x_1 = N(z_i : z_1, \cdots, z_c)$ the set χ_c is given by $\chi_c = z_c \bar{x}_1 \chi$, where χ is the set of all finite sequences over $(A\bar{1}, \cdots, A\bar{m}, B\bar{1}, \cdots, B\bar{n}, A1, \cdots, Am, B1, \cdots, Bn)$ such that $\chi_1 \cdots \chi_c$ is a well formed merged distance sequence.

This is a key result since the occurrence of a particular merged type stack (from which miss ratios are calculated) can be interpreted as a property of a merged distance sequence. Thus, letting $q(j)$ denote the subset of realizations of the merged distance sequence of length j, such that the associated well formed merged distance sequence is a member of $\Omega(z)$,

$$\Pr\{T_j(M, c) = z\} = \Sigma \Pr\{P_1^*(M) = q_1, \cdots, P_j^*(M) = q_j\},$$
$$z \in Z(M, c), \qquad (4)$$

where the summation is over all $(q_1, \cdots, q_j)$ such that the twin $(q_1', \cdots, q_j') \in q(j)$. Since any merged distance sequence is a permutation of its twin and since $\{P_j^*(M); j \ge 1\}$ are i.i.d. random variables,

$$\Pr\{P_1^*(M) = q_1, \cdots, P_j^*(M) = q_j\}$$
$$= \Pr\{P_1^*(M) = q_1', \cdots, P_j^*(M) = q_j'\},$$

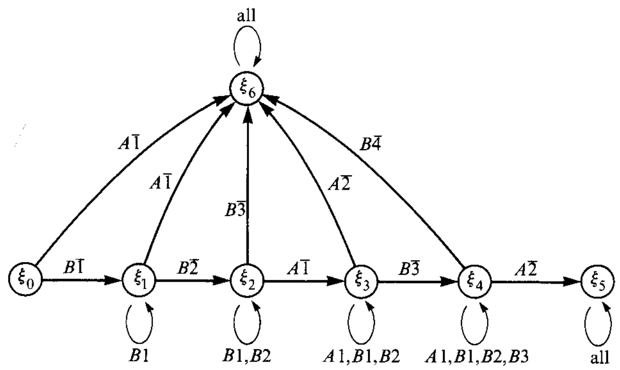

Figure 2 State transition diagram for Markov chain $H(z)$ for $z = (B, B, A, B, A)$.

where $q_1, \cdots, q_j$ and $q'_1, \cdots, q'_j$ are twins; Eq. (4) can also be written

$$\Pr\{T_j(M, c) = z\} = \sum_{(q_1, \cdots, q_j) \in q(j)} \Pr\{P_1^*(M) = q_1, \cdots, P_j^*(M) = q_j\}, z \in Z(M, c). \quad (4')$$

• *Markov chain* $H(z)$

To compute the right hand side of Eq. (4′) it is convenient to introduce for $z \in Z(M, c)$ a Markov chain $H(z) = \{H_k(z); k \geq 0\}$ based on the merged distance sequence process $P^*(M)$. The chain has a state ξ_c such that $H_j(z) = \xi_c$ if and only if $p_1^*(M), \cdots, p_j^*(M)$ is such that its well formed counterpart $p_1(M), \cdots, p_j(M) \in \Omega(z)$.

Definition 6.4 For $z = (z_1, \cdots, z_c) \in Z(M, c)$, let $H(z) = \{H_k(z); k \geq 0\}$ be a Markov chain where

1. the state space is $\{\xi_0, \cdots, \xi_{c+1}\}$;
2. the initial state is ξ_0, i.e., $\Pr\{H_0(z) = \xi_0\} = 1$;
3. for $l = 1, \cdots, c$, $\Pr\{H_k(z) = \xi_l | H_{k-1}(z) = \xi_{l-1}\}$

$$= \begin{cases} \Pr\{U_l(M) = A, D_l^*(U_l(M)) > N(A: z_1, \cdots, z_{l-1})\} \text{ if } z_l = A; \\ \Pr\{U_l(M) = B, D^*(U_l(M)) > N(B: z_1, \cdots, z_{l-1})\} \text{ if } z_l = B; \end{cases}$$

4. for $l = 1, \cdots, c-1$, $\Pr\{H_k(z) = \xi_l | H_{k-1}(z) = \xi_l\}$

$$= \Pr\{U_l(M) = A, D_l^*(U_l(M)) \leq N(A: a_1, \cdots, z_l)\} + \Pr\{U_l(M) = B, D^*(U_l(M)) \leq N(B: z_1, \cdots, z_1)\};$$

