Lecture 2
Reconstruction Attacks

Today:

1. Census Reconstruction [Garfinkel, Abowd, Martindale '19]
2. Dinur-Nissim Database Reconstruction [Dinur-Nissim '03]
3. DN Reconstruction in Practice [Cohen-Nissim '20]

Census Reconstruction

Setup

(age, sex, race) microdata

sex: male / female; race: black / white

Ages: A, B, C ($A \leq B \leq C$)

Constraints

- Whole numbers
- $0 \leq A \leq B \leq C \leq 125$

$(126)(126)(126) = 126^3 \approx 300k$

$B = 30$

$\frac{1}{3}(A+B+C) = 44 \rightarrow \frac{1}{3}(A+C) = 34$

$A = 0, B = 30, C = 102$

$A = 1, B = 30, C = 101$, →

31 options

Statistic	Group	Age Count	Age Median	Age Mean
1A	Total Population	7	30	38
2A	Female	4	30	33.5
2B	Male	3	30	44
2C	Black or African American	4	51	48.5
2D	White	3	24	24
3A	Single Adults	(D)	(D)	(D)
3B	Married Adults	4	51	54
4A	Black or African American Female	3	36	36.7
4B	Black or African American Male	(D)	(D)	(D)
4C	White Male	(D)	(D)	(D)
4D	White Female	(D)	(D)	(D)
5A	Persons Under 5 Years	(D)	(D)	(D)
5B	Persons Under 18 Years	(D)	(D)	(D)
5C	Persons 64 Years or Over	(D)	(D)	(D)

Note: Married persons must be 15 or over

A	B	C	A	B	C	A	B	C
1	30	101	11	30	91	21	30	81
2	30	100	12	30	90	22	30	80
3	30	99	13	30	89	23	30	79
4	30	98	14	30	88	24	30	78
5	30	97	15	30	87	25	30	77
6	30	96	16	30	86	26	30	76
7	30	95	17	30	85	27	30	75
8	30	94	18	30	84	28	30	74
9	30	93	19	30	83	29	30	73
10	30	92	20	30	82	30	30	72

Attack

1. Generate constraints
2. Find a feasible point
   - NP Hard
   - SAT Solvers. Integer Programming

Real Attack

Linkage Attack

Tweetorial by Abowd

Differential Privacy

Dinur-Nissim Database Reconstruction

Databases

Row ⟺ datapoint

Column ⟺ dim, feature

$d \in \{0,1\}^n$

identifiers
(name, postal code, dob, sex)
secret bit $\{0,1\}$

Name	Postal Code	Date of Birth	Sex	Has Disease?
Alice	K8V7R6	5/2/1984	F	1
Bob	V5K5J9	2/8/2001	M	0
Charlie	V1C7J	10/10/1954	M	1
David	R4K5T1	4/4/1944	M	0
Eve	G7N8Y3	1/1/1980	F	1

Setting

Analyst - Trying to get answers

How many rows satisfying (conditions) have 'Has Disease=1'?

"Name = Alice OR Name = Bob OR Name = Eve"

True answer = 2

Queries $S \subseteq [n]$, $S \in \{0,1\}^n$. $S_i = 1$ if i is in subset, $= 0$ else. $S = [1,1,0,0,1]$

Subset queries

true answer: $A(S) = d \cdot S$, $[1,1,0,0,1] \cdot [1,0,1,0,1] = 2$

Curator - Respond, but "private"

Receive S, respond $r(S)$.

Option: $r(S) = A(S)$ ✗

$S = [1,0,0,\ldots,0]$

Add noise:

- return $r(S)$ s.t. $|r(S) - A(S)| \leq E$

Blatant Non-Privacy ($d \in \{0,1\}^n$)

Definition 1. *An algorithm is* blatantly non-private *if an adversary can construct a database $c \in \{0,1\}^n$ such that it matches the true database d in all but $o(n)$ entries.*

Fairly general schemes are blatantly nonprivate

Inefficient Attack

Theorem 2 ([DN03]). *If the analyst is allowed to ask 2^n subset queries, and the curator adds noise with some bound E, then based on the results, the adversary can reconstruct the database in all but $4E$ positions.*

$E = \frac{n}{401} \Rightarrow$ reconstruct in 99% of entries

$E = o(n) \rightarrow$ b.n.p.

Attack

1. Analyst ask all 2^n subset queries
2. For all $c \in \{0,1\}^n$,

2a. $\exists$? set S s.t. $|\sum c_i - r(S)| > E$.
↳ if so, rule out c

2b. Output any c not ruled out

$S = [1,0,0,\ldots]$
$= [0,1,0,\ldots]$
$= [1,1,0,\ldots]$
$= [0,0,1,\ldots]$

Analysis

- d wouldn't be ruled out
- $I_0 = \{i \mid d_i = 0\}$, $I_1 = \{i \mid d_i = 1\}$

Suppose c output

$|\sum_{i \in I_0} c_i - r(I_0)| \le E$

$|\sum_{i \in I_0} d_i - r(I_0)| \le E$

$\Rightarrow c$ and d differ by $\le 2E$ indices in I_0

$\le 2E$ differences in I_1.

$\le 4E$ diffs. □

Efficient Attack

Theorem 3 ([DN03]). *If the analyst is allowed to ask $O(n)$ subset queries, and the curator adds noise with some bound $E = O(\alpha\sqrt{n})$, then based on the results, a computationally efficient adversary can reconstruct the database in all but $O(\alpha^2)$ positions.*

2^n queries, $O(n)$

$O(n)$ queries, $O(\sqrt{n})$ noise

Attack

1. Analyst asks $O(n)$ random queries. S is chosen u.a.r from $\{0,1\}^n$
2. Find any db c consistent
   ↳ use an LP

Analysis (Intuition)

Domain: $c, d, S \in \{-1,+1\}^n$

Suppose c and d differ in $\Omega(n)$ coords.

$(c-d)\cdot S = \sum (c_i - d_i) S_i$, where S uar from $\{-1,+1\}^n$

if $c_i = d_i$, $(c_i - d_i)S_i = 0$

o.w., $(c_i - d_i)S_i = \begin{cases} +2 & \text{w.p } \frac{1}{2} \\ -2 & \text{w.p } \frac{1}{2}. \end{cases}$

$\left(\sum (c_i - d_i) S_i\right) \sim \text{Bin}(\Omega(n), \frac{1}{2})$ (rescale, shifting)

$(c-d)\cdot S$ has: mean $= 0$, Var $= \Omega(n)$

$\Rightarrow |(c-d)\cdot S| \geq \Omega(\sqrt{n})$ with "large" prob (anti-concentration)

Curator: $E \leq o(\sqrt{n}) \Rightarrow$ Analyst can rule out c (fixed candidate)

- Take $\Omega(n)$ queries, rule out any c w.h.p.
- Union bound over all c which are far from d
- Only remaining are close to d

Attack 1: 2^n q's, $O(n)$ noise

2: $O(n)$ q's, $O(\sqrt{n})$ noise ⇐ [Dwork, McSherry, Talwar '07]

"tight" by DP: $O(m)$ q's, $O(\sqrt{m})$ noise (safe), $m \ll n$

Database Reconstruction in Practice [Cohen-Nissim '20]

Diffix and Aircloak Challenge

2017

$5000

Differences in Setting

- Subset q's — Name = Alice (1) or Bob (2) or Eve (3)
- $|E| \approx \sqrt{\#\text{ of conds.}}$
- Other conditions: No OR statements

(1, 2)

Modifying the Attack

Dinur-Nissim Attack: Picking random sets
- low # of conds

client-ID

mult, exp, d, pred ← T/F q'n

(mult, exp: numbers)

Does the d-th digit of $(\text{mult} * \text{client-id})^{\text{exp}}$ satisfy pred?

mult = 17, client-id = 1, exp = 0.5, d = 3, pred = "Even?"

$(17 \cdot 1)^{0.5} = 4.1231\ldots$ Yes.

```
SELECT count(clientId)
FROM loans
WHERE floor(100 * ((clientId * 2)^0.7) + 0.5) = floor(100 * ((clientId * 2)^0.7))
AND clientId BETWEEN 2000 and 3000
AND loanStatus = 'C'
```

100% ≈ no noise