Finding Witnesses by Peeling

Yonatan Aumann[1] Moshe Lewenstein[1] Noa Lewenstein[2]
Dekel Tsur[3]

[1]Bar-Ilan University

[2]Netanya Academic College

[3]Ben-Gurion University

Witness Problems

- The k-mismatches problem is given P and T, to find all substrings T' of T with $|T'| = |P|$ and $\mathrm{Hamming}(T', P) \leq k$.
- Suppose that we also want to find for every T' that satisfies the requirements above, all the indices i s.t. $T'[i] \neq P[i]$.
  This problem can be solved in $O(kn)$ time using the kangaroo method.

Witness Problems

- The k-mismatches problem is given P and T, to find all substrings T' of T with $|T'| = |P|$ and $\mathrm{Hamming}(T', P) \leq k$.
- Suppose that we also want to find for every T' that satisfies the requirements above, all the indices i s.t. $T'[i] \neq P[i]$.
  This problem can be solved in $O(kn)$ time using the kangaroo method.
- In the k-matches problem, we need to find all substrings T' with $|T'| = |P|$ and $\mathrm{Hamming}(T', P) \leq |P| - k$. We also want for each such T', all the indices i s.t. $T[i] = P[i]$.

k-Aligned Ones

The k-aligned ones problem

Input Strings P and T over alphabet $\Sigma = \{0,1\}$ of lengths m and n.

Output For each substring T' of T with $|T'| = m$ and $\sum_{i=1}^{m} P[i]T'[i] \leq k$, output all indices i s.t. $P[i] = T[i] = 1$
(all locations of 1-against-1 when P is aligned with T')

Example

$T = 101101, \quad P = 10111, \ k = 2$
For $T' = T[1..5] = 10110, \ \sum P[i]T'[i] = 3$
For $T' = T[2..6] = 01101, \ \sum P[i]T'[i] = 2$, output is $3, 5$.

Results for k-Aligned Ones

	Deterministic	Randomized
Previous	$O(nk^3 \cdot \log m \cdot \log k)$ [AF95]	$O(nk \cdot \log^4 m)$ [Mut95]
New	$O(nk \cdot \log^{O(1)} m)$	$O(nk \cdot \log m + n \cdot \log^2 m \cdot \log k)$

$$n = |T|$$
$$m = |P|$$

The Reconstruction Problem

Unknown sets $S_1, \ldots, S_n \subseteq U$, with $|U| = m$.

Goal: Reconstruct k_i elements from S_i using few queries of the following types:

- $\textsc{ISize}(A) = \langle |S_1 \cap A|, \ldots, |S_n \cap A| \rangle$.
- $\textsc{ISum}(A) = \langle \sum_{u \in S_1 \cap A} u, \ldots, \sum_{u \in S_n \cap A} u \rangle$.

k-reconstruction $k_i = \min(k, |S_i|) = \begin{cases} |S_i| & \text{if } |S_i| \leq k \\ k & \text{otherwise} \end{cases}$

bounded k-reconstruction $k_i = \begin{cases} |S_i| & \text{if } |S_i| \leq k \\ 0 & \text{otherwise} \end{cases}$

Solving the k-Aligned Ones Problem

Let P and T be strings of lengths m and n.

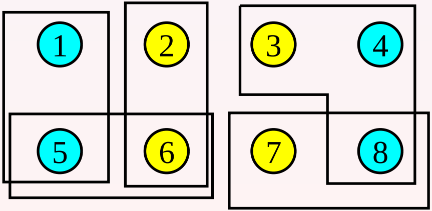
- Define $U = \{1, \ldots, m\}$.
- Define $S_i = \{j \in U : T[i-1+j] = P[j] = 1\}$.
  (all locations of 1-against-1 when P is aligned with the i-th substring of T of length m)
- For a set $A \subseteq U$, let b_A, c_A be vectors of length m:
  - $b_A[i] = \begin{cases} 1 & \text{if } i \in A \\ 0 & \text{otherwise} \end{cases}$.
  - $c_A[i] = \begin{cases} i & \text{if } i \in A \\ 0 & \text{otherwise} \end{cases}$
- The convolution $T \circ b_A$ gives $\text{ISIZE}(A)$.
- The convolution $T \circ c_A$ gives $\text{ISUM}(A)$.

k-Separators

Definition

Let $S \subseteq U$ and $F =$ set of subsets of U.
F is a k-separator for S if there are sets $A_1, \ldots, A_{\min(k,|S|)} \in F$ s.t.