5. $\Pr\{H_k(z) = \xi_c | H_{k-1}(z) = \xi_c\} = 1$;
6. for $l = 0, \cdots, c-1$, $\Pr\{H_k(z) = \xi_{c+1} | H_{k-1}(z) = \xi_l\}$

$$= \begin{cases} \Pr\{U_l(M) = A, D_l^*(U_l(M)) > N(A: z_1, \cdots, z_l)\} \text{ if } z_{l+1} = B; \\ \Pr\{U_l(M) = B, D_l^*(U_l(M)) > N(B: z_1, \cdots, z_l)\} \text{ if } z_{l+1} = A; \end{cases}$$

7. $\Pr\{H_k(z) = \xi_{c+1} | H_{k-1}(z) = \xi_{c+1}\} = 1$; and
8. all other one-step transition probabilities are zero.

As an example of the chain $H(z)$, a state transition diagram for $z = (B, B, A, B, A)$ is shown in Figure 2. The transitions are labeled by the associated elements of the well formed merged distance sequence. Thus, $p^*(M) = (B4, B1, B5, B1, A3, B2, A1, A3, \cdots)$ would cause the state transition sequence $(\xi_0, \xi_1, \xi_1, \xi_2, \xi_2, \xi_3, \xi_3, \xi_3, \xi_6, \xi_6, \cdots)$.

In the chain $H(z)$, states ξ_c and ξ_{c+1} are absorbing states representing, respectively, that the well formed merged reference sequence is or is not an element of $\Omega(z)$. The remaining states are transient and indicate that the question of membership in $\Omega(z)$ has not yet been determined.

Proposition 6.5 For $z \in Z(M, c)$ and $j \geq 1$

$$\Pr\{T_j(M) = z\} = \Pr\{H_j(z) = \xi_c\}. \quad (5)$$

Since $Z(M, c)$ is an irreducible closed set of recurrent states, for all $z \in Z(M, c)$, $\pi(z) = \lim_{j \to \infty} \Pr\{T_j(M) = z\}$ exists such that $\pi(z) > 0$ and

$$\sum_{z \in Z(M, c)} \pi(z) = 1.$$

Lemma 6.6 For all $z \in Z(M, c)$

$$\lim_{j \to \infty} \Pr\{H_j(z) = \xi_c\} = \pi(z).$$

Thus $\pi(z)$ is the probability in $H(z)$ of ultimate absorption in state ξ_c; this quantity can be easily determined by inspection from the structure of $H(z)$.

Recall from Eq. (1) that

$$\Pr\{U_j(M) = A, D_j^*(U_j(M)) = k\} = \delta\alpha_k, \quad 1 \leq k \leq m, j \geq 1,$$

and

$$\Pr\{U_j(M) = B, D_j^*(U_j(M)) = k\} = (1 - \delta)\beta_k, \quad 1 \leq k \leq n, j \geq 1,$$

and let

$$\alpha'_i = \begin{cases} \sum_{j=i}^{m} \alpha_j, \; i = 1, \cdots, m, \\ 0, \text{ otherwise}; \end{cases}$$

$$\beta'_i = \begin{cases} \sum_{j=i}^{n} \beta_j, \; i = 1, \cdots, n, \\ 0, \text{ otherwise}. \end{cases}$$

Note that $\delta\alpha'_i$ is the joint probability that $U_j(M) = A$ and $D_j^*(U_j(M)) \geq i$, i.e., the probability that the well formed realization has $p_j(M) = A\bar{i}$ given that $p_l(M) = A\overline{i-1}$ and

$p_{l+1}(M) \neq A\bar{i}, \cdots, p_{j-1}(M) \neq A\bar{i}$ for some $l < i$. Now, for $z = (z_1, \cdots, z_c) \in Z(M, c)$ and $j = 1, \cdots, c$, let

$$G(z, j) = \begin{cases} \dfrac{\delta\alpha'_{N(A:z_1,\cdots,z_j)}}{\delta\alpha'_{N(A:z_1,\cdots,z_j)} + (1-\delta)\beta'_{N(B:z_1,\cdots,z_j)+1}} & \text{if } z_j = A; \\ \dfrac{(1-\delta)\beta'_{N(B:z_1,\cdots,z_j)}}{\delta\alpha'_{N(A:z_1,\cdots,z_j)+1} + (1-\delta)\beta'_{N(B:z_1,\cdots,z_j)}} & \text{if } z_j = B. \end{cases} \quad (6)$$

