A Parallel Search Game

Navin Goyal[1] Michael Saks[2]

Abstract

We answer in negative a question of Gál and Miltersen [3] about a combinatorial game arising in the study of time-space trade-offs for data structures.
Keywords: Combinatorial Games, Permutations, Data Structures

1 Introduction

The following game is due to Gál and Miltersen [3].

We have b boxes labeled $0, \ldots, b-1$ and a slips of paper with $a \leq b$, labeled $0, \ldots, a-1$. The game is played between player I and a team consisting of players $\mathrm{II}_0, \ldots, \mathrm{II}_{a-1}$. Player I secretly colors each slip of paper either red or blue and puts each slip in a different box. Now player II_i can look in at most $b/2$ boxes using any adaptive strategy and based on this must make a guess about the color of the slip labeled i. This is done by each of the player on the team without communication or observation between them. The team wins if every player in the team correctly announces the color of his slip.

In their paper, Gál and Miltersen suggested the following hypothesis about the above game. If this hypothesis were true, it would imply good time-space trade-offs for a certain data structure problem (see section 3).

> Let $b \geq 2a$. Suppose player I adopts the strategy of coloring each slip of paper uniformly at random and independently putting them at random into a boxes chosen at random. Then no matter which strategy the team adopts, the probability that they win is at most $2^{-\Omega(a)}$.

The intuition behind this hypothesis is the following. It is easy to see that if $b = 2a$, then each individual player can guess the color of his slip correctly with probability $3/4$, and that is the best he can do. And since the players cannot communicate after the game starts, it seems that their success or failure in guessing their slip colors are more or less independent. We give a strategy which shows that this intuitive reasoning is incorrect. Under this strategy, the individual probabilities of success are $3/4$. But now when the team looses, many players guess incorrectly, thereby increasing the number of inputs on which everybody guesses correctly. This phenomenon is akin to the one underlying the *hat problem* of T. Ebert (see [1] and [2]). Apart from this high level similarity, the two problems appear to be rather different.

We consider slightly more general games by introducing one more parameter $k \geq 1$ into the above game, and requiring that each player in the team can open at most b/k boxes. Also we make the requirements the team needs to satisfy more stringent: Player II_i is required to continue searching in the boxes until he finds the slip labeled i (as opposed to guessing the color of his slip). The team wins if each player opens at most b/k boxes and finds his own slip. It is clear that the probability of winning this game is a lower bound on the probability of winning when the players are allowed to guess. We call this game *GM-game*. The game considered by Gál and Miltersen corresponds to the case $k = 2$.

The following theorem about the GM-game implies that the above hypothesis is not true. $\lg$ stands for logarithm to base 2.

[1]Department of Computer Science, Rutgers University, Piscataway, NJ 08854, USA, email: ngoyal@cs.rutgers.edu. Supported in part by NSF grant CCR-9988526.
[2]Department of Mathematics, Rutgers University, Piscataway, NJ 08854, USA, email: saks@math.rutgers.edu. Supported in part by NSF grant CCR-9988526.

Theorem 1 *Let positive integers a, b, k, the parameters for the GM-game, be such that $b = da$ and $a = 2kn^2$ for some positive integers $d > 1$ and n. If player I adopts the random strategy as in the hypothesis above, then there is a strategy for the team that succeeds with probability at least $2^{-9\sqrt{ka}\lg a - k}$, for some positive constant c independent of a, b, and d.*

The requirements that b/a be an integer and a be of the form $2kn^2$ are not important, and are adopted to make the proof cleaner. We will use the following well-known approximation of $n!$ due to Stirling (see, eg, [6]).

$$\sqrt{2\pi}\, n^{n+1/2}e^{-n}e^{\frac{1}{12n+1}} \leq n! \leq \sqrt{2\pi}\, n^{n+1/2}e^{-n}e^{\frac{1}{12n}}. \tag{1}$$

The rest of this paper is organized as follows. In the next section we prove the theorem. In the final section we describe its consequence for proving time-space trade-offs for data structures.

