Deanonymization: Aggregated Data

Dr. Balázs Pejó

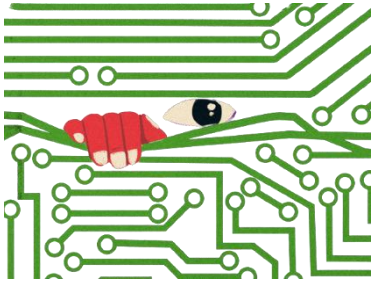
www.crysys.hu

Agenda

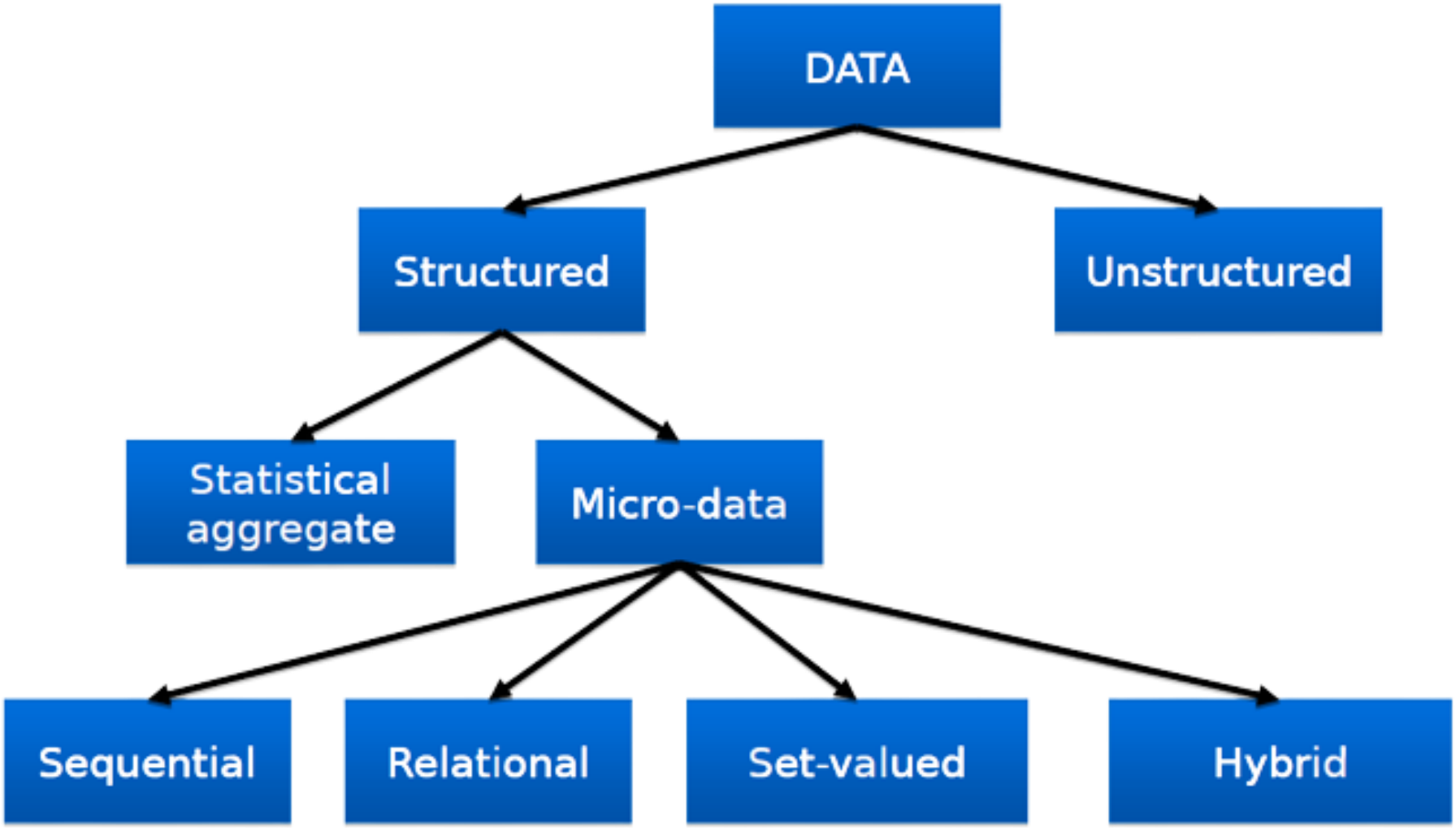
- Dark Patterns
- Tracking
- GDPR
- Deidentification
- Machine Learning
- Anonymization
- Cryptography
- Entropy
- Database Reconstruction
- Query Auditing
  - On-line & Off-line
  - Theorems & Attack
- SAT
- Location Data
- Federated Learning

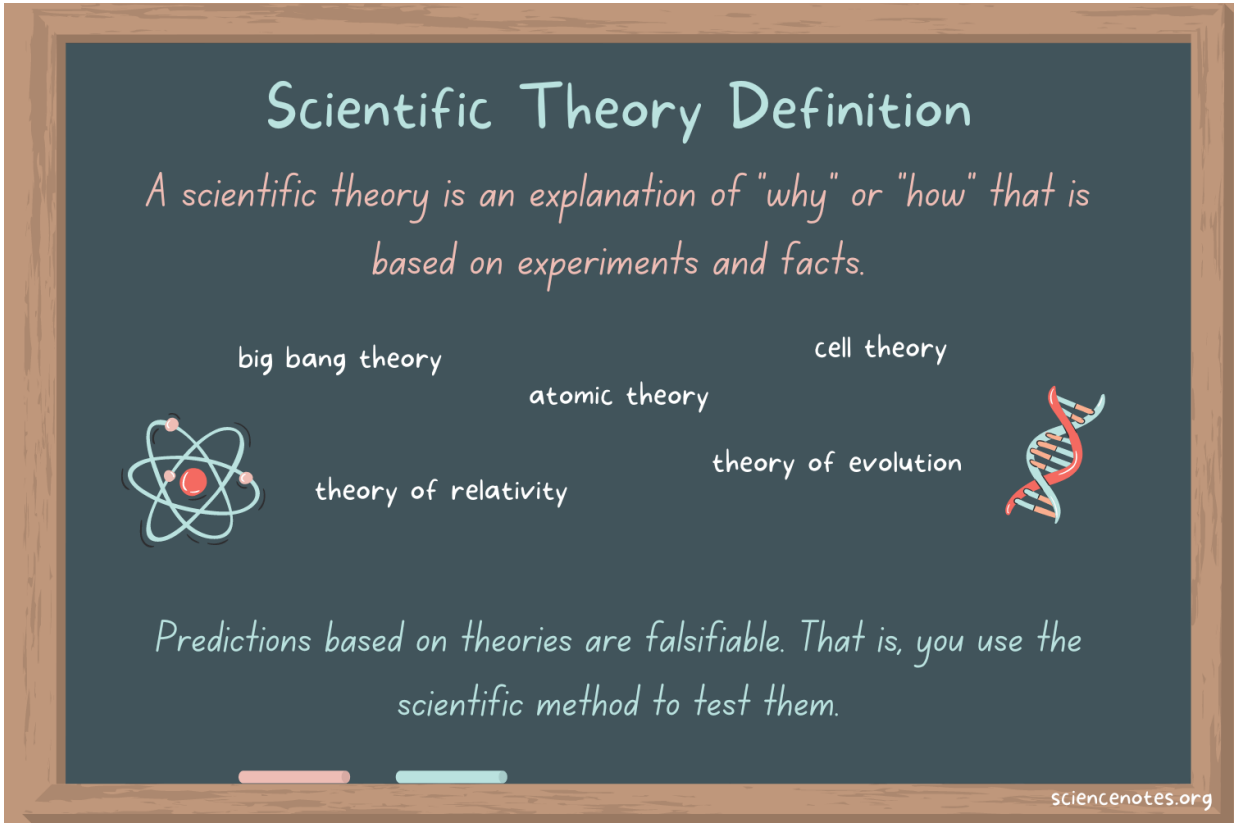
Recap

- All data type can be attacked.
  - Targeted / Untargeted / Mass Attack
- GDPR: if de-anonymization is reasonably possible, the data is personal (even it is encrypted).
  - By linking it to a physical person.
- Systematic (and automatic) de-anonymization is done by machine learning.
  - User is linked to anonymous data with a classifier.
  - One vs One / One vs Rest / All vs All



Data Types


DATA
Structured
Unstructured
Statistical aggregate
Micro-data
Sequential
Relational
Set-valued
Hybrid


Scientific Theory Definition
A scientific theory is an explanation of "why" or "how" that is based on experiments and facts.
big bang theory
cell theory
atomic theory
theory of evolution
theory of relativity
Predictions based on theories are falsifiable. That is, you use the scientific method to test them.
sciencenotes.org

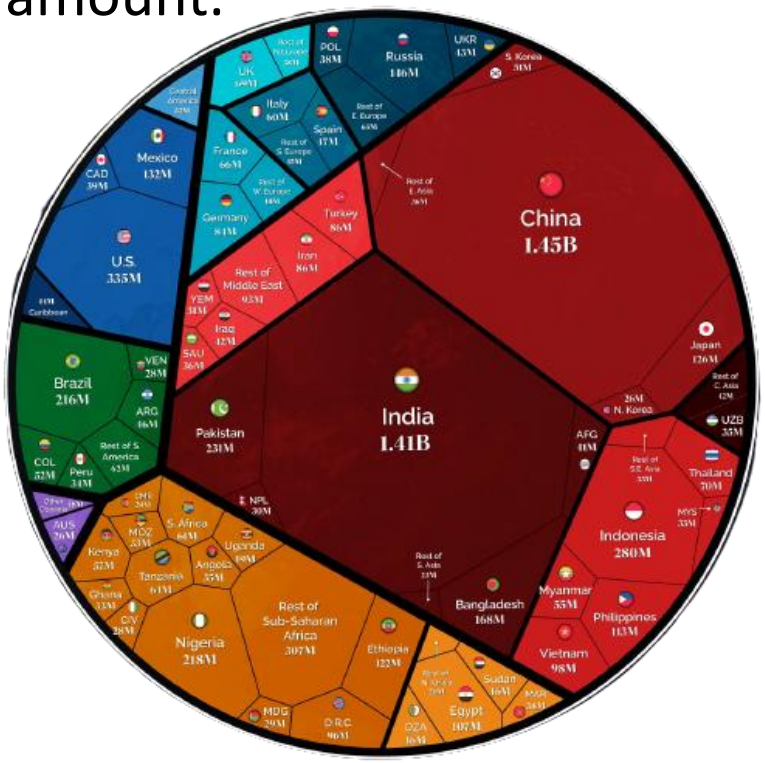
Entropy

Entropy

- There are around 8 billion humans on the planet.
- The identity of a random, unknown person contains just under 33 bits of entropy ($2^{33} \approx 8$ billion).
- When we learn a new fact about a person, that fact reduces the entropy of their identity by a certain amount.

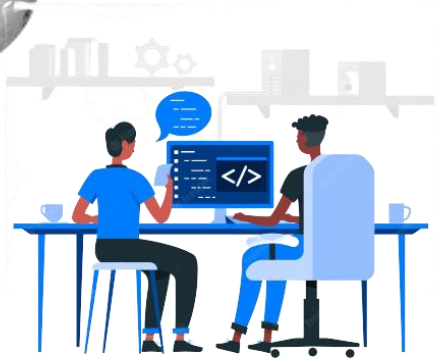
- $\Delta S = - \log_2 \Pr(X=x)$

- ΔS is the reduction in entropy, measured in bits.
- $\Pr(X=x)$ is simply the probability that the fact would be true of a random person.



Example

- Western / Chinese Zodiac Sign: Ram & Goat
  - $\Delta S = -\log_2 \Pr(\text{Starsigns} = \text{Capricorn}) = -\log_2[1/(12\text{x}12)] = 7.2$
- City: Szolnok
  - $\Delta S = -\log_2 \Pr(\text{ZIP} = 5000) = -\log_2[70 \text{ thousand} / 8 \text{ billion}] = 16.8$
- Job: Software Engineer
  - $\Delta S = -\log_2 \Pr(\text{Job} = \text{SE}) = -\log_2 (30 \text{ million} / 8 \text{ billion}) = 8.0$
- Starsigns & City & Job
  (assuming they are independent)
  - 7.2 + 16.8 + 8.0 = 33 bits



- Knowing all three pieces of information,
  we can probably say exactly who the person is!
  - By combining several facts, you might learn nothing, e.g., a starsign does not reduce the entropy if the birthday is already known.

Database Reconstruction

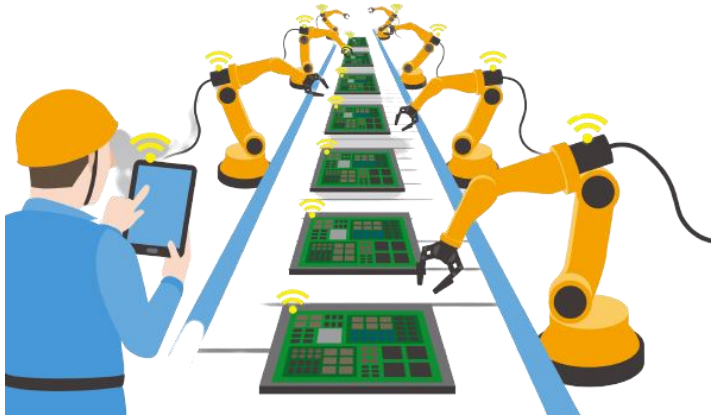
Attack & Naïve Solution

- Ad-hoc anonymization techniques were automated from 1960.
  - Increased ability to analyze & the risk to isolate individuals.

- Attack: if the answer is 1 below, we singled out the victim.
  - `SELECT count(*) FROM table WHERE bday = '1990-01-01' AND gender = 'F' AND zip = 12345`
  - This can lead to further attribute disclosure.

```
SELECT salary, count(*) FROM table WHERE bday =
'1990-01-01' AND gender = 'F' AND zip = 12345
GROUP BY salary
```

- *K-threshold mechanism*: a certain number K of individuals must be present in aggregated data for the aggregate to be released.



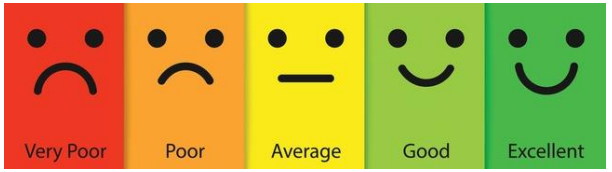
Following Attack & Naïve Solution

- *Intersection Attack*
  - Suppose that there are N people in the CS department and N-1 of them are men). Let K > 2.
  - `SELECT salary, count(*) FROM table WHERE dep = 'CS'`
  - `SELECT salary, count(*) FROM table WHERE dep = 'CS' AND gender = 'M'`

- Solution: distorting answers, e.g., by adding noise.

More Attacks & Naïve Solution

- *Averaging Attack*
  - If the analyst can make an unlimited number of queries to the database, the added noise can be removed through averaging.



- Defense: use sticky noise.
  - The same query would then produce the same noise.

```
# Trust me, I used a die
rand = 4
```

- *Chaff Attack*
  - The analyst can still average out the noise by generating multiple different queries that all produce the same result, e.g., by adding conditions like age < 1000.

- And so on …

U.S. Census History

- 1790: First census (without any safeguards).
- 1850: 'Business data' is not published directly.
- 1920: Manual suppression & compression is used to prevent indirect disclosure of 'business data'.
  - Extended to 'people data' in 1940.
- 1930: Stops publishing small area data.
  - Cannot prevent indirect disclosure.
- 1970: Whole data table suppression about housing.
  - Extended to more tables in 1980.
- 1990: Data swapping, blank & impute protection is added.
  - To enable less whole table suppression.
- 2000: Rounding, top-coding, and more is added as further protection mechanisms.
- 2020: Differential Privacy is replaced all previous ad-hoc protection mechanism.

Example

- Database is private.
- Statistics are public.
  - NA means not published because the value is < 3.
- Is this safe?
- Can we reconstruct the private dataset from the published statistics?

Age	Sex	Race	Marital status
8	F	B	S
18	M	W	S
2[illegible]	F	W	S
0	M	W	M
3	F	B	M
66	F	B	M
84	M	B	M

PRIVATE

Age	Count	Median	Mean
Total population	7	30	38
Female	4	30	3[illegible]
Male	3	3[illegible]	4[illegible]
Black	4	5[illegible]	48.5
White	3	24	24
Single adults	[illegible]	NA	NA
Married adults	4	51	54
Black female	3	36	36.7

PUBLIC

Feasibility

- Private data
  - Each record has 3 binary attributes (3 bits) and one multi-value attribute (age: 0 – 128 -> 7 bits).
  - For 7 records: $(3 + 7) \cdot 7 = 70$ bits in total.
- Public statistics
  - Median: 7 bits
  - Mean: 7 bits
  - Count: 3 bits
  - Each query is released for 7 groups, which means $(7+7+3) \cdot 7 = 119$ bits.

- We release 119 bits about a dataset which contains only 70 bits of information, so reconstruction seems possible!

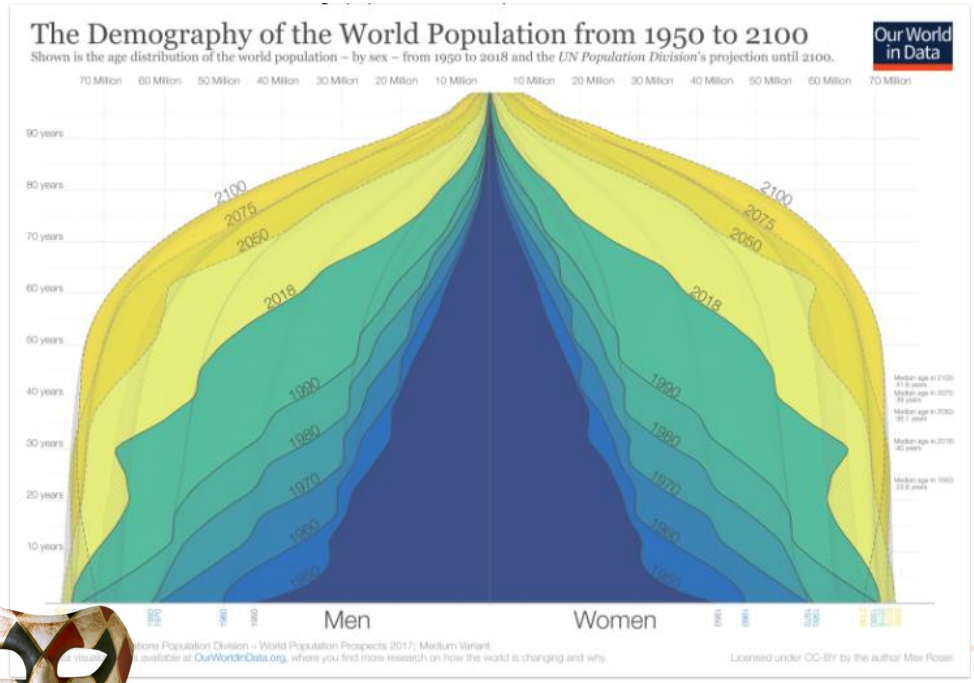
Reconstruction Idea

- We know that there are 3 whites in the private dataset.
  - If 0 < Age < 128, then we have $128^3 / 3 = 699051$ possible combinations.
- However, out of these, there are only 24 with mean 24 and median 24.
  - Applying the constraints imposed by the released statistics, the possible combinations can be substantially reduced!
- The more constraints (i.e., released statistics) we have the more likely we can find the private attributes!

Age	Count	Median	Mean
Total population	7	30	38
Female	4	30	33.5
Male	3	30	40
Black	4	51	48.5
White	3	24	24
Single adults	NA	NA	NA
Married adults	4	51	54
Black female	3	36	36.7

Background Knowledge

- Instead of selecting an option uniformly from the possible options, the attacker can use other available (population level) statistics to further increase the guessing chance.

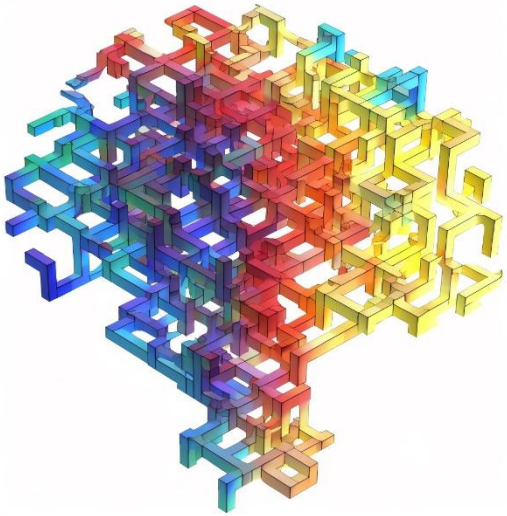


A	B	C	A	B	C	A	B	C
1	24	47	9	24	39	17	24	31
2	24	46	10	24	38	18	24	30
3	24	45	11	24	37	19	24	29
4	24	44	12	24	36	20	24	28
5	24	43	13	24	35	21	24	27
6	24	42	14	24	34	22	24	26
7	24	41	15	24	33	23	24	25
8	24	40	16	24	32	24	24	24

Principle

- Too many statistics published too accurately from a confidential database exposes the entire database with near certainty.

- Database can be reconstructed by treating the attributes of the persons as a collection of variables.
  - A set of constraints is then extracted from the published table.
- The database reconstruction finds a set of attributes that are consistent with the constraints.
  - There must be at least one solution because the statistics are known to be computed from a real database.
  - If the statistics are highly constraining, then there will be a single possible reconstruction.



Problem Formalization

- Mean: $A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 = 7 \cdot 38$
  - Assumed that $0 < A_i < 128$
- Median: $A_4 = 30$
  - Assumed that $A_1 \le A_2 \le A_3 \le A_4 \le A_5 \le A_6 \le A_7$
- Count: $S_1 + S_2 + S_3 + S_4 + S_5 + S_6 + S_7 = 4$
  - Assumed that female → 1 & male → 0
- Count: $R_1 + R_2 + R_3 + R_4 + R_5 + R_6 + R_7 = 4$
  - Assumed that black → 1 & white → 0
- Count: $M_k + M_{k+1} + \ldots + M_7 = 3$
  - Assumed that single → 0 & married → 1
  - Assume that $18 \le A_k \le A_{k+1} \le \ldots \le A_7$
- …

Age	Count	Median	Mean
Total population	7	30	38
Female	4	30	33.5
Male	3	30	40
Black	4	51	48.5
White	3	24	24
Single adults	NA	NA	NA
Married adults	4	51	54
Black female	3	36	36.7

Age	Sex	Race	Marital status
A1	S1	R1	M1
A2	S2	R2	M2
A3	S3	R3	M3
A4	S4	R4	M4
A5	S5	R5	M5
A5	S6	R6	M6
A7	S7	R7	M7

U.S. Census Attack

- Solving such a system of constraints is in general NP hard.
  - Using heuristics like SAT solvers and integer programming solvers.
- Same approach was effective against the full 2010 US Census.
- Internally they were able to reconstruct the microdata exactly for 46% of the population.
  - 71% if allowing errors in age of ±1 year.
- Linking this with commercial databases allowed them to correctly re-identify over 50 million individuals by name.
- This large-scale reconstruction attacks lead to the adoption of differential privacy for disclosure avoidance in the 2020 U.S. census.
  - To be discussed in detail.

United States® Census Bureau

Query Auditing

Problem

- Objective: given a database with some disclosure policy.
  - Attribute X is private (e.g., diagnosis), but aggregated values of X over different subsets of the records may be available (e.g., SUM, COUNT, MEDIAN, MAX).
  - Detect or prevent violations of the disclosure policy (i.e., disclosure of private information).

- SUM(Blood S.) 
- SUM(Blood S.) WHERE Sex = Female 
- SUM(Blood S.); SUM(Blood S.) WHERE Sex = Male 

Name	Sex	ZIP	Blood sugar	Diagnosis
John S.	Male	1123	4.3	Meningitis
John D.	Male	1123	5.2	Crohn
Jerry K.	Male	1114	6.1	Alzheimer
Jack. D.	Male	8423	3.2	Crohn
Eve A.	Female	1234	7.1	Facture

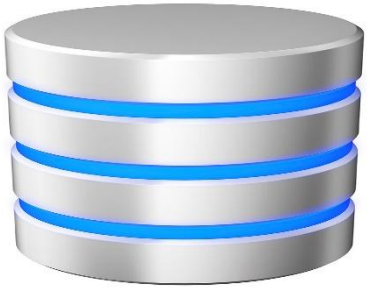
Detection vs Prevention

- Detection aka off-line query auditing
  - The process of examining queries that were answered in the past to determine whether answers to these queries could have been used to obtain confidential information forbidden by the disclosure policy.
- Prevention aka on-line query auditing
  - Examine current query in real-time and deny queries that could potentially cause a breach of privacy.

- Both ca be realized by suppressing information.

On-line vs Off-line

Database



1. Answer
2. Answer
   …
n. Answer

1. Query
2. Query
   …
n. Query

- Query 1
- Query 2
- …
- Query n

- Answer 1
- Answer 2
- …
- Answer n

Auditor

1. Answer
2. No Reply
   …
n. Answer

1. Query
2. Query
   …
n. Query

- Query 1
- Query 2
- …
- Query n

- No Reply

OR

- Answer 1
- …
- Answer n

Querier

Off-line Auditor for SUM over Reals

- SUM(Blood S.) WHERE Sex = Male
  - $x_1 + x_2 + x_3 + x_4 = 18.8$
- SUM(Blood S.) WHERE ZIP = 1123
  - $x_1 + x_2 = 9.5$
- SUM(Blood S.) WHERE ZIP > 1200
  - $x_4 + x_5 = 10.3$

Name	Sex	ZIP	Blood sugar
John S.	Male	1123	4.3
John D.	Male	1123	5.2
Jerry K.	Male	1114	6.1
Jack. D.	Male	8423	3.2
Eve A.	Female	1234	7.1

- Queries can be represented by a binary matrix.
- An auditor essentially needs to solve a system of linear equations.
  - If the matrix is invertible, all x_i's are disclosed (complexity $O(n^3)$).
  - Otherwise, the matrix can be diagonalized (complexity $O(n^3)$), where some x_i's are disclosed.

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 18.8 \\ 9.5 \\ 10.3 \end{bmatrix}$$

Off-line Auditor for SUM over Boolean

- SUM(HIV) WHERE Sex = Male
  - $x_1 + x_2 + x_3 + x_4 = 2$
- SUM(HIV) WHERE ZIP = 1123
  - $x_1 + x_2 = 1$
- SUM(HIV) WHERE ZIP > 1200
  - $x_4 + x_5 = 1$

Name	Sex	ZIP	HIV
John S.	Male	1123	True
John D.	Male	1123	False
Jerry K.	Male	1114	True
Jack. D.	Male	8423	False
Eve A.	Female	1234	True

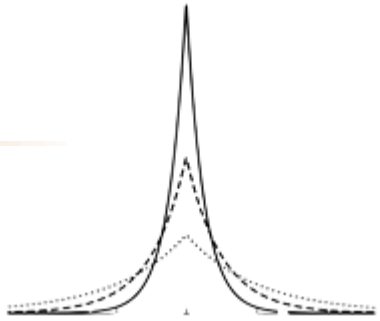
- Instead of solving a system of linear equations over integers, it is possible to relax the constraints for reals.
  - The final solutions need to be rounded: $x_i = 1$ if $x_i > ½$, otherwise $x_i=0$.

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$$

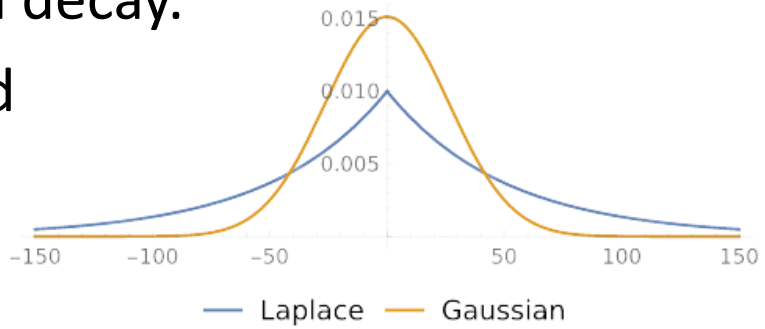
~~where $x_i \in \{0, 1\}$ for all i~~

where $0 \leq x_i \leq 1$ for all i

Perturbation



- Instead of denying answers, add noise to the result.
- Noise has zero mean and exponential decay.
- The variance of the noise is calibrated to the desired level of privacy.
  - Larger variance → stronger privacy.
  - Smaller variance → weaker privacy.



- Suppose the absolute error caused by the noise is at most E.

- This can still be solved efficiently with any LP-solver.

$$\begin{bmatrix} 2-E \\ 1-E \\ 1-E \end{bmatrix} \leq \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} \leq \begin{bmatrix} 2+E \\ 1+E \\ 1+E \end{bmatrix}$$

$$\text{where } 0 \leq x_i \leq 1 \text{ for all } i$$

Model

- Dataset D is a collection of n records, i.e., D = $\{d_1, d_2, \ldots, d_n\}$.
- Each record consist of some identifying information z and some secret b, i.e., $d_i = \{z_i, b_i\}$.
  - Identifying information: ZIP, gender, Bday, etc.
  - Secret: disease, sexual orientation, etc.

$$\left[\begin{array}{c|c} z_1 & b_1 \\ \vdots & \vdots \\ z_n & b_n \end{array}\right]$$

- The attacker's goal is to learn some piece of secret information about as many individuals as possible.

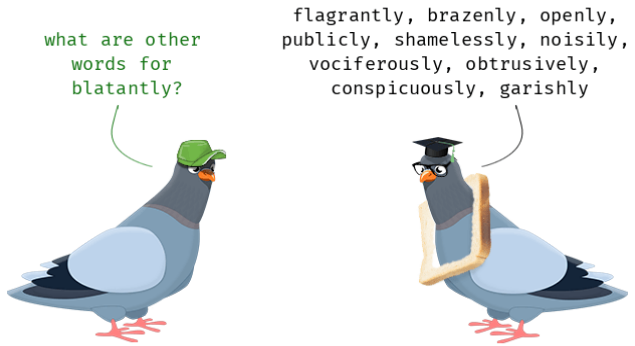
- Simplification
  - b_i is binary, i.e., $b_i \in \{0,1\}$.
  - z_i is unique for all $1 \leq i \leq n$.
  - Attacker knows z_i for all $1 \leq i \leq n$.

Dwork and Roth

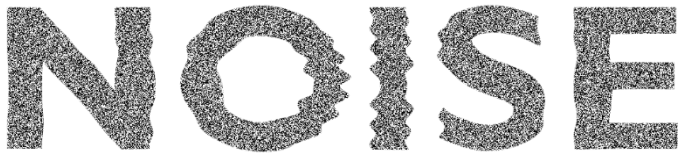
- Fundamental Law of Information Recovery
  - Giving overly accurate answers to too many questions will inevitably destroy privacy.
- The querier can specify queries which are a subset of [n].
- Subset query: $S \in \{0, 1\}^n$ vector with 1 for the indices included in the subset, and 0 otherwise.
  - For instance, "NAME = John S. OR Jack D."
- The true answer to query S is A(S).
- The auditor's output is r(S).
- r(S) = A(S) could easily lead to privacy violations.
  - Query $S = \{s_k\}$ where $s_k=1$ if k=i and 0 otherwise would reveal the secret bit of individual i.

Name	Sex	ZIP	Blood sugar	Diagnosis
John S.	Male	1123	4.3	Meningitis
John D.	Male	1123	5.2	Crohn
Jerry K.	Male	1114	6.1	Alzheimer
Jack. D.	Male	8423	3.2	Crohn
Eve A.	Female	1234	7.1	Facture

Dinur and Nissim