Given that $H_k(z) = \xi_{j-1}$, $G(z, j)$ is seen to be the probability that $H_l(z) = \xi_j$ for some $l > k$, i.e., of ever making the transition $\xi_{j-1} \to \xi_j$. For the example in Fig. 2,

$$G(z, 1) = \frac{(1-\delta)\beta'_1}{\delta\alpha'_1 + (1-\delta)\beta'_1} = 1 - \delta,$$

corresponding to $\Pr\{U_1(M) = B\}$.

It follows from this that the probability of ultimate absorption in ξ_c, and thus $\pi(z)$, is given by the following lemma.

Lemma 6.7 For $z \in Z(M, c)$, $\pi(z) = \prod_{i=1}^{c} G(z, i)$. Recall that the M-distance sequence $d(M) = \{d_j(M); j \geq 1\}$ is the well formed distance sequence associated with the well formed merged stack sequence $s(M)$. The corresponding stochastic process $D(M) = \{D_j(M); j \geq 1\}$ is seen for $c = 1, \cdots, m + n$ to have the property

$$\begin{aligned} \Pr\{D_j(M) = c\} &= \Sigma \Pr\{T_{j-1}(M) = z\} \\ &\quad \times \Pr\{U_j(M) = A, D_j^*(U_j(M)) = N(A:z)\} \\ &\quad + \Sigma \Pr\{T_{j-1}(M) = z\} \\ &\quad \times \Pr\{U_j(M) = B, D_j^*(U_j(M)) = N(B:z)\}, \end{aligned} \quad (7)$$

where the first summation is over all $z \in Z(M, c)$ such that $z = (z_1, \cdots, z_{c-1}, A)$, and the second summation is over all $z \in Z(M, c)$ such that $z = (z_1, \cdots, z_{c-1}, B)$. In other words, an M-distance of c is observed just when a type stack $z \in Z(M, c)$ exists and the cth entry in the stack is "referenced."

Theorem 6.8 For $c = 1, \cdots, m + n$ and $k \geq 1$, let $M_k(c)$ be the fraction of M-distances $D_j(M)$, $1 \leq j \leq k$, greater than c. Then the (long-run) *expected miss ratio function* $M(c) = \lim_{k\to\infty} \mathrm{E}\{M_k(c)\}$ for the merged stream exists and is given by

$$M(c) = \sum_{z \in Z(M,c)} \pi(z)(\delta\alpha'_{N(A:z)+1} + (1-\delta)\beta'_{N(B:z)+1}). \quad (8)$$

To see this, observe that $\Pr\{D_j(M) > c\}$ can be obtained from Eq. (7) by replacing the second term in both sums by $\delta\alpha'_{N(A:z)+1} + (1-\delta)\beta'_{N(B:z)+1}$. The sums can then be combined into a single sum over $z \in Z(M, c)$ and since $Z(M, c)$ is an irreducible aperiodic recurrent class (cf. [6]), letting $j \to \infty$ we obtain Eq. (8).

Theorem 6.8, together with Lemma 6.7, completes the proof of Theorem 3.1 for the case of $J = 2$ streams. Extension of the proof to the case $J > 2$ involves the preliminaries of Section 4 and the same sequence of steps used in Sections 5 and 6. The merged sequences of Section 5 become multivariate (J-type) sequences. The stochastic process of merged type stacks can be shown to be a Markov chain (as in Proposition 6.2) and under the conditions $c \leq \Sigma_{j=1}^{J} m_j$, $\alpha_{m_j}(A_j) > 0$ and $\delta_j > 0$ for $j = 1, \cdots, J$, the chain has a single irreducible closed set of recurrent states $Z(M, c) = \{z; z \in \{A_1 \cup \cdots \cup A_J\}^c, N(A_j{:}z) \leq m_j \text{ for } j = 1, \cdots, J\}$. Lemma 6.3 relating the occurrence of a type stack to well formed merged distance sequences; the definition of the Markov chains $H(z)$ for $z \in Z(M, c)$; and Proposition 6.5, Lemma 6.6, and Lemma 6.7 relating to the determination of $\{\pi(z)\}$ all extend directly for $J > 2$.