2 Proof

In this section we prove theorem 1. We begin with some definitions. $\oplus$ and $\ominus$ are used for arithmetic $\bmod\ b$; so, for example, $x \oplus y$ stands for $x + y \mod b$. Imagine the boxes as arranged in a cycle. For $s, t \in \{0, \ldots, b-1\}$, $[s, t]$ denotes set $\{s, s+1, \ldots, t\}$ if $s \leq t$, and it denotes set $[s, b-1] \cup [0, t]$ if $s > t$. So $[s, t]$ denotes the set we get when we go from s to t in the cyclic order ($s, s \oplus 1, s \oplus 2, \ldots$). As usual, for positive integer n, $[n]$ denotes the set $\{1, \ldots, n\}$.

A special case. First we look at the case $b = a$ and $k = 2$. For this case Gál and Miltersen mentioned that a strategy by S. Skyum achieves probabilty of success roughly 0.3 independent of a. Here we briefly describe Skyum's strategy and its analysis. Our strategy for the general case will be built upon this.

Since we are considering the $d = 1$ case, the placement of the slips in the boxes induces a permutation σ on $[0, a-1]$ taking $i \in [0, a-1]$ to $\sigma(i)$, which denotes the label of the slip in box i. The strategy of the players is to follow the cycles of this permutation:

> For $i = 0, \ldots, a-1$, player II_i starts by looking into box i. If $\sigma(i) = i$ then he is done, else he looks into the box numbered $\sigma(i)$, and repeats (by looking into boxes $\sigma(\sigma(i)), \sigma(\sigma(\sigma(i)))$ and so on) until he finds the slip numbered i.

The team wins the game if all the cycles are of length at most $a/2$. The probability of this happening is about 0.3 for large a. We sketch the proof here. For $i > \lfloor \frac{a}{2} \rfloor$, the probability that a uniformly randomly chosen permutation σ on a elements contains a cycle of length i is $1/i$. So the probability that σ has a cycle of length $> \lfloor \frac{a}{2} \rfloor$ is $1/a + 1/(a-1) + \ldots + 1/(\lfloor \frac{a}{2} \rfloor + 1)$. This converges to $\ln 2$ as a becomes large. So the probability that all the cycles in σ have length $\leq \lfloor \frac{a}{2} \rfloor$ is about $1 - \ln 2 \approx 0.3$.

The general case. Next we turn to the general case. The idea is to simulate the strategy for the $d = 1$ case, where the placement of the slips naturally defined a permutation of the set $[0, a-1]$. We first want to define such a permutation π when $d > 1$.

We need several definitions. For a given placement of the slips into the boxes by player I, define $\mathsf{occupied}[s, t]$ to be the number of boxes whose labels are in $[s, t]$ (from now on we will just say "boxes in $[s, t]$") and which contain slips; and define the *surplus* of $[s, t]$ by $\mathsf{surplus}[s, t] = \mathsf{occupied}[s, t] - |[s, t]|/d$. In other words, $\mathsf{surplus}[s, t]$ is the difference between the number of occupied boxes in $[s, t]$ and the expected number of occupied boxes in $[s, t]$. In the sequel the placement will often be implicit.

We partition the set of boxes into a sets called *bins*: $B_0, \ldots, B_{a-1}$, where B_i contains boxes in $[di, di + d - 1]$ for $i \in [0, a-1]$. The number of bins is the same as the number of slips.

We will use the following three properties of surplus which follow immediately from the definition.

S1 For $s \in [0, b-1]$,

$$\mathsf{surplus}[s, s \ominus 1] = 0. \tag{2}$$

That is, if we start from box s and go a full circle and come to box $s \ominus 1$ then its surplus is 0. This is because the total number of slips is equal to the expected total number of occupied boxes.

S2 It is additive: For $s, t, u \in [0, b-1]$,

$$\mathsf{surplus}[s, u] = \mathsf{surplus}[s, t] + \mathsf{surplus}[t \oplus 1, u]. \tag{3}$$

S3 For $i, j \in [0, a-1]$, $\mathsf{surplus}[di, dj \ominus 1]$ is an integer.

