A Brief History Of

History Independent Data Structures

Maverick Woo
Credit: South China Morning Post
Oblivious Data Structures
Daniele Micciancio

Oblivious data structures: Applications to cryptography

- STOC 1997

An attempt to solve the privacy problem in incremental cryptography
M. Bellare, O. Goldreich, S. Goldwasser

()-> Incremental Cryptography
and Application to Virus Protection

- STOC 1995
**Why Incremental?**

Most cryptographic primitives act on the document as a whole.

*MD5 Signature*

<table>
<thead>
<tr>
<th>T</th>
<th>Th</th>
<th>Th</th>
</tr>
</thead>
<tbody>
<tr>
<td>H9ECE18C950AFBFA 6B0FDBFA4FF731D3</td>
<td>EEE89A88B45DD351 D9EC0EB4ACCE66CE</td>
<td>A4704FD35F030828 7F29378A3ECCF5FE</td>
</tr>
</tbody>
</table>

Re-compute from scratch is wasteful

Ideally, the running time should be $O(f(|\text{change}|))$. 

This is a survey on the area of history independent data structures. I will show you how this area developed in the last decade [...].

This is a survey on the area of history independent data structures. I will show you how this area developed in the last decade [...].
The Free Food Cam

At 15 frame per second, 1 M pixel per frame, 3 byte per pixel

http://zimbs4.srv.cs.cmu.edu/coke/ffc.html

(Google for the phrase “free food cam”)
How do you know these pictures are authentic?
Tree Signatures [BGG95]

- Represent the document as a 2-3 Tree whose leaves contain constant-size block of characters.

- Non-incrementally sign each leaf and each internal node.

- Each change only takes $O(\log n)$ signature updates.
Privacy of Tree Signatures

An Apparent Paradox

The very nature of digital signature ⇒
there can be no secret about the document
and its signature...

How can privacy be an issue of when your original intent is
to publish the document and its signature?
Oct 2000: The Wall Street Journal reports that a candidate running for the U.S. Senate began receiving anonymous emails containing messages written in MS Word criticizing and attacking the candidate. A savvy aide looked at the document properties and discovered they were authored by the chief-of-staff of the opposing party.

Feb 2003: A dossier on Iraq’s security and intelligence organizations, cited by Colin Powell and published by 10 Downing Street, is discovered to have been plagiarized from a U.S. researcher on Iraq. Since the dossier was published in MS Word format, researchers also discovered the four people in the British government who edited the document. They were subsequently called to Parliament for a hearing.

Mar 2004: SCO Group, seller of UNIX and Linux, sent out a warning letter to 1,500 of the world’s largest companies threatening legal liability for using Linux if they failed to obtain a license from the Utah-based company. After filing suit against Daimler-Chrysler, metadata in a MS Word document revealed that the SCO’s attorneys had originally identified Bank of America as the defendant.

From http://www.abanet.org/tech/ltrc/publications/metadata.html
Metadata in Tree Signatures

Represent the document as a 2-3 Tree whose leaves contain constant-size block of characters

Non-incrementally sign each leaf and each internal node

Where is the metadata?

How can a 2-3 tree leak information?
The Wedding Guest List

As you may know, some of us are getting married soon!

☞ As hard-core computer scientists, of course you will maintain the guest list in sorted order (using an object-oriented multimedia relational XML database).

Congrads!

A, B

C

D

E

A, B

C

D

E
Manual Sorting Is Hard
And as the guest list gets compiled, there will always be someone who gets added to the list LAST...

“\text{You added me} \textit{after} \text{you have added} \{\text{Foo}\}?”
Between B and D is C...ontention

Initial

- Insert("D")

Initial

- Insert("B")
Moni Naor, Vanessa Teague

Anti-persistence:
History Independent Data Structures

- STOC 2001
Daniele Micciancio

Oblivious data structures:
Applications to cryptography
- STOC 1997

An attempt to solve the privacy problem in incremental cryptography
Oblivious Data Structures

Informal definition

A data structure is said to be *oblivious* if it does not give out any knowledge about the sequence of update operations that have been applied to it other than the final result of the operations.
Formal Definition [Mic97]

Let $M$ be a set of operations, and $S$ be a set of algorithms implementing them.