7. Other results

Calculation of the expected miss ratios for $J = 2$ and capacity c from Eq. (8) involves a sum over $z \in Z(M, c)$ and thus is of complexity 2^c (J^c in general). This can be a severe practical limitation. However, there is a less complex computational procedure whereby $\pi(z)$ need not be determined for each $z \in Z(M, c)$. The procedure, which is not described in detail here, involves partitioning the states of $Z(M, c)$ into classes such that z and z' are in the same class if and only if $N(A{:}z) = N(A{:}z')$. For $c \geq 1$ let $Q(k, c)$ denote the long-run probability that a merged type stack contains exactly k type A entries, $\max(0, c-n) \leq k \leq \min(m, c)$. By expressing $Q(k; c)$ as a sum over the appropriate $\pi(z)$ the following lemma can be established.

Lemma 7.1

1. For $c = 1$, $Q(0; 1) = 1 - \delta$ and $Q(1; 1) = \delta$. (9)
2. For $c \geq 2$ and $\max(0, c - n) \leq k \leq \min(m, c)$,

$$\begin{aligned} Q(k; c) &= Q(k; c-1)\left[\frac{(1-\delta)\beta'_{c-k}}{\delta\alpha'_{k+1} + (1-\delta)\beta'_{c-k}}\right] \\ &\quad + Q(k-1; c-1)\left[\frac{\delta\alpha'_k}{\delta\alpha'_k + (1-\delta)\beta'_{c-k+1}}\right]. \end{aligned} \quad (10)$$

The following theorem can then be demonstrated.

Theorem 7.2 For $c \geq 1$

$$M(c) = \sum_{k=\max(0,\,c-n)}^{\min(m,\,c)} Q(k; c)[\delta\alpha'_{k+1} + (1-\delta)\beta'_{c-k+1}].$$

Equations (9) and (10) and Theorem 7.2 constitute the less complex procedure for calculating $M(c)$, which is seen to have complexity c^2. Additionally, $\{Q(k; c)\}$

provides the long-run distribution of the number of type A blocks in a first-level device, which is also of interest. For $J > 2$ and $c \geq 1$ let $Q(k_1, \cdots, k_J; c)$, where $\Sigma_{j=1}^{J} k_j = c$, denote the long-run probability that a merged type stack contains exactly k_i type A_i entries, $1 \leq i \leq J$.

Theorem 7.3

1. For $c = 1$ and $j = 1, \cdots, J$, $Q(k_1, \cdots, k_J; 1) = \delta_j$ when $k_j = 1$ and $k_i = 0$, $i \neq j$.
2. For $c \geq 2$ and all $k_1, \cdots, k_J$ such that $0 \leq k_j \leq m_j$ and $\Sigma_{i=1}^{J} k_i = c$,

$$Q(k_1, \cdots, k_J; c) = \sum_{j=1}^{J} \frac{\delta_j \mu_{k_j}(A_j)}{\sum_{l=1}^{J} \delta_l \mu_{k_l+1-\Delta_{jl}}(A_l)} \times Q(k_1 - \Delta_{1j}, \cdots, k_J - \Delta_{Jj}; c-1),$$

where $\Delta_{ij} = 1$ if $i = j$ and 0 otherwise.

3. $M(c) = \sum Q(k_1, \cdots, k_J; c)\left[\sum_{i=1}^{J} \delta_i \mu_{k_i+1}(A_i)\right]$,

where the summation is over all $k_1, \cdots, k_J$ such that $0 \leq k_j \leq m_j$ and $\Sigma_{i=1}^{J} k_i = c$.

Use of Theorem 7.3 for calculating miss ratios involves complexity c^J. This is a substantial improvement over the use of Theorem 3.1, which involves complexity J^c.

For fixed capacity c it is interesting to consider the range of values of the expected merged miss ratio $M(c)$ when only the rates λ_i of the set of access streams are varied. The following lemma establishes that the minimum value of $M(c)$ is equal to the minimum expected miss ratio of the individual streams at capacity c.