Now for $i \in [0, a-1]$, as we go around the cycle starting from box di, let $m(i)$ be the first integer j such that $\mathsf{surplus}[di, j]$ is nonnegative. This is well-defined by **S1**. Another way to think of $m(i)$ is as follows: If some box in bin B_i contains a slip then the index of the first such box is the value of $m(i)$; if there is no such box then we look into the next bin B_{i+1}, and $m(i)$ is the index of the second occupied box (if it exists) in B_{i+1} (the first occupied one is assigned to B_{i+1}); if it does not exist then we look for the third occupied box in B_{i+2}, and so on (indexes $i+1$ etc are modulo b here).

By the minimality of $m(i)$, the box $m(i)$ must contain a slip whose number we denote by $\pi(i)$. We say that the slip $\pi(i)$ is *associated* with the bin B_i. We observe:

Proposition 2 *π is a permutation of* $\{0, \ldots, a-1\}$.

Proof. Suppose not and that there are distinct $i, j \in [0, a-1]$ such that $\pi(i) = \pi(j)$ and so $m(i) = m(j) = m$. We derive a contradiction from this. Assume without loss of generality that in the cyclic order starting from m, di precedes dj.

We have $\mathsf{surplus}[di, m] = \mathsf{surplus}[di, dj \ominus 1] + \mathsf{surplus}[dj, m]$. Now, $\mathsf{surplus}[di, dj \ominus 1] \leq -1$ since it is an integer and is negative by definition of $m(i)$. We also have, $\mathsf{surplus}[dj, m] < 1$. This is because $\mathsf{surplus}[dj, m] = \mathsf{surplus}[dj, m \ominus 1] + \mathsf{surplus}[m, m] \leq \mathsf{surplus}[m, m] < 1$, where the first inequality uses **S1** if $dj = m$, and uses the definition of m if $dj \neq m$. ■

Next we note that, given i, $\pi(i)$ can be computed by the following algorithm: Sequentially probe the boxes starting from di, keeping track of the surplus, until the surplus is first nonnegative.

The algorithm followed by the players is now a natural extension of the case $d = 1$ relative to permutation π. For $i \in [0, a-1]$, player II_i uses the above subroutine to successively compute $\pi(i), \pi(\pi(i)), \pi(\pi(\pi(i))), \ldots$ until step k such that $\pi^{(k)}(i) = i$.

Analysis. In the rest of the proof we analyze the performance of this strategy. For evaluating $\pi(i)$, $|[di, m(i)]|$ probes are needed. Let M be the maximum of $|[di, m(i)]|$ for $i \in [0, a-1]$. The number of probes needed for any player in the team to find its slip using the above algorithm is upper bounded by M times the length of the longest cycle in π. It is possible that a player finds his slip before he finishes executing the above algorithm, but that only works in our favor.

The integers $m(h)$ and the permutation π are randomly determined by the choices of player I. Since the integers $m(h)$ depend only on the set of occupied boxes, and if we condition on this set of boxes, the labeling of the slips is uniformly distributed over all permutations, we conclude that the integers $m(h)$ are independent of π. We will separately lower bound the probabilities of the events that all cycles in π are small, and that to evaluate $m(h)$ not many boxes need be opened. Then the product of these two probabilities will give us the lower bound claimed in the theorem.

Let $p_1(\alpha)$ be the probability that all cycles in π have length $\leq \alpha$ for $\alpha \in [a]$. Clearly $p_1(\alpha)$ is at least the probability that all the cycles in π have length α (we assume that α divides a), which is, by Stirling's approximation (1),

$$\frac{1}{\alpha^{a/\alpha}(a/\alpha)!} \geq \frac{e^{a/\alpha}}{e\sqrt{2\pi}a^{a/\alpha}\sqrt{a/\alpha}}.$$

(While this estimate of p_1 probably can be improved, we do not attempt it here because the final probability of the team winning the game obtained with this estimate is perhaps not too far from what would be obtained by having a better bound for p_1. And also because it appears to be difficult. See, for instance, [4], where explicit formulas are derived for p_1 but to the best of our knowledge no usable asymptotic estimates are known.)