1. for each i, $|S \cap A_i| = 1$.
2. for $i \neq j$, $S \cap A_i \neq S \cap A_j$.

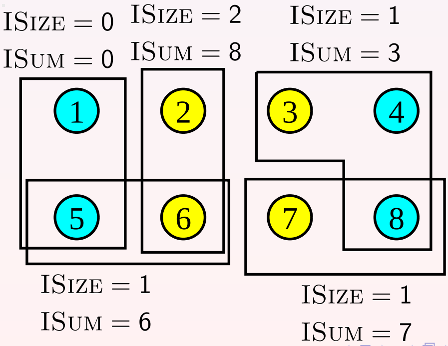


k-Separators

Definition

Let $S \subseteq U$ and $F =$ set of subsets of U.
F is a *k-separator for S* if there are sets $A_1, \ldots, A_{\min(k,|S|)} \in F$ s.t.

1. for each i, $|S \cap A_i| = 1$.
2. for $i \neq j$, $S \cap A_i \neq S \cap A_j$.

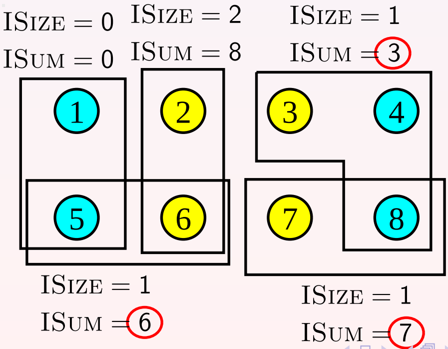


k-Separators

Definition

Let $S \subseteq U$ and $F =$ set of subsets of U.
F is a *k-separator for S* if there are sets $A_1, \ldots, A_{\min(k,|S|)} \in F$ s.t.

1. for each i, $|S \cap A_i| = 1$.
2. for $i \neq j$, $S \cap A_i \neq S \cap A_j$.



k-Separators

Let F be a k-separator for $S_1, \ldots, S_n$.
The k-reconstruction problem is solved as follows:

```
For each A ∈ F do
    z_A^1, ..., z_A^n ← ISize(A)
    m_A^1, ..., m_A^n ← ISum(A)
For i = 1, ..., n do
    For each A ∈ F do
        If z_A^i = 1 then output m_A^i
```

Construction of k-Separators

Theorem

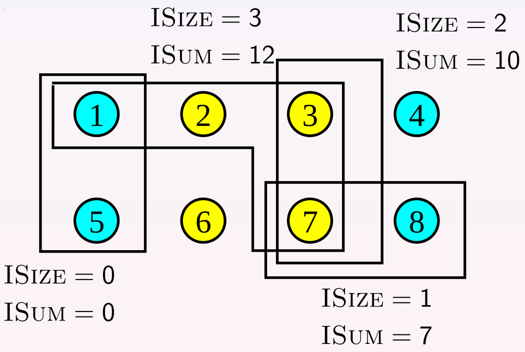
There is a deterministic algorithm s.t. for every collection of sets $S_1, \ldots, S_n$ with $|S_i| \geq k \log^c m$ for all i (where c is some constant), it constructs a k-separator F for $S_1, \ldots, S_n$ with $|F| = O(k \cdot \text{polylog}(mn))$.
The algorithm makes $O(k \cdot \text{polylog}(mn))$ calls to the procedure $\textsc{ISize}(\cdot)$.

