# Decision Making
## *in Robots and Autonomous Agents*

## Security, Privacy and Relevance to Robotics

Subramanian Ramamoorthy
School of Informatics

20 March, 2018

# Correctness vs. Security

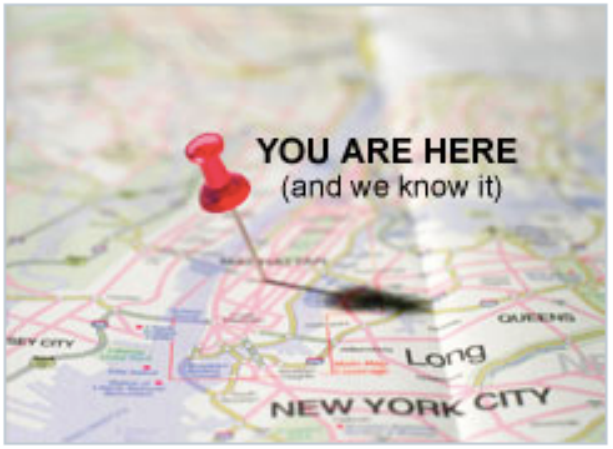
- Program or system correctness:
  program satisfies specification
  - For reasonable input, get reasonable output
- Program or system security:
  program properties preserved in the face of attack
  - For unreasonable input, output not completely disastrous
- Main difference: adversary
  - Active interference from a malicious agent
  - It is very difficult to come up with a model that captures all possible adversarial actions
    - Hence for the need for discussion around models

# An Ongoing Situation

## Senate introduces mobile location privacy bill

Consumer Reports News: June 16, 2011 02:23 PM

Yesterday, two U.S. senators proposed a bill that would require companies to obtain consent from mobile device users before sharing location-based information with third parties.




The bill, called the Location Privacy Protection Act of 2011, is sponsored by Senators Al Franken (D-Minn.) and Richard Blumenthal (D-Conn.). It would require companies to tell consumers when information about their location is being collected, and allow consumers to decide if they want that information to be shared, according to a press release on Franken's official website.

The legislation seeks to close a loophole in the Electronic Communications Privacy Act that allows smartphone companies, app companies, and phone companies that offer wireless Internet to disclose consumers' location information to third parties.

# What kinds of data does Microsoft collect?

Microsoft collects data to help you do more. To do this, we use the data we collect to operate and improve our software, services, and devices, provide you with personalized experiences and to help keep you safe. These are some of the most common categories of data we collect.

