**The Word Probe Model: Membership**
**Exposition by William Gasarch**

# 1 Introduction

This is an exposition of parts of [1].

**Def 1.1** $S_n$ is the set of all bijections from $[n]$ to $[n]$. Members of $S_n$ are called *permutations*.

**Def 1.2** Let $U, n, q \in \mathbb{N}$. A $(U, n; q)$-*WPDSY for membership* (WPDSY=Word Probe Data Structure-Yao. Henceforth we drop the "for membership") consists of the following.

1. A function *STOREPERM* that will, given $A \in \binom{[U]}{n}$, output $\sigma \in S_n$. We interpret this as follows. Let $A = \{a_1 < \cdots < a_n\}$. We store $A$ in (exactly) $n$ cells by putting $a_{\sigma(i)}$ into $CELL[i]$.
2. (Assume that the function in step 1 has been done.) A process that does the following: Given $u \in U$, the process generates $i_1$ and asks "What is CELL$[i_1]$?" Upon getting that answer the process then generates another index $i_2$ (which may depends on $u$ and CELL$[i_1]$) and asks "What is CELL$[i_2]$?" This goes on for $q$ rounds. Then, given the content of the requested cells, the process outputs YES or NO. If $u \in A$ then the output is YES, if $u \notin A$ then the output is NO.

**Note 1.3**

1. The function and process in Definition 1.2 need not be computable. Hence our lower bounds will be very strong.
2. The process in Definition 1.2 was adaptive: questions asked could depend on prior answers given. If all the questions are asked at once, that is called non-adaptive.

**Example 1.4** For all $U, n$ there is a $(U, n; \lceil \lg(n+1) \rceil)$-WPDSY. Take $A$ and sort it. Formally for all $A \in \binom{[U]}{n}$, $STOREPERM(A) = id$, the identity permutation. The query algorithm is binary search.

Is there an WPDSY that has $q < \lceil \lg(n+1) \rceil$? Here is a very cheap answer: If $U = n$ then there is a $(U, n; 0)$-WPDSY- the query algorithm always say YES. What if $n \ll U$?

We show the following:

1. Assume $U \geq 2n - 1$. If there is a $(U, n; q)$-WPDSY and a $\sigma \in S_n$ such that, for all $A \in \binom{[U]}{n}$, $STOREPERM(A) = \sigma$, then $q \geq \lceil \lg(n+1) \rceil$.
2. Assume $U \geq R(n, 2n-1, n!)$ (Ramsey number for $n$-hypergraphs, $2n-1$-sized monochromatic set, $n!$ colors). If there is a $(U, n; q)$-WPDSY then $q \geq \lceil \lg(n+1) \rceil$.

## 2 If the $(U, n; q)$-WPDSY uses SORTING

**Lemma 2.1** *Let $U, n, q \in \mathbb{N}$ such that $U \geq 2n - 1$. Assume there is a $(U, n; q)$-WPDSY such that all sets $A \in \binom{[U]}{n}$ are stored sorted. Imagine that the membership query "$n \in A$?" is asked. There exists a sequence of answers to the probes such that (1) the sequence is of length $\lceil \lg(n+1) \rceil - 1$, and (2) the sequence is consistent with both $n \in A$ and $n \notin A$. (Note: If this lemma is true then it is still true in some form if the universe is $\{a+1, \ldots, a+U\}$ and we make the membership query "$a + n \in A$?" it will still be true. We will need this when we use the induction hypothesis.)*

**Proof:**

We prove this by induction on $n$.

**Base Case:** $n = 2$. Then $U \geq 3$ and $\lceil \lg(n+1) \rceil = 2$. We simulate what happens when the membership query "$2 \in A$?" is asked. There are two cases.

1. If the first probe asked is $CELL[1]$ then answer $CELL[1] = 1$. Since $CELL[2]$ could be 2 or 3, a second question must be asked.

2. If the first probe asked is $CELL[2]$ then answer $CELL[2] = 3$. Since $CELL[1]$ could be 1 or 2, a second question must be asked.

*Induction Hypothesis:* The lemma is true for $n' < n$.

*Induction Step:* Assume $U \geq 2n - 1$ and there is a $(U, n; q)$-WPDSY. We take $n$ to be even ($n$ odd is similar).

We simulate what happens when the membership query "$n \in A$?" is asked. There are two cases (we will only do one of them, the other is similar). Assume the first probe is $p$, so we are asking "What is in $CELL[p]$?"

**Case 1:** $p \leq n/2$. Then answer $p$. Hence we know that the $CELL$ array looks like this:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | $\cdots$ | $p-1$ | $p$ | $p+1$ | $\cdots$ | $n/2$ | $n/2+1$ | $\cdots$ | $n$ |
| 1 | 2 | 3 | $\cdots$ | $p-1$ | $p$ | ? | $\cdots$ | ? | ? | $\cdots$ | ? |

Look at $CELL[n/2]$, ..., $CELL[n]$. It is consistent with the $CELL[p] = p$ for this subarray to contain any $n/2$-sized subset of $\{n/2+1, \ldots, U\}$. Hence we can view the current situation as follows: We are looking for $n$, the universe is $\{n/2+1, \ldots, U\}$ which is size $U - n/2$, and the set is size $n/2$. By shifting everything down by $n/2$ this is equivalent to looking for $n/2$ in the universe $\{1, \ldots, U - n/2\}$ and the set is size $n/2$. Note that

$$U - n/2 \geq 2(n/2) - 1$$

Hence the induction hypothesis applies. So there is a sequence of answers to probes of length $\lceil \lg(n/2+1) \rceil - 1$ which does not reveal $n \in A$ or not. Add this to the 1 probe made so far an you have a sequence of length

$$1 + \lceil \lg(n/2+1) \rceil - 1 \leq \lceil \lg(n+1) \rceil - 1.$$

∎

The following follows easily.

**Theorem 2.2** *Let $U, n, q \in \mathbb{N}$ such that $U \geq 2n-1$.*

1. *Assume there is a $(U, n; q)$-WPDSY such that all sets $A \in \binom{[U]}{n}$ are stored sorted. Then $q \geq \lceil \lg(n+1) \rceil$.*
2. *Assume there is a $(U, n; q)$-WPDSY and a $\sigma \in S_n$ such that for all sets $A \in \binom{[U]}{n}$, $STOREPERM(A) = \sigma$. Then $q \geq \lceil \lg(n+1) \rceil$.*

# 3 If $n \ll U$ then $q \geq \lceil \lg(n+1) \rceil$

**Def 3.1** Let $R(k, r, t)$ be the least $R$ such that for all $t$-colorings of $\binom{[R]}{r}$ there exists $X \subseteq [R]$, $|X| = k$, such that $X$ is monochromatic. Note that such an $R$ exists by Ramsey's theorem.

**Theorem 3.2** *If $U \geq R(2n-1, n, n!)$ and there is a $(U, n; q)$-WPDSY. Then $q \geq \lceil \lg(n+1) \rceil$.*

**Proof:** Assume $U \geq R(2n-1, n, n!)$ and that there is a $(U, n; q)$. Define the following coloring of $\binom{[U]}{n}$:

The color of $A$ is $STOREPERM(A)$.

Let $H$ be the monochromatic set of size $2n-1$. By the definition of the coloring there is a $\sigma \in S_n$ such that, for all $A \in \binom{H}{n}$, $STOREPERM(A) = \sigma$. Restrict the $WPDSY$ to $H$. By Theorem 2.2 $q \geq \lceil \lg(n+1) \rceil$. ∎

# References

[1] A. Yao. Should tables be sorted? *Journal of the ACM*, 28, 1981. Earlier version in FOCS78.