For $\beta \in [a]$, let $p_2(\beta)$ be the probability that $M \leq 2d\beta$. As noted earlier, M depends only on the subset of a occupied boxes and not on the labels on the slips; thus we can view our probability space as the uniform distribution over the $\binom{da}{a}$ subsets. We lower bound $p_2(\beta)$ by identifying a large set of subsets, each of these subsets has M at most $2d\beta$. We partition the set of boxes into a/β sets (called *groups*) of size $d\beta$ (assume that β divides a): Boxes with labels in $[0, d\beta - 1]$ and bins with labels in $[0, \beta - 1]$ belong to the first group, etc. We first restrict to the placements of slips in the boxes in which each group gets exactly β slips.

We make a further restriction to how, in each group, exactly β slips are placed in its boxes. There are $\binom{d\beta}{\beta}$ ways to place β (unlabeled) slips in the $d\beta$ boxes of a group. Let G be a group containing boxes in $[di, di + d\beta \ominus 1]$. Let us classify placements of slips as above in G according to the maximum of $\mathsf{surplus}[di, dj \ominus 1]$, where j is such that $dj \in [di, di + d\beta]$. In words, we look at the maximum of the surplus of sets containing a prefix the bins in G. Since there are β slips, that maximum is at most $\beta - 1$ and at least 0. So we get a partition of the set of placements into β subsets according to the maximum surplus. At least one of these sets, call it S, has size at least $\frac{1}{\beta}\binom{d\beta}{\beta}$. Denote the maximum surplus achieved in S by s. The restriction on the placements of slips in the boxes is that for each group, its placement comes from S.

In such placements no bin requires more than $2d\beta$ boxes to be opened. This is because whatever bin we start with in a group, we are going to have its surplus nonnegative in the same or the next group:

Proposition 3 *If the placements of slips in groups all come from the set S defined above, then the slip associated with a bin is either in its group or the next one.*

Proof. Let us fix a group G_1 and a bin B in it. Let G_2 be the group next to group G_1 in the cyclic order. We want to show that the slip associated with bin B is either in G_1 or G_2. Denote the index of the first box in G_1 by i_1, and that of the first box in G_2 by i_2. Let j be the index of a box in G_2 such that $\mathsf{surplus}[di_2, j] = s$. Let dx be the index of some box in G_1, we show that $\mathsf{surplus}[dx, j] \geq 0$, thus the slip associated with bin B_x occurs in $[dx, j]$.

By **S2** we have, $\mathsf{surplus}[dx, j] = \mathsf{surplus}[di_1, di_2 \ominus 1] - \mathsf{surplus}[di_1, dx \ominus 1] + \mathsf{surplus}[di_2, j]$. Now, by our restrictions on the placements of the slips, we have $\mathsf{surplus}[di_1, di_2 \ominus 1] = 0$, $\mathsf{surplus}[di_1, dx \ominus 1] \leq s$, and we also know that $\mathsf{surplus}[di_2, j] = s$. Hence, $\mathsf{surplus}[dx, j] \geq 0$, completing the proof of the proposition. ■

Now the number of assignments of slips to the boxes so that each group gets its assignment from S is

$$\left(\frac{1}{\beta}\binom{d\beta}{\beta}\right)^{a/\beta}.$$

For y a positive integer, we have,

$$\binom{dy}{y} = \frac{dy!}{y!(dy-y)!} \geq \frac{\sqrt{2\pi}(dy)^{dy+1/2}e^{-dy}}{\sqrt{2\pi}y^{y+1/2}e^{-y}\sqrt{2\pi}[(d-1)y]^{(d-1)y+1/2}e^{-(d-1)y}} \cdot \frac{e^{1/(12dy+1)}}{e^{1/(12y)}e^{1/(12(d-1)y)}}$$
$$= \frac{(dy)^{dy+1/2}}{\sqrt{2\pi}y^{y+1/2}[(d-1)y]^{(d-1)y+1/2}} \cdot \frac{e^{1/(12dy+1)}}{e^{1/(12y)}e^{1/(12(d-1)y)}}.$$