## Information from device sensors

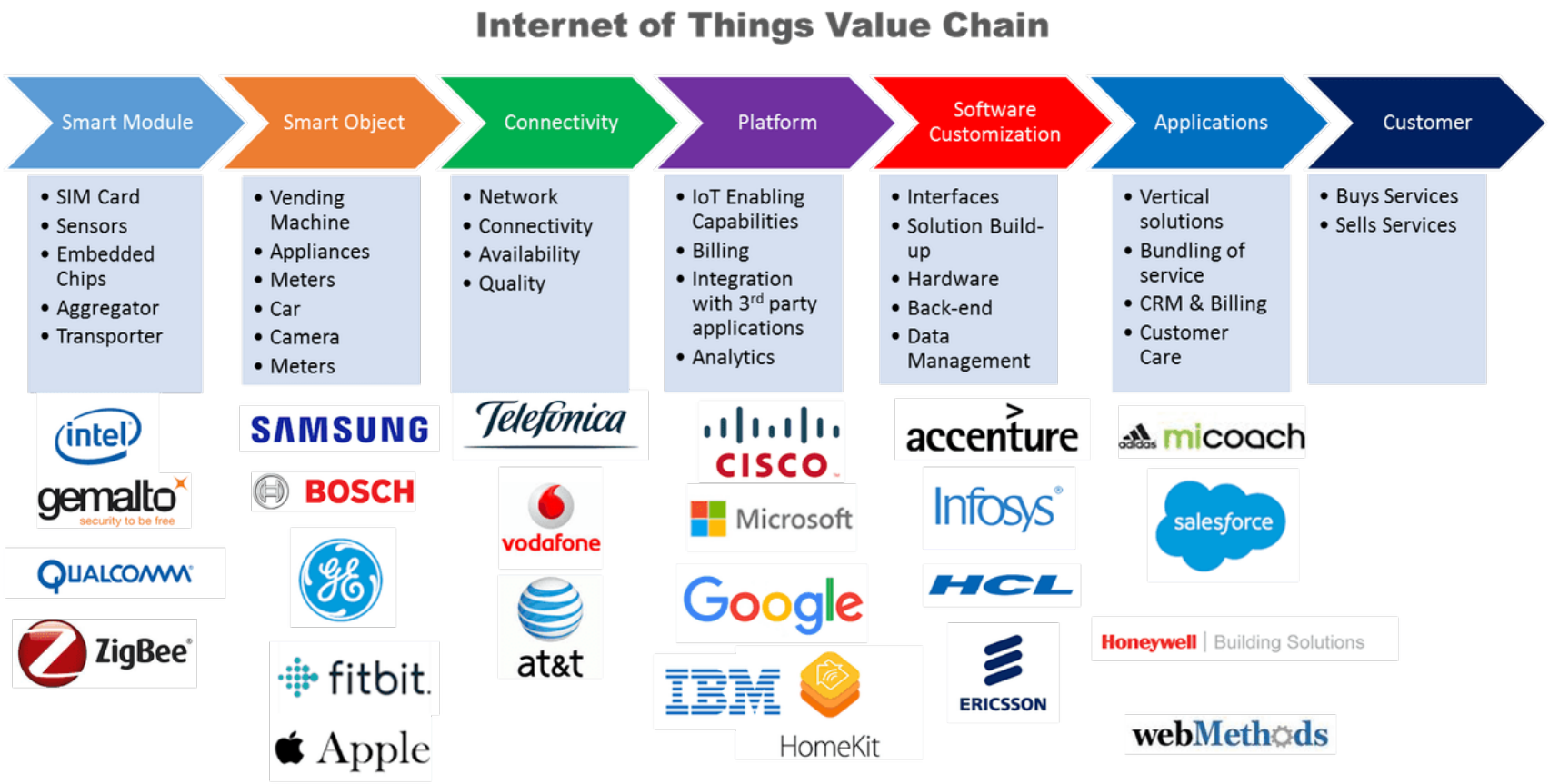
Windows 10 phones, tablets, and PCs come with
d. That can be your phone's microphone or
', an internal GPS sensor, and more.

## Places you go

Location information helps us give you directions to the places you want to go and show you info relevant to where you are. For this, we use the locations you provide or that we've detected using technologies like GPS or IP addresses.

Detecting location also helps us protect you. For example, if you almost always sign in from Tokyo, and suddenly you're signing in from London, we can check to make sure it's really you.

# Why do we Give it Away?!
# Location Based Services




*Note, the above is not an exhaustive list of companies and any company may have play in more than one component of value chain*
*Copyright: Telecomcircle.com*

# Robots are Becoming Connected too! What are the implications?

[http://blog.ncpad.org/wp-content/uploads/2011/05/carebot.jpg]

Already in your home:

https://www.youtube.com/watch?v=H0h20jRA5M0

# Discuss…

1. What is “privacy”? How will you model it?

2. How will you ensure it through computational means? (We will not spend much time discussing regulation, social engineering, etc. – topic of a whole other course!)

# Location Privacy

- "… the ability to prevent other parties from learning one's current or past location" [Beresford + Stajano]
- Principle is that the person whose location is being measured should control who can know it
- Many ways in which location information can be revealed:
  1. *When*: A subject may be more concerned about current or future location being revealed than past locations
  2. *How*: User may be comfortable if friends can manually request location but not want alerts sent automatically
  3. *Extent*: User may prefer to have location reported as ambiguous region rather than precise point

# Computational Threats

- Consequences of location leak can range from uncomfortably creepy (being watched), to unwanted revelation (e.g., AIDS clinic, political locations), to actual physical harm.
- Computational attacks include:
  - Analysis of movement patterns, e.g., GPS traces
  - “Inference” attacks
  - Context inference

# Analysis of Movement Patterns

Examples (often benign) from the literature:

