0x1A Great Papers in Computer Security

Vitaly Shmatikov

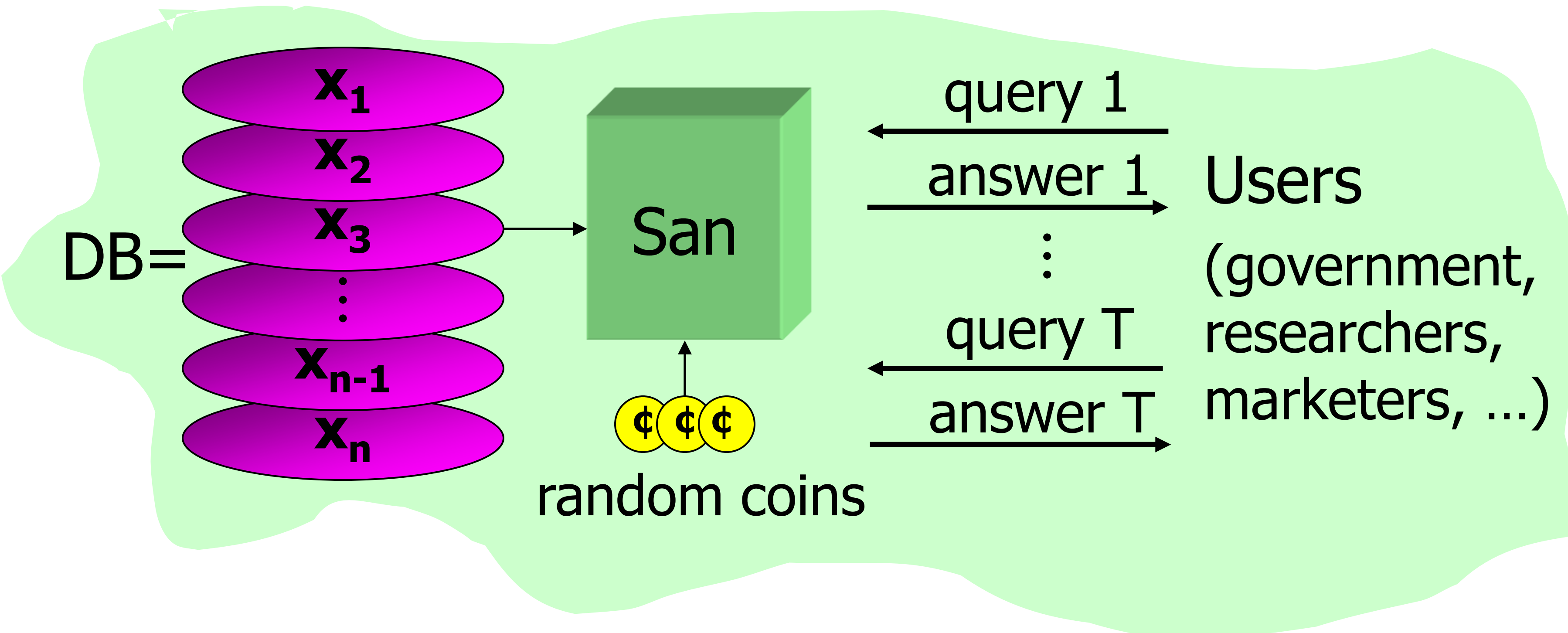
http://www.cs.utexas.edu/~shmat/courses/cs380s/

C. Dwork

Differential Privacy

(ICALP 2006 and many other papers)

Basic Setting



Examples of Sanitization Methods

- Input perturbation
  - Add random noise to database, release
- Summary statistics
  - Means, variances
  - Marginal totals
  - Regression coefficients
- Output perturbation
  - Summary statistics with noise
- Interactive versions of the above methods
  - Auditor decides which queries are OK, type of noise

Strawman Definition

- Assume $x_1, \ldots, x_n$ are drawn i.i.d. from unknown distribution
- Candidate definition: sanitization is safe if it only reveals the distribution
- Implied approach:
  - Learn the distribution
  - Release description of distribution or re-sample points
- This definition is tautological
  - Estimate of distribution depends on data… why is it safe?

Clustering-Based Definitions

- Given sanitization S, look at all databases consistent with S
- Safe if no predicate is true for all consistent databases
- k-anonymity
  - Partition D into bins
  - Safe if each bin is either empty, or contains at least k elements
- Cell bound methods
  - Release marginal sums

	brown	blue	Σ
blond	2	10	12
brown	12	6	18
Σ	14	16	

	brown	blue	Σ
blond	[0,12]	[0,12]	12
brown	[0,14]	[0,16]	18
Σ	14	16	

Issues with Clustering

- Purely syntactic definition of privacy
- What adversary does this apply to?
  - Does not consider adversaries with side information
  - Does not consider probability
  - Does not consider adversarial algorithm for making decisions (inference)

Classical Intution for Privacy

- "If the release of statistics S makes it possible to determine the value [of private information] more accurately than is possible without access to S, a disclosure has taken place." [Dalenius 1977]
  - Privacy means that anything that can be learned about a respondent from the statistical database can be learned without access to the database
- Similar to semantic security of encryption
  - Anything about the plaintext that can be learned from a ciphertext can be learned without the ciphertext

Problems with Classic Intuition

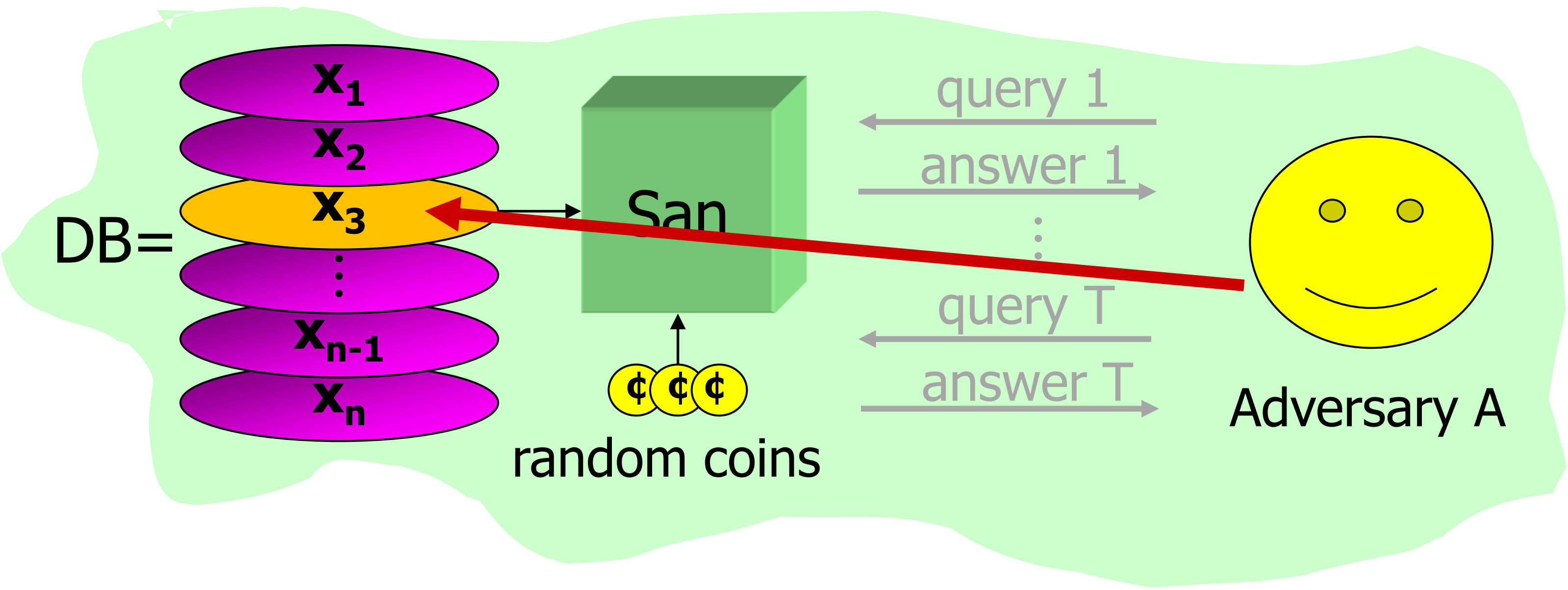
- Popular interpretation: prior and posterior views about an individual shouldn't change "too much"
  - What if my (incorrect) prior is that every UTCS graduate student has three arms?
- How much is "too much?"
  - Can't achieve cryptographically small levels of disclosure <u>and</u> keep the data useful
  - Adversarial user is <u>supposed</u> to learn unpredictable things about the database

Absolute Guarantee Unachievable

[Dwork]

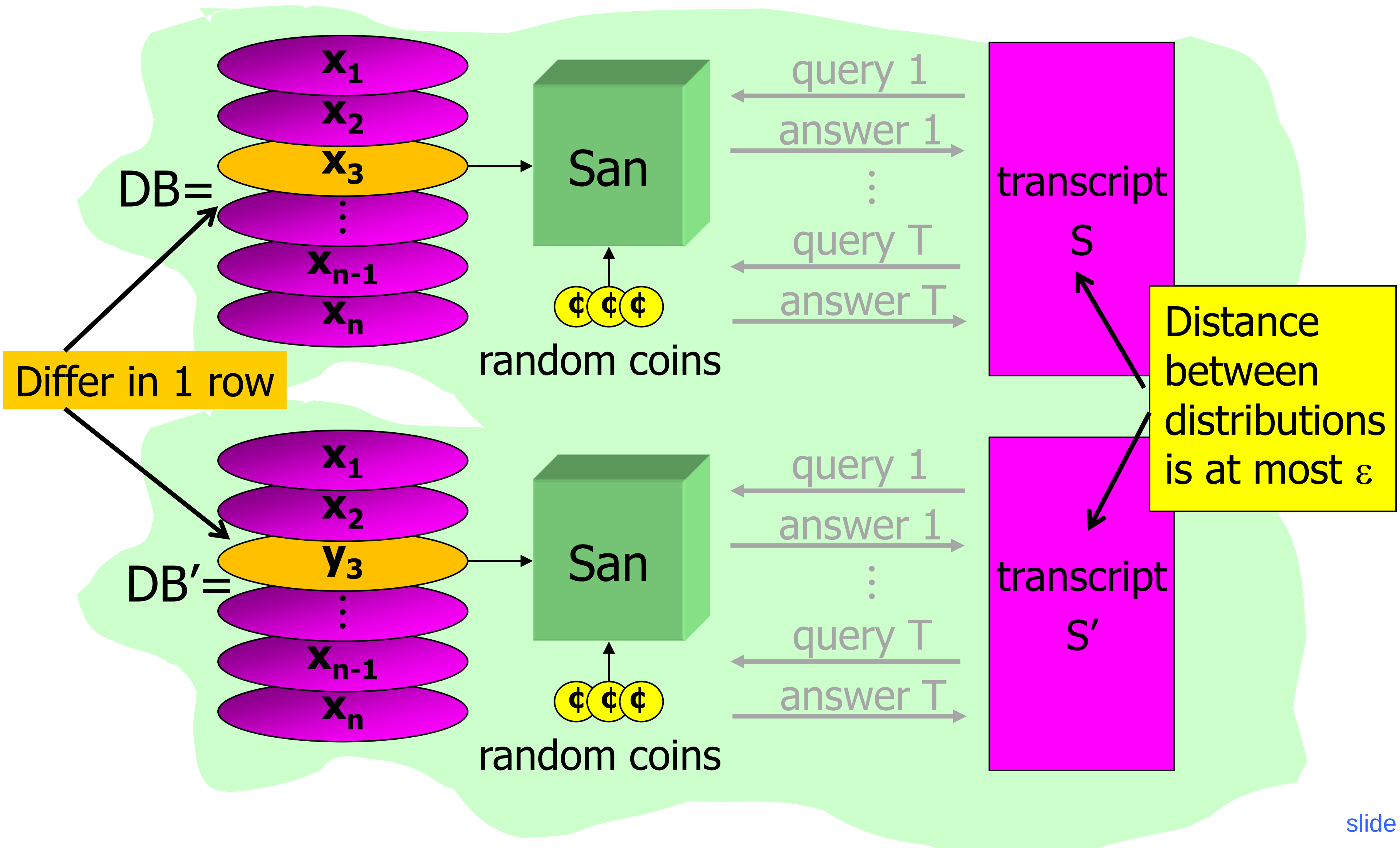
- Privacy: for some definition of "privacy breach,"
  $\forall$ distribution on databases, $\forall$ adversaries A, $\exists$ A′
  such that Pr(A(San)=breach) – Pr(A′()=breach) $\leq \varepsilon$
  - For reasonable "breach", if San(DB) contains information about DB, then some adversary breaks this definition
- Example
  - Vitaly knows that Chad is 2 inches taller than the average Russian
  - DB allows computing average height of a Russian
  - This DB breaks Chad's privacy according to this definition… even if his record is not in the database!

Differential Privacy



- Absolute guarantees are problematic
  - Your privacy can be "breached" (per absolute definition of privacy) even if your data is not in the database
- Relative guarantee: "Whatever is learned would be learned regardless of whether or not you participate"
  - Dual: Whatever is already known, situation won't get worse

Indistinguishability



Which Distance to Use?

- Problem: ε must be large
  - Any two databases induce transcripts at distance $\leq n\varepsilon$
  - To get utility, need $\varepsilon > 1/n$
- Statistical difference $1/n$ is not meaningful!
  - Example: release a random point from the database
    - $San(x_1,\ldots,x_n) = (j, x_j)$ for random j
  - For every i, changing x_i induces statistical difference $1/n$
  - But some x_i is revealed with probability 1
    - Definition is satisfied, but privacy is broken!

Formalizing Indistinguishability

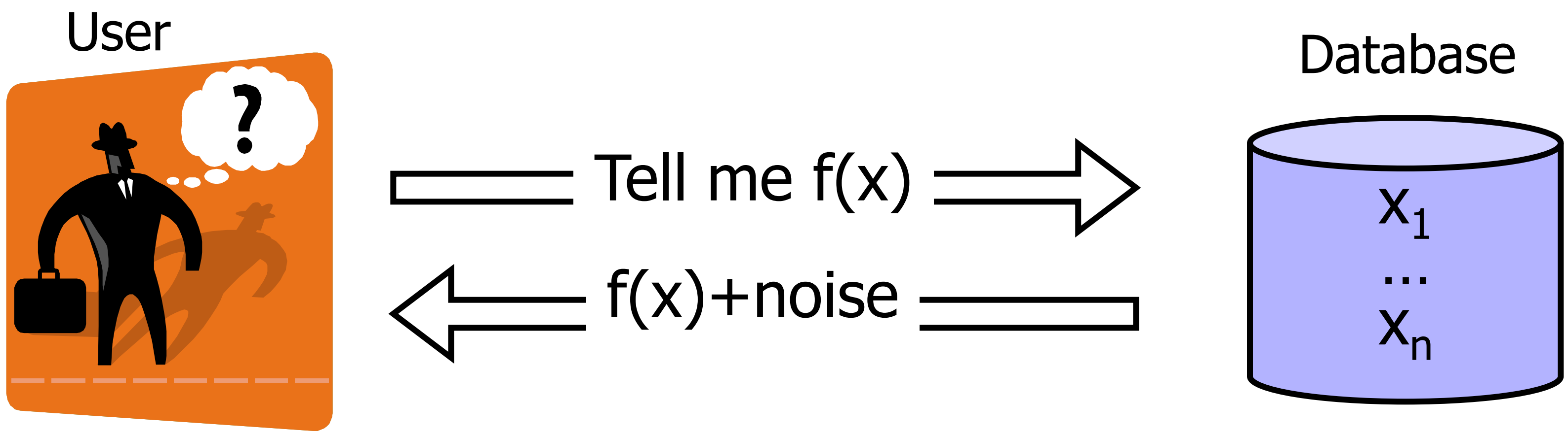
Definition: San is ε-indistinguishable if

$\forall$ A, $\forall$ DB, DB′ which differ in 1 row, $\forall$ sets of transcripts S

$$p(\text{ San(DB)} \in S\ ) \in (1 \pm \varepsilon)\ p(\text{ San(DB')} \in S\ )$$

Equivalently, $\forall$ S: $\dfrac{p(\text{ San(DB)} = S\ )}{p(\text{ San(DB')} = S\ )} \in 1 \pm \varepsilon$

Laplacian Mechanism



- Intuition: f(x) can be released accurately when f is insensitive to individual entries $x_1, \ldots x_n$
- Global sensitivity $GS_f = \max_{\text{neighbors } x,x'} ||f(x) - f(x')||_1$
  - Example: $GS_{average} = 1/n$ for sets of bits

Lipschitz constant of f

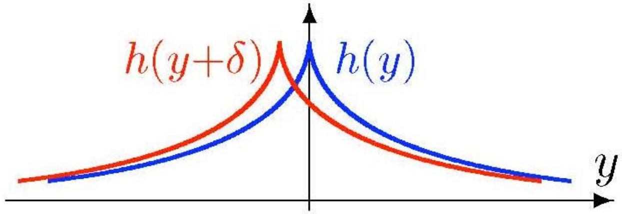
- Theorem: $f(x) + Lap(GS_f/\varepsilon)$ is ε-indistinguishable
  - Noise generated from Laplace distribution

Sensitivity with Laplace Noise

Theorem

If $A(x) = f(x) + \mathsf{Lap}\left(\frac{\mathsf{GS}_f}{\varepsilon}\right)$ *then* A *is* ε*-indistinguishable.*

Laplace distribution $\mathsf{Lap}(\lambda)$ has density $h(y) \propto e^{-\frac{\|y\|_1}{\lambda}}$



Sliding property of $\mathsf{Lap}\left(\frac{\mathsf{GS}_f}{\varepsilon}\right)$: $\frac{h(y)}{h(y+\delta)} \leq e^{\varepsilon \cdot \frac{\|\delta\|}{\mathsf{GS}_f}}$ for all y, δ

Proof idea:

$$A(x)\text{: blue curve}$$

$$A(x')\text{: red curve}$$

$$\delta = f(x) - f(x') \leq \mathsf{GS}_f$$

Differential Privacy: Summary

- San gives ε-differential privacy if for all values of DB and Me and all transcripts t:

$$\frac{\Pr[\,San(\mathrm{DB} - \mathrm{Me}) = t\,]}{\Pr[\,San(\mathrm{DB} + \mathrm{Me}) = t\,]} \le e^{\varepsilon} \approx 1 \pm \varepsilon$$