Similarly,

$$\binom{dy}{y} \leq \frac{(dy)^{dy+1/2}}{\sqrt{2\pi}y^{y+1/2}[(d-1)y]^{(d-1)y+1/2}} \cdot \frac{e^{1/(12dy)}}{e^{1/(12y+1)}e^{1/(12(d-1)y+1)}}.$$

By the above discussion we have,

$$p_2(\beta) \geq \frac{(\frac{1}{\beta}\binom{d\beta}{\beta})^{a/\beta}}{\binom{da}{a}} \geq \frac{a^{1/2}}{\beta^{3a/2\beta}}\left(\frac{d}{2\pi(d-1)}\right)^{\frac{a}{2\beta}-\frac{1}{2}} \cdot \frac{\exp(\frac{a}{\beta}[\frac{1}{12d\beta+1} - \frac{1}{12\beta} - \frac{1}{12(d-1)\beta}])}{\exp(\frac{1}{12da} - \frac{1}{12a+1} - \frac{1}{12(d-1)a+1})}$$
$$\geq \frac{a^{1/2}}{\beta^{3a/2\beta}}\left(\frac{d}{2\pi(d-1)}\right)^{\frac{a}{2\beta}-\frac{1}{2}} e^{-\frac{a}{6\beta^2}}.$$

If both events associated with $p_1(\alpha)$ and $p_2(\beta)$ happen then each player needs to open at most $2d\alpha\beta$ boxes before he finds his slip and answers correctly. Hence, if $2d\alpha\beta \leq b/k$ then the team wins with probability $\geq p_1(\alpha)p_2(\beta)$. Since $d = b/a$, condition $2d\alpha\beta \leq b/k$ is same as $\alpha\beta \leq a/2k$. We choose $\alpha = \sqrt{a/2k}$ and $\beta = \sqrt{a/2k}$ which are integers as we assumed in the hypothesis of the theorem that $a = 2kn^2$. Then, using $d > 1$ and lower bounding with the goal of getting a simple final expression,

$$\begin{aligned}
p_1(\sqrt{a/2k}) \cdot p_2(\sqrt{a/2k}) &\geq \frac{e^{\sqrt{2ka}}}{e\sqrt{2\pi}a^{\sqrt{2ka}}(2ka)^{1/4}} \cdot \frac{a^{1/2}}{(a/2k)^{\sqrt{9ka/8}}}\left(\frac{d}{2\pi(d-1)}\right)^{\sqrt{ka/2}-\frac{1}{2}} e^{-k/3} \\
&\geq \frac{1}{a^{\sqrt{2ka}}(2ka)^{1/4}a^{\sqrt{9ka/8}}8^{\sqrt{ka/2}}}2^{-k/(3\lg e)} \\
&\geq \frac{1}{2^{\sqrt{2ka}\lg a + \lg(2ka)/4 + \sqrt{9ka/8}\lg a + 3\sqrt{ka/2} + k/(3\lg e)}} \\
&\geq 2^{-9\sqrt{ak}\lg a - k}.
\end{aligned}$$

This completes the proof of the theorem.

3 Conclusion

Application to data structure lower bounds. The GM-games and the associated hypothesis were introduced by Gál and Miltersen for proving time-space trade-offs for the so-called *systematic data structures* for the well-known *substring search problem*. Here we briefly describe the problem, some relevant results, and what our result implies. For more details we refer to [3]; the definitions below are verbatim from there.

Substring search problem is the following. Given a string x in $\{0,1\}^n$ we want to store it in a data structure so that given a query string y of length m, we can tell whether y is a substring of x by inspecting the data structure.

A systematic data structure is a storage scheme satisfying $\phi(x) = x \cdot \phi^*(x)$ for some map ϕ^*, i.e., we require that the original data is kept "verbatim" in the data structure.

We work in the bit probe model of data structures (see [5] for a survey on this area). We are interested in proving trade-offs between the space taken by a systematic data structure for the substring search problem and the number of probes needed to answer queries in the worst case.

If a data structure for the substring search problem takes space s, then its *redundancy* r is defined by $s-n$. Assume that in the worst case number of bit probes needed to answer a query is t.

It was proved in [3] that if m and n are such that $2\log n+5 \le m \le 5\log n$, then we have $(r+1)t \ge \frac{1}{800}n/\log n$. In [3] the hypothesis mentioned in the introduction was claimed to imply a better bound of the form $t < n/\mathrm{poly}\log n \Rightarrow r > n/\mathrm{poly}\log n$. Our theorem implies that such an approach cannot give a trade-off better than $t < n/\mathrm{poly}\log n \Rightarrow r^2 > n/\mathrm{poly}\log n$ (here m and n are related as in the first line of this paragraph). We skip the details, as the methods are the same as in [3].

Variants. We briefly consider two variants of the GM-game. We can ask what happens if player I is allowed to make a worst case choice of the placement of the slips. We get at least two variants depending on whether the players are allowed to share random bits or not. In our first variant the team can generate a common random string (which is not seen by player I) to use in its startegy. In the second variant the team can not generate a common random string, but each individual player can generate his own random string.

The first variant is equivalent to the GM-game because the team can use the common randomness to choose a random permutation ϕ of $[0, b-1]$, and then follow the strategy for the GM-game, but this time when a player needs to look into box i, he actually looks into box $\phi(i)$. It is now easy to see that the probability that the team wins for this variant is the same as for the GM-game.

In the second variant the team wins with probability at most $(1/k)^n$. For proof consider a matrix with rows indexed by the players and the columns indexed by placements of the slips into the boxes. An entry in this matrix is the probability of the corresponding player reading the color of his slip when the input is the corresponding placement. For each row the average of its entries is at most $1/k$, because if player I chooses the placement at random as in the GM-game then a player has chance at most $1/k$ of reading his slip irrespective of his strategy. Hence there is a column with probabilities in it summing to $\le a/k$. If player I chooses the placement corresponding to this column as input then the team can win with probability at most $(1/k)^n$.

Open problems. Some interesting problems remain open: Improve the analysis of our strategy. Our analysis treated the cycles in π and M separately. We did not look at the cases where the team wins (using our strategy) because when a cycle in π is large corresponding evaluations of $m(i)$ do not require many boxes to be opened; or because i's for which evaluating $m(i)$ is costly lie in small cycles. Exploiting this dependence may lead to better analysis of our strategy.

And of course the main problem is to find optimal strategies for the GM game.

Acknowledgements. We thank the two anonymous referees, Sachin Lodha, and Peter Bro Miltersen for helpful comments on previous versions of this paper.

References

[1] T. Ebert, $<$ `http://www.cecs.csulb.edu/ ebert/` $>$.

[2] T. Ebert, W. Merkle, H. Vollmer, On the Autoreducibility of Random Sequences, *SIAM J. Comput.* 32 (2003), no. 6, 1542–1569.

[3] A. Gál and P. B. Miltersen, The cell probe complexity of succinct data structures, *Proceedings of 30th International Colloquium on Automata, Languages, and Programming (ICALP) 2003*, 332–344.

[4] S. Golomb and P. Gaal, On the number of permutations of n objects with greatest cycle length k, *Adv. in Appl. Math*, 20 (1998), no. 1, 98–107.

[5] P. B. Miltersen, Cell probe complexity - a survey, Advances in Data Structures (Pre-conference workshop of FSTTCS'99), available at $<$ `http://www.daimi.au.dk/~bromille/Papers/survey3.ps` $>$.

[6] H. Robbins, A remark on Stirling's formula, *Amer. Math. Monthly*, 62, (1955), 26–29.