- Look for places where GPS signal is lost three or more times within a given radius
  - Often happens because a building blocks the signal, so prompt user to enter location
  - Cluster such places and treat as labels
- Look for combinations of dwell time, breaks in time or distance, and periods of low GPS accuracy – treat as potentially significant locations
- Fingerprinting through the use of repeatable sets of in-range GSM and Wi-Fi base stations

# “Inference Attacks”

Use inference algorithms (e.g., Bayesian inference) to go from evidence (observed traces of movement) to latent variables (e.g., locations of interest acting as goals)

Examples of how attacks get carried out:

- Using location measurements from an indoor sensor, examine where people in an office building spent their time, including e.g., who spent more time than anyone else at a given desk?
- Using week-long GPS traces from drivers in a city, algorithmically determine the home locations of drivers
  - Can be done to up to 85% accuracy

# “Context” Inference

We can infer many more things beyond home location

Examples:

- Use GPS traces to infer, in real time, a moving person’s mode of transportation (bus/foot/car)
- Predict their route based on historical movement data
- Very common to predict potential routes and target destinations (highly developed due to Uber, etc.)
- Can look at multi-agent data to identify events such as meetings and stopover locations

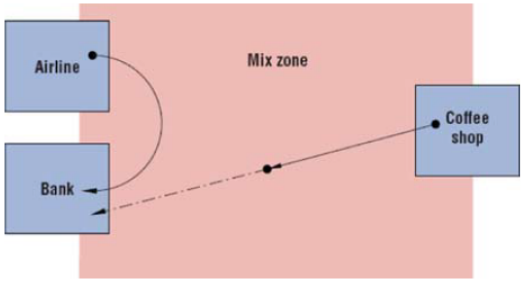
# Computational Countermeasures: What could One Do?

Four main routes:

- Regulatory strategies: government rules on what is OK
- Privacy policies: trust-based agreements between individuals and those receiving the data
- Anonymity: use a pseudonym and create ambiguity by grouping with other people
- Obfuscation: reduce the quality of location data

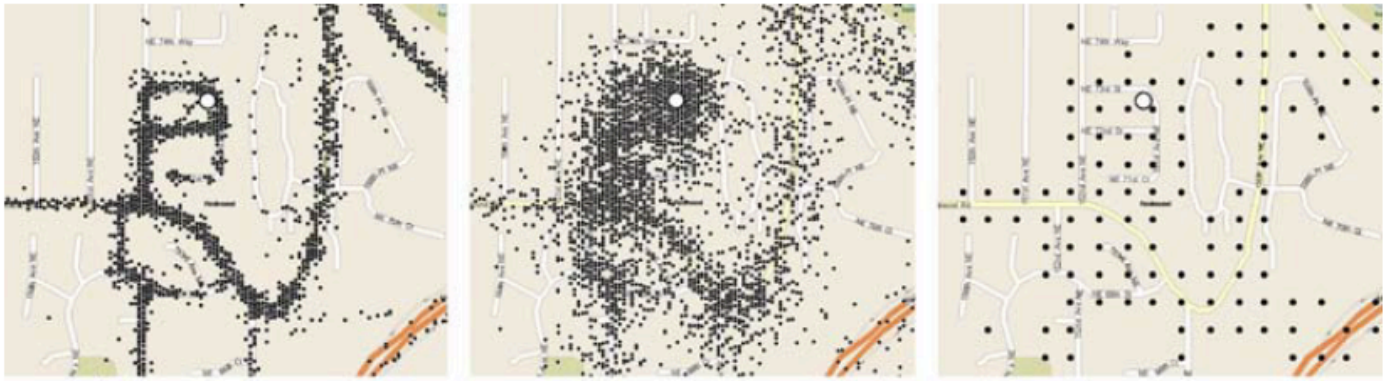
# Anonymity

- Replace associated name with an untraceable ID, i.e., pseudonym (could be long-term or frequently changing)
  - What is the benefit of frequent change?
  - What is the practicality of using this as protection?
- Algorithmic ideas:
  1. K-anonymity: Instead of pseudonymously reporting exact location, person reports a region containing k-1 other people
     - Person can not be distinguished from k-1 other people
     - May need historical k-anonymity (when attacker can use traces)
  2. Mix zones: Give new pseudonym in regions
     - Defined as outside of well known labelled areas
     - Hard for attacker to guess identity in this zone