We say $S$ is **oblivious** if:

- for any two sequences of operations $p_1, p_2, \ldots, p_n$ and $q_1, q_2, \ldots, q_m$ that lead to the same set of values, the execution of these sequences have identical output probability distributions.
Oblivious 2-3 Tree

Issue: How do 2-3 trees “leak”?
- Degrees of nodes give out too much information

Solution: Randomize the degrees!
- Degrees should split uniformly between 2 and 3
Results in [Mic97]

Oblivious 2-3 Trees

- Insertion and Deletion in expected $O(\log n)$ time whp
- Search in worst-case $O(\log n)$ time

Incremental Signature Scheme

- Security: Tamper Proof, as defined in [BGG95]
- Privacy: Private (hmm... we will see :P)
- Expected $O(\log n)$ signature updates per change whp
For More Information

- FOCS Tidbits
  - PDF
  - PPT
  - 2003-10-29 Theory Lunch

- Understanding Combinatorial Preconditioners
  - PDF
  - PPT
  - 2003-09-10 Theory Lunch

- Communication Complexity 101
  - PDF
  - PPT
  - 2003-03-23 Theory Workshop

- SODA Bites
  - PDF
  - PPT
  - 2003-01-29 Theory Lunch

- Minimizing Congestion in General Networks
  - PDF
  - PPT
  - 2002-10-16 Theory Lunch

- A Collection of Probability Tricks
  - PDF
  - PPT
  - 2002-09-27 Theory Workshop

- MST and The Discrepancy Method
  - PDF
  - PPT
  - 2002-08-30 Theory Workshop

- Oblivious Search Trees
  - PDF
  - PPT
  - 2002-03-06 Theory Lunch

- Space Efficient Finger Search
  - PDF
  - PPT
  - 2001-10-31 Theory Lunch

- Ideas On Treaps
  - PDF
  - PPT
  - 2001-05-02 Theory Lunch
History Independence
Are We Done Yet?

In an oblivious 2-3 tree, the degrees indeed don’t tell you anything about the update sequence.

What else can leak information?
Interface Matters

Dictionary ADT

- Insert(key)
- Delete(key)
- Search(key)

Does this allow you to ask

“is $k_1$ is inserted some time before $k_2$?”

Principle

When privacy is concerned, if some piece of information cannot be retrieved via the legitimate interface of a system, then it should not be retrievable even with full access to the system.
Moni Naor, Vanessa Teague

Anti-persistence: History Independent Data Structures
- STOC 2001
Memory Representation

Most memory allocators have a tendency to allocate newer objects at higher addresses.

Shape

Memory
History Independence [NT01]

Definition

A data structure implementation is history independent if any two sequences $S_1$ and $S_2$ that yield the same content induce the same distribution on the memory representation.

Similar to Micciancio’s definition, except this is specified on memory representation.
Single vs. Multiple Observations

Thief

Hacker
Definition

Let $P_1 = \{i_1^1, i_2^1, \ldots, i_l^1\}$ and $P_2 = \{i_1^2, i_2^2, \ldots, i_l^2\}$ be two lists of points such that for all $b \in \{1, 2\}$ and $1 \leq j \leq l$, we have that $1 \leq i_b^j \leq |S_b|$ and the content of data structure following $S_1$ up to $i_1^j$ and $S_2$ up to $i_2^j$ are identical.

A data structure implementation is **strongly history independent** if for any $S_1, P_1, S_2, P_2$ the distributions of the memory representations at the points $i_1^j$ and $i_2^j$ are identical.
Strong History Independence [NT01]

P₁
S₁
D₁

P₂
S₂
D₂

...
Weak History Independence [NT01]

Definition

A data structure implementation is *weakly history independent* if any two sequences \( S_1 \) and \( S_2 \) that yield the same content induce the same distribution on the memory representation.