Proof.

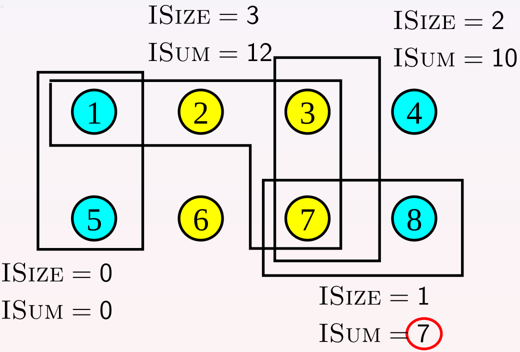
Follows from results of Alon & Naor (1996) and Ta-Shma, Umans, & Zuckerman (2001). □

k-separator allows us to solve the k-reconstruction problem when $S_1, \ldots, S_n$ are sets of size $\geq k \cdot \log^c m$.

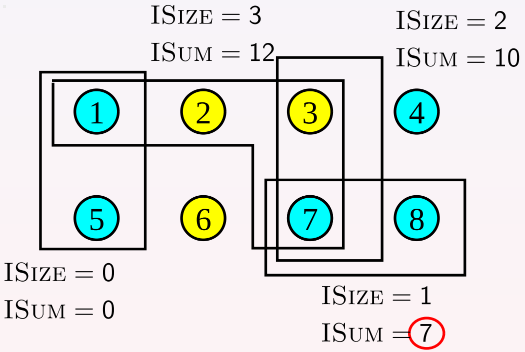
The Peeling Procedure



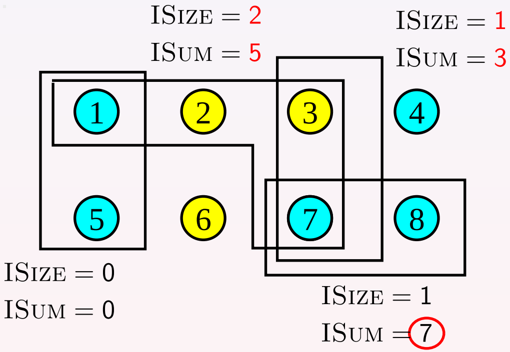
The Peeling Procedure



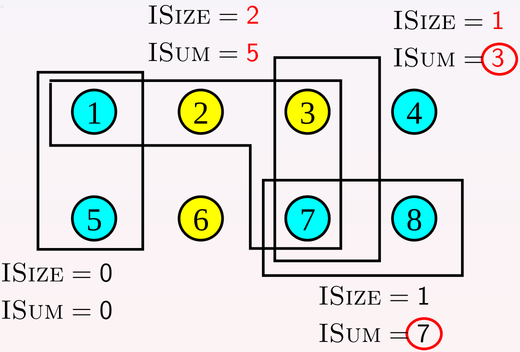
The Peeling Procedure



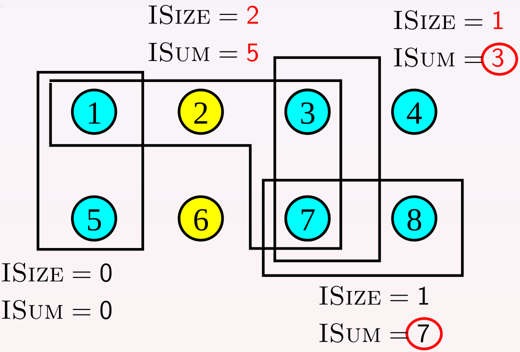
The Peeling Procedure



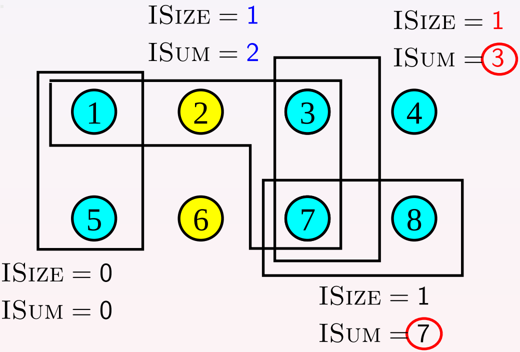
The Peeling Procedure



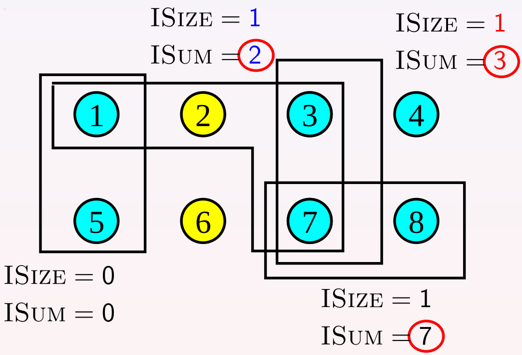
The Peeling Procedure



The Peeling Procedure



The Peeling Procedure



The Peeling Procedure

For each $A \in F$ do

  $z_A^1, \ldots, z_A^n \leftarrow \textsc{ISize}(A)$

  $m_A^1, \ldots, m_A^n \leftarrow \textsc{ISum}(A)$

For $i = 1, \ldots, n$ do

  $\mathrm{Ones} \leftarrow \{A \in F : z_A^i = 1\}$.

  While Ones is not empty do

    Choose any $A \in \mathrm{Ones}$ and remove A from Ones

    Output m_A^i

    For each $B \in F$ such that $m_A^i \in B$ do

      $z_B^i \leftarrow z_B^i - 1$

      $m_B^i \leftarrow m_B^i - m_A^i$

      If $z_B^i = 1$ then add B to Ones

k-Peelers

Definition