# Obfuscation

- Degrading the quality of location measurements may reduce threats to location privacy
- Inaccuracy: give measurement different from actual
- Add additive noise and/or quantization
  - Is this enough? Discuss when and how much…



**(a)** Original GPS data **(b)** Additive Gaussian noise **(c)** Discretized to points on grid

[J. Krumm, Inference attacks on location tracks. In Pervasive 2007, pp 127–143]

# Modelling Location Privacy

What to include in the model:

- Set of mobile users
- Set of all possible traces (motion trajectories)
- Location Privacy Preserving Mechanism (LPPM) – the protocol
- Set of all observable traces

- Specification of the “Adversary”
- Specification of an evaluation metric, i.e., when is an adversary considered to have succeeded

# Location Privacy Preserving Mechanisms

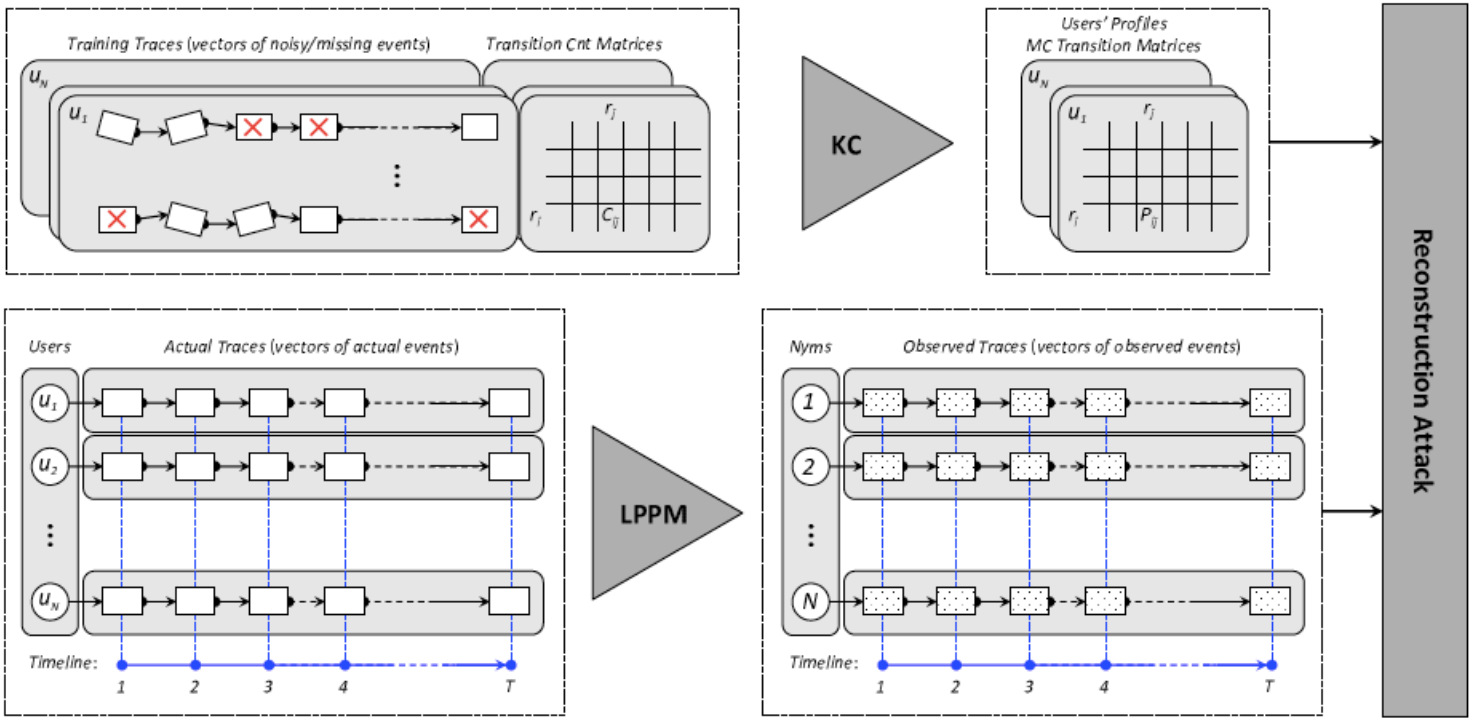
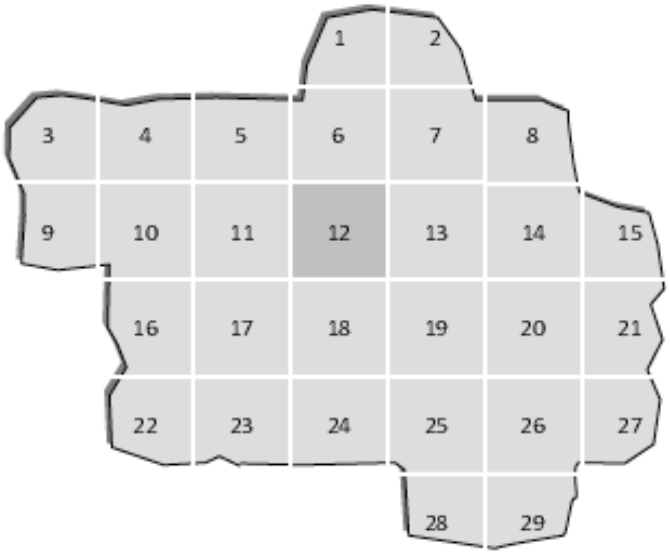



Figure 1. Elements of the proposed location-privacy framework. The users produce actual traces, which are then anonymized and obfuscated by the LPPM to produce anonymous observed traces. The attacker uses a set of training traces to create, via the knowledge construction (KC) mechanism, a mobility profile for each user in the form of a Markov Chain transition probability matrix. Having the user mobility profiles and the observed traces, the adversary tries to reconstruct (infer) the actual traces. The only element of the framework not shown here is the metric that evaluates the success of the adversary's reconstruction attack by comparing the results of the attack with the users' actual traces.

[R. Shokri et al., Quantifying location privacy, IEEE Symp. Sec. Privacy 2011]

# Obfuscating Location

- Consider user u whose actual location is region r12
- Different obfuscation methods will replace r12 with a different location pseudonym r′
  - Perturbation: r′ = {14}
  - Add dummy regions: r′ = {12, 15, 26}
  - Reduce precision, r′ = {9, 10, 11, 12, 13, 14, 15}
  - Location hiding, r′ = ∅



# Main Inference Problem in Tracking Attacks

- Attacker has partial traces of location, possibly after some kind of obfuscation
- They need to solve an inference problem (for parameters of a Markov Chain, P), which involves first completing the traces
- Direct computation is intractable (sum of terms whose number grows exponentially with length of trace)
  - Use sampling based approximations

$$\Pr(P|TT,TC) = \sum_{ET} \Pr(P,ET|TT,TC).$$

Training traces  Transition counts  Estimated Completion of TT

# Many Variations on this Theme of Inference Attacks

- Tracking attacks: Maximimize a quantity of the form,

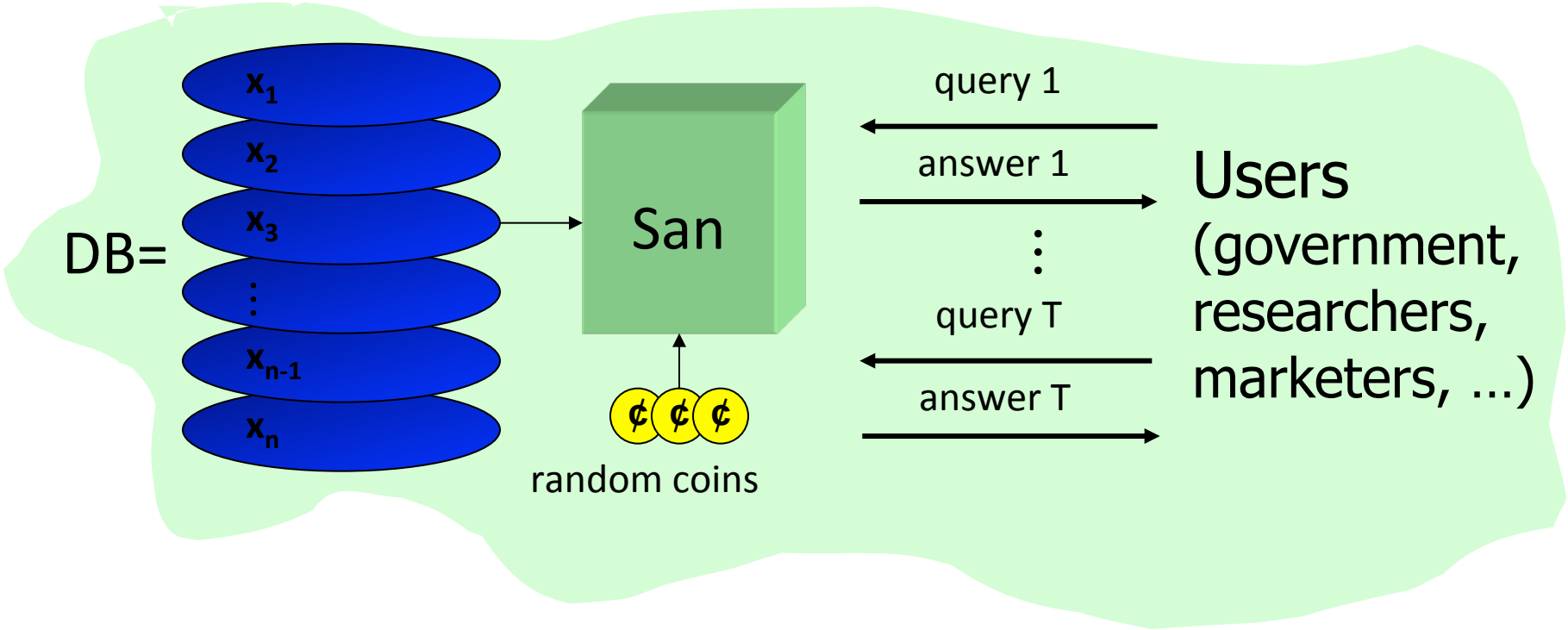
$$\arg\max_{\sigma,A} \Pr(\sigma, A|O).$$

to determine pseudonym permutation assignments ($\sigma$) and actual traces ($A$) given observed traces ($O$)

- Localization attack: More specifically, determine the probability of a user being at a location at a specific time (given some knowledge of user profile $P^u$),

$$\Pr\{a_u(t) = r|o_u, P^u\}$$

# A More General View of Privacy in Data Anslysis

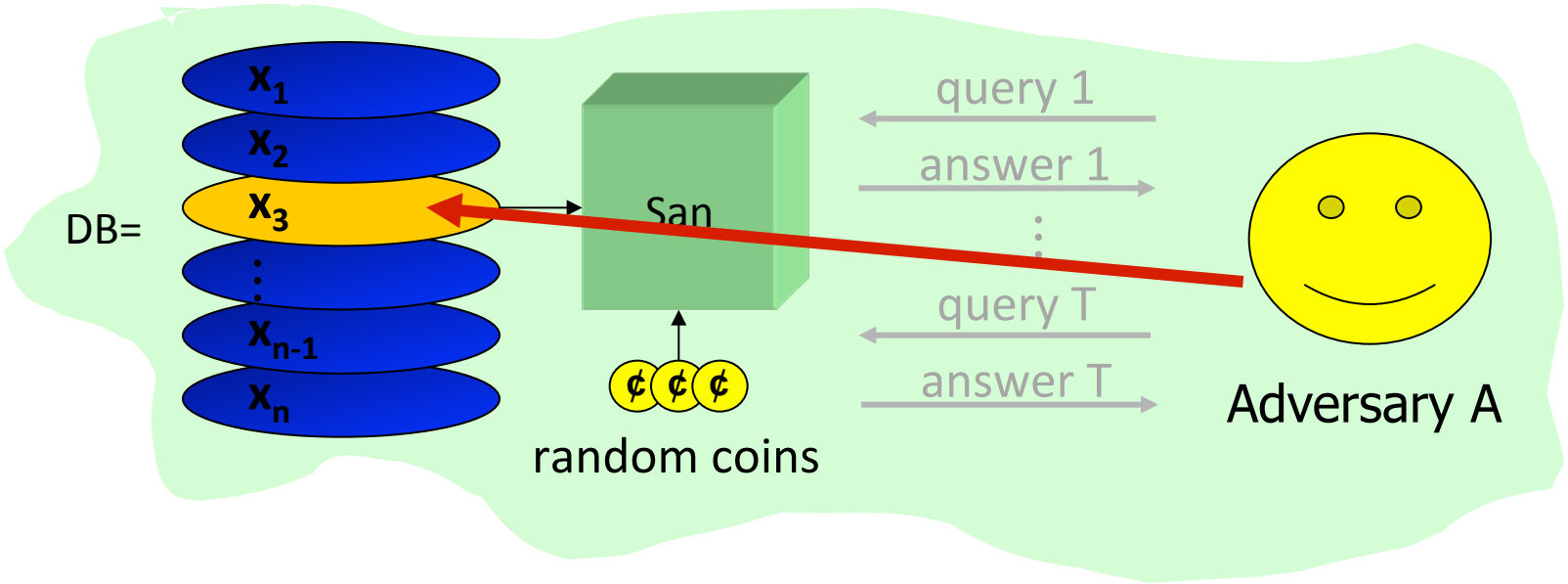



San: Sanitization mechanism; DB: database

# 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

# Differential Privacy

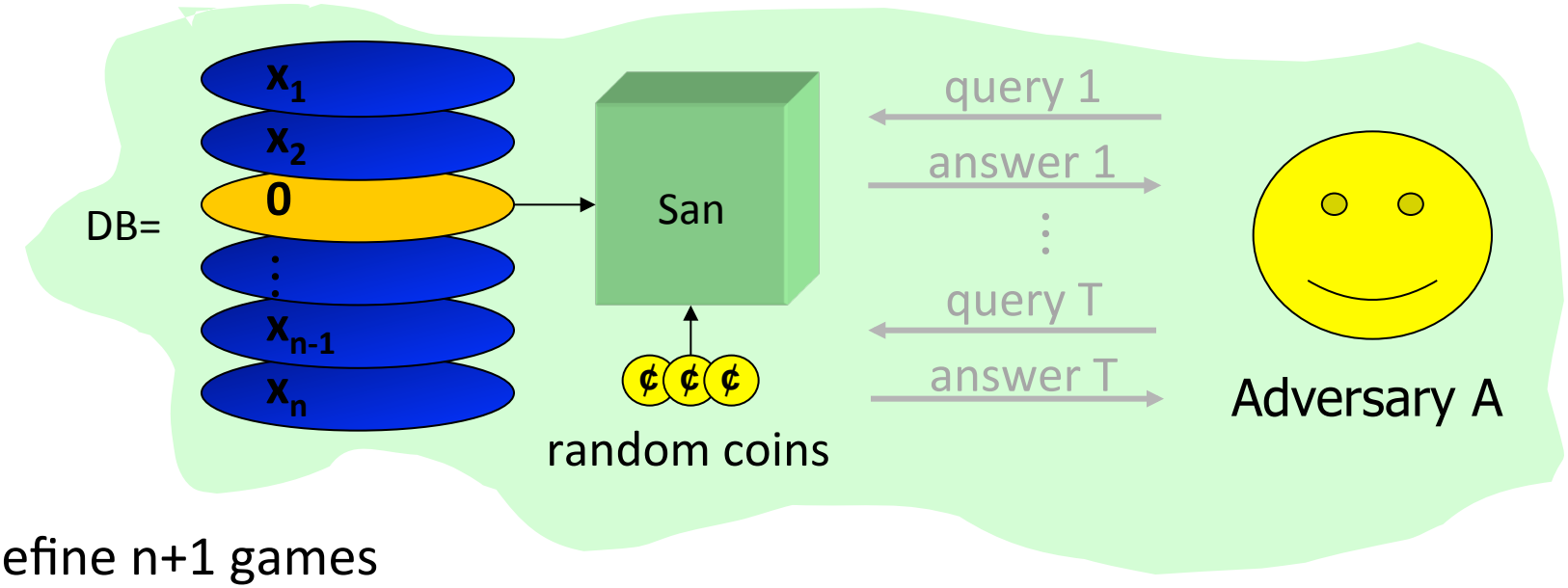



Adversary learns Alex's height even if he is not in the database

Intuition: "Whatever is learned would be learned regardless of whether or not Alex participates"

Dual: Whatever is already known, situation will not get worse

# Differential Privacy




Define n+1 games

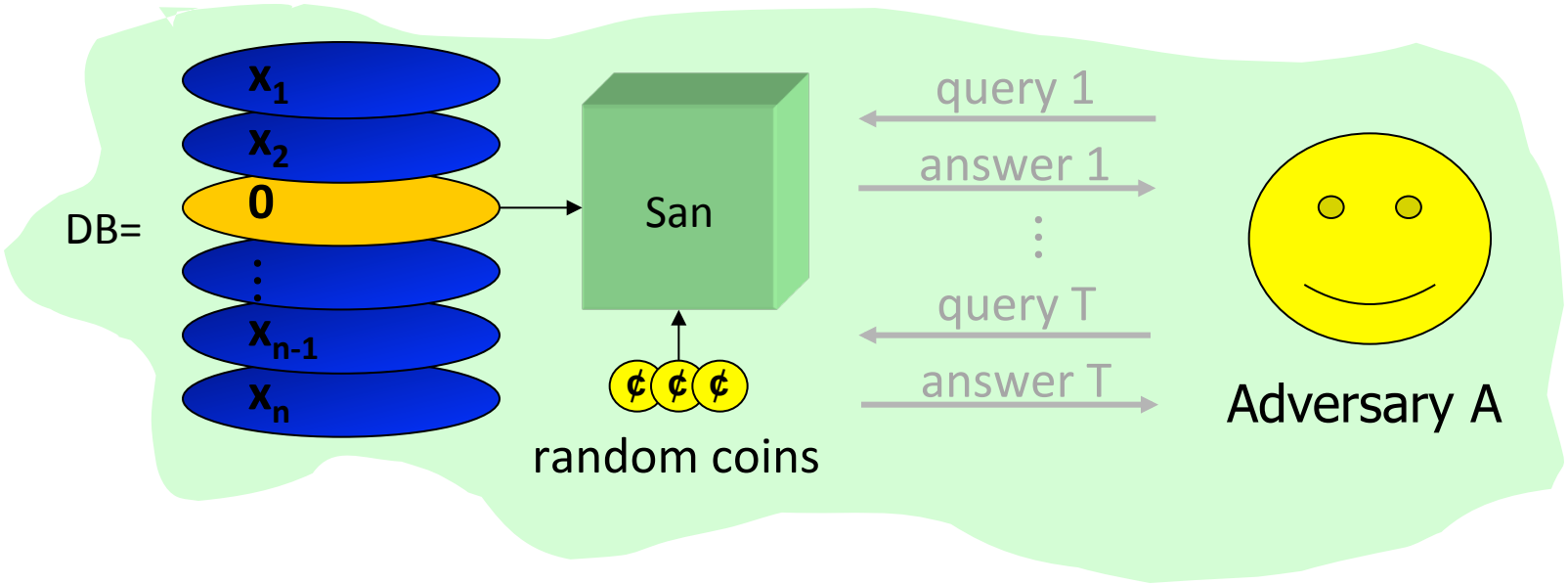
Game 0: Adv. interacts with San(DB)

Game i: Adv. interacts with San($DB_{-i}$); $DB_{-i} = (x_1,\ldots,x_{i-1},0,x_{i+1},\ldots,x_n)$

Given S and prior p() on DB, define n+1 posterior distrib's

$$p_i(DB|S) = p(DB|S \text{ in Game } i) = \frac{p(San(DB_{-i}) = S) \times p(DB)}{p(S \text{ in Game } i)}$$

# Differential Privacy




San is safe if

∀ prior distributions p(¢) on DB,

∀ transcripts S, ∀ i =1,…,n

$$\text{StatDiff}( p_0(¢|S) , p_i(¢|S) ) \leq \varepsilon$$

# (Statistical) Indistinguishability