Lemma 7.4 Consider $J \geq 2$ access streams $A_1, \cdots, A_J$ satisfying the assumptions of Section 3 where, for $j = 1, \cdots, J$ and $i \geq 1$, $\mu_i(A_j)$ is the miss ratio function (evaluated at capacity $i - 1$) for the jth stream and $\delta_j = \lambda_j / \Sigma_{i=1}^{J} \lambda_i$. For $1 \leq j \leq J$, the minimum value of the expected merged miss ratio $M(c)$ over $0 \leq \delta_j \leq 1$ and $\Sigma_{j=1}^{J} \delta_j = 1$ is given by $\mu_{c+1}(A_k)$ for any k such that for $1 \leq j \leq J$, $\mu_{c+1}(A_k) \leq \mu_{c+1}(A_j)$.

To see this, suppose that $\mu_{c+1}(A_1) \leq \mu_{c+1}(A_j)$ for $j \geq 2$. From 3 of Theorem 7.3 it follows that for any $\delta_1, \cdots, \delta_J$,

$$M(c) \geq \sum Q(k_1, \cdots, k_J; c)\left[\sum_{i=1}^{J} \delta_i \mu_{c+1}(A_i)\right] = \sum_{i=1}^{J} \delta_i \mu_{c+1}(A_i).$$

Furthermore,

$$\sum_{i=1}^{J} \delta_i \mu_{c+1}(A_i) \geq \sum_{i=1}^{J} \delta_i \mu_{c+1}(A_1) = \mu_{c+1}(A_1),$$

which implies that $\delta_1 = 1$ and $\delta_j = 0$ for $j \geq 2$ minimizes $M(c)$.

The set of rates that yields the maximum value of $M(c)$ must generally be determined empirically. Surprisingly, when all the access streams have identical LRU stack distance processes, the maximum expected merged miss ratio is not necessarily achieved by choosing equal rates for all streams.

8. Remarks

The results of this paper provide a method for predicting the miss ratio function for a multistream environment from miss ratio functions for individual streams. The method involves representing each access stream by a sequence of LRU stack distances evolving in time as a Poisson process and viewing the composite stream as the superposition of these processes.

Interesting extensions to this work lie in varying either the representation of the individual access streams, the merging mechanism, or both. For example, the reference sequence of each access stream could be represented by a (finite order) Markov chain of stack distances (cf., [7]). With the same merging mechanism, namely the superposition of Poisson processes, the merged stack sequence would still be a Markov chain.

Alternatively, other merging mechanisms could be studied. The merging mechanism considered in this paper has the property that the number of successive accesses from a given stream in a merged reference sequence is geometrically distributed. This may well be unrealistic and it would be of interest to compare results obtained from merging mechanisms that more closely represent the access patterns observed in actual systems. A system incorporating "time slicing," for example, might be usefully represented by forming (for positive integral N) a merged reference sequence from N accesses from stream A_1, N accesses from stream A_2, etc.

An important extension of the results of this paper would be to relax the assumption that the sets of blocks accessed by the individual streams are mutually disjoint. (Such "sharing" exists to some extent in most real systems and can have a significant effect on the miss ratio function.) Several formulations of sharing are possible. Examination of actual system structures might suggest an appropriate formulation. For example, it is often the case that only certain blocks (e.g., those of compilers and system code) are shared, the remaining blocks being private to the individual streams. Prediction of composite miss ratio functions for such situations, however, appears to be difficult.

References

1. R. L. Mattson, J. Gecsei, D. R. Slutz, and I. L. Traiger, "Evaluation Techniques for Storage Hierarchies," *IBM Syst. J.* **9**, 78 (1970).
2. P. H. Oden and G. S. Shedler, "A Model for Memory Contention in a Paging Machine," *Comm. ACM* **15**, 761 (1972).

3. J. E. MacDonald, D. L. Moorer, and K. L. Sigworth, IBM System Products Division, Poughkeepsie, New York, private communication, December 1975.
4. D. R. Cox and P. A. W. Lewis, "Multivariate Point Processes" in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, edited by L. LeCam, J. Neyman, and E. L. Scott, University of California Press, Berkeley, California, 1972, p. 401.
5. E. Çinlar, "Superposition of Point Processes" in *Stochastic Point Processes*, edited by P. A. W. Lewis, John Wiley & Sons, Inc., New York, 1972, p. 549.
6. E. Parzen, *Stochastic Processes*, Holden-Day, Inc., San Francisco, CA, 1962.
7. G. S. Shedler and C. Tung, "Locality in Page Reference Strings," *SIAM J. Comput.* **1**, 218 (1972).

Received January 19, 1976

The authors are located at the IBM Research Laboratory, 5600 Cottle Road, San Jose, CA 95193.