Similar to definition [Mic97], except this is specified on memory representation.
Checkpoint

✓ How HIDS got started
✓ Why obliviousness is inadequate
✓ What WHI and SHI are

The rest of this talk:
- WHI is bearable
- SHI is hard
Results in [NT01]

Besides the definitions,

- WHI object allocators
  - Applied to Dynamic perfect hashing

- SHI hash table
  - Currently does not support deletion

- WHI Union-Find (sketch)
Incremental Card Shuffling

Insert(new card x)
- Put the x at the end
- Swap x with another card, chosen u.a.r.

Delete(card at position j)
- Swap card with the last card
- Discard the (new) last card

Lemma
The permutation of cards remains random with these procedures.
Fixed Size Object Allocator

Turn any bounded-indegree, fixed-size record oblivious data structure into WHI

- Record allocation takes $O(1)$ time
- Need to be careful when moving blocks

Examples
- Treaps
- Oblivious 2-3 Trees
Variable Size Object Allocator

Create buckets for objects of geometrically increasing size

\( B_1, B_2, B_4, B_8 \ldots \) (exercise :P)

Record allocation takes \( O(s \log s) \) time

From here, adapt the dynamic perfect hashing into WHI
Going Back To Card Shuffling

Insert(new card x)
- Put the x at the end
- Swap x with another card, chosen u.a.r.

Don’t store the randomness and we get WHI

But that also rules out SHI
Example

\[ \text{D}_1 \]

- Ins(4)
  - 3, 1, 2
- Ins(5)
  - 3, 4, 2, 1
- Ins(5)
  - 3, 4, 5, 1, 2
- Del(4)
  - 3, 2, 5, 1

\[ \text{D}_2 \]

- Ins(5)
  - 3, 1, 2
- Ins(5)
  - 3, 5, 2, 1
Ordered Hashing

In open addressing, how do you break tie in a collision?

- Resolve by alphabetical order

O. Amble, D. Knuth

- Ordered hash tables

- The Computer Journal 17(2), 1974
SHI Hash Table

Looks like Ordered Hashing on Steroid

Define a priority function $p(i, x, y)$:
$$\{0, \ldots, N-1\} \times U \times U \rightarrow \{\text{True, False}\}$$

Round # \hspace{1cm} Contestants

Theorem

If $p(i, x, y)$ defines a total order, then the hash table is SHI.
**SHI Hash Table**

**Youth-rules**

true if $\text{age}(i,x) < \text{age}(i,y)$

$p(i, x, y) = \begin{cases} 
\text{true if } \text{age}(i,x) = \text{age}(i,y) \text{ and } x < y \\
\text{false otherwise}
\end{cases}$

**Theorems**

For any sequence of insertions the expected amortized insertion probe-time for an element is $1/(1-\alpha)$.

For any element, the expected probe-time for successful or unsuccessful search is $1/(1-\alpha)$.
Observation in [NT01]

“It is interesting to note that all the SHI data structures we have found have the property that each data structure content has a unique representation.”

 Państwo: 

SHI ⇒ Unique Representation?
SHI $\Rightarrow$ Unique Representation
J. Hartline, E. S. Hong, A. E. Mohr, W. R. Pentney, E. C. Rocke

Characterizing History Independent Data Structures

- ISSAC 2002
Results in [H.02]

A (somewhat) simpler definition of WHI and SHI

Show \( \text{SHI} \Rightarrow \text{Unique Representation} \)

Relax SHI to a new definition: \( \text{SHI}^* \)

The adversary can distinguish between empty and non-empty sequences of operations

Show how to resize WHI data structures
Hysteresis

Capacity

0

2n

n

0.75n

n

#Items

Your wedding is running out of budget?
Niv Buchbinder, Erez Petrank

Lower and Upper Bounds on Obtaining History Independence

- Crypto 2003
Results In [BP03]

SHI \Rightarrow Unique Representation

Predated by [H.02]

SHI Heap Lowerbound (in the comparison-based model)

Insert, Extract, Increase-Key: $\Omega(n)$

WHI Heap Upperbound

Build-Heap: $O(n)$; Increase-Key: $O(\log n)$

Insert and Extract-Max: expected $O(\log n)$
Heap Review

Binary Heap

10
9
8
3
6
2

5

7
4
1

10 9 7 8 5 4 1 3 6 2
Heapify

Heapify⁻¹
Heap $\rightarrow$ Permutation

Build-Heap$^{-1}(H)$

- If size of $H$ is 1, Return $H$
- Choose a node $i$ u.a.r. among all nodes in $H$
- $H \leftarrow \text{heapify}^{-1}(H, i)$
- Return $\text{MakeNode}(\text{root}(H),$
  
  $\text{Build-Heap}^{-1}(H_L),$
  $\text{Build-Heap}^{-1}(H_R))$
**WHI Heap**

\[ \Omega_1 \]: Pick u.a.r. a heap among all possible heaps with values \( v_1, \ldots, v_n \).

\[ \Omega_2 \]: Pick u.a.r. a permutation on the values \( v_1, \ldots, v_n \).
   
   Place them in an (almost) full tree and call Build-Heap.

**Theorem**

\( \Omega_1 \) and \( \Omega_2 \) are actually the same distribution.
Simulation

Build-Heap \[\rightarrow\] Build-Heap\(^{-1}\]

Insert

Build-Heap

8, 2, 4, 5, 3, 11, 10, 1, 7, 6, 9

8, 2, 4, 5, 3, 9, 10, 1, 7, 6

10, 9, 7, 8, 5, 4, 1, 3, 6, 2

Dynamizing Static Algorithms with Applications to Dynamic Trees and History Independence

- SODA 2004

Also [ABH02], [ABH03]
Dynamiziation

10
9
7
8
5
4
3
6
2

Build-Heap 10,9,7,8,5,4,1,3,6,2

Permute 8,2,4,5,3,9,10,1,7,6

Insert 8,2,4,5,3,11,10,1,7,6,9

Build-Heap 8,2,4,5,3,11,10,1,7,6,9
Uniquely Represented Dictionaries
Lawrence Snyder

On Uniquely Represented Data Structures

- FOCS 1977
### Motivating Questions

<table>
<thead>
<tr>
<th>Data Structure</th>
<th>Uniquely Represented</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linked list</td>
<td>yes</td>
<td>$O(n)$</td>
</tr>
<tr>
<td>Heap</td>
<td>yes</td>
<td>$O(n)$</td>
</tr>
<tr>
<td>AVL tree</td>
<td>no</td>
<td>$O(\log n)$</td>
</tr>
<tr>
<td>2-3 tree</td>
<td>no</td>
<td>$O(\log n)$</td>
</tr>
</tbody>
</table>

- 🧶 unique representation $\Rightarrow$ linear cost? ×
- 🧶 logarithmic cost $\Rightarrow$ redundant representation? ✓
Jellyfish

Update and Search take $O(n^{1/2})$ time
Snyder’s Lowerbound

Theorem

If a search tree uniquely represents its data, then

$$\max(\text{Search}(n), \text{Insert}(n), \text{Delete}(n)) = \Omega(\sqrt{n})$$

“The argumentation of this lemma is quite involved and will not be presented here.”
[A091]

Arne Andersson, Thomas Ottmann

New Tight Bounds on Uniquely Represented Dictionaries

- FOCS 1991
Fig. 6. A jump list containing 74 elements and the embedded binary search tree.
Fig. 7. Performing an insertion at position 41 in the data structure of Figure 6.
Rajamani Sundar, Robert E. Tarjan

Unique Binary Search Tree Representations and Equality-testing of Sets and Sequences

- STOC 1990
Future
Open Problems

- SHI Hash table with deletion (Working on it)
- SHI* seems to be a very reasonable definition
  - See what can be done in this model
Google is Hiring

Amitabh Sinha

Ke Yang
Google

Nikhil Bansal
IBM Research

2004-5-15
• They are hiring!

• Talk to Ke Yang if you are interested...

  yangke@google.com

  (917) 881-2797

  He’ll be around until 2004-Oct-1.

  Ask him for a GMail account.
Color Test