Let $P = (A_1, \ldots, A_k)$ be a sequence of sets. P is a peeling sequence for S if for $i = 1, \ldots, k$, $|A_i \cap (S - \cup_{j=1}^{i-1} A_j)| = 1$.

Definition

Let S be a set and F a collection of sets. We say that F is a k-peeler for S if F contains a peeling sequence for S of length $k' = \min(k, |S|)$.

Theorem

If F is a k-peeler for S then the peeling procedure finds $\min(k, |S|)$ *distinct elements of S.*

Construction of k-Peelers

Theorem (Indyk 2002)

There is a deterministic algorithm s.t. given U and k, it constructs a collection of sets F such that

1. *For every set $S \subseteq U$ of size at most k, F is a k-peeler of S.*
2. $|F| = O(k \cdot \text{polylog}(m))$.

Solving the k-Reconstruction Problem

$t =$ time for computing $\textsc{ISize}(A)$ or $\textsc{ISum}(A)$.

Theorem

The bounded k-reconstruction problem can be solved in $O(tk \cdot \text{polylog}(m))$ time.

Proof.

Use a k-peeler and the peeling procedure. □

Solving the k-Reconstruction Problem

$t =$ time for computing $\textsc{ISize}(A)$ or $\textsc{ISum}(A)$.

Theorem

The bounded k-reconstruction problem can be solved in $O(tk \cdot \text{polylog}(m))$ time.

Proof.

Use a k-peeler and the peeling procedure. □

Theorem

The k-reconstruction problem can be solved in $O(tk \cdot \text{polylog}(mn))$ time.

Proof.

Sets of size $\leq k \cdot \log^c m$: use a $k \cdot \log^c m$-peeler.

Sets of size $> k \cdot \log^c m$: use a k-separator. □

Randomized Algorithms for k-Reconstruction

Theorem

The bounded k-reconstruction problem can be solved in time

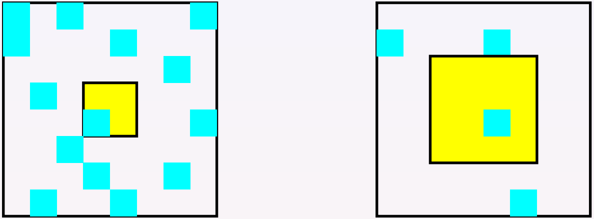
- $O(tk \cdot \mathrm{polylog}(m))$ — *deterministic.*
- $O(t(k + \log k \cdot \log n) + nk \log(mn))$ — *randomized.*

Theorem

The k-reconstruction problem can be solved in time

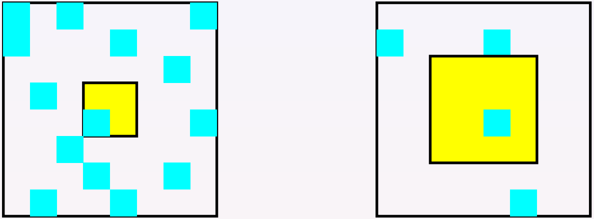
- $O(tk \cdot \mathrm{polylog}(mn))$ — *deterministic.*
- $O(t(k(\log m + \log k \log\log m) + \log n \log m) + nk \log(mn))$ — *randomized.*

Randomized Construction of k-Peelers



- A p-random set is a subset of U that contains each element of U with probability p.
- For $S \subseteq U$ and $p \approx \frac{1}{|S|}$, if A is p-random then $|A \cap S| = 1$ with probability $\geq \frac{1}{6}$.

Randomized Construction of k-Peelers



- A *p-random set* is a subset of U that contains each element of U with probability p.
- For $S \subseteq U$ and $p \approx \frac{1}{|S|}$, if A is p-random then $|A \cap S| = 1$ with probability $\geq \frac{1}{6}$.
- Suppose that $|S| = k$. We build collections $F^{(0)}, F^{(1)}, \ldots, F^{(\lceil \log k \rceil)}$ of sets, where $F^{(i)}$ contains $\frac{1}{2^i}$-random sets.
- Using $F^{(\lceil \log k \rceil)}$, S is peeled down to a set of size $k/2$.
- Using $F^{(\lceil \log k \rceil - 1)}$, S is peeled down to a set of size $k/4$.

Application — All-Pairs Shortest Paths

$G =$ an undirected graph.
$M(n) =$ time for multiplying two $n \times n$ boolean matrices.

- The distance between every pair of vertices can computed in $O(M(n) \log n)$ time [Alon et al. 92, Seidel 95]

Application — All-Pairs Shortest Paths

$G =$ an undirected graph.
$M(n) =$ time for multiplying two $n \times n$ boolean matrices.

- The distance between every pair of vertices can computed in $O(M(n) \log n)$ time [Alon et al. 92, Seidel 95]
- Finding a shortest path between any u and v can be done in $O(\mathrm{dist}(u, v))$ time, after preprocessing whose time is
  - $O(M(n) \log n)$ randomized [Alon et al. 92, Seidel 95].
  - $O(M(n)\,\mathrm{polylog}(n))$ deterministic [Alon & Naor 96].

The results above rely on algorithms for the 1-reconstruction problem.

Application — All-Pairs Shortest Paths

$G =$ an undirected graph.
$M(n) =$ time for multiplying two $n \times n$ boolean matrices.

- The distance between every pair of vertices can computed in $O(M(n) \log n)$ time [Alon et al. 92, Seidel 95]
- Finding a shortest path between any u and v can be done in $O(\mathrm{dist}(u, v))$ time, after preprocessing whose time is
  - $O(M(n) \log n)$ randomized [Alon et al. 92, Seidel 95].
  - $O(M(n)\,\mathrm{polylog}(n))$ deterministic [Alon & Naor 96].

  The results above rely on algorithms for the 1-reconstruction problem.
- **New result:** Finding k shortest paths between u and v can be done in $O(k \cdot \mathrm{dist}(u, v))$ time, after preprocessing whose time is
  - $O(M(n)(k + \log n))$ randomized.
  - $O(M(n) \cdot k \cdot \mathrm{polylog}(n))$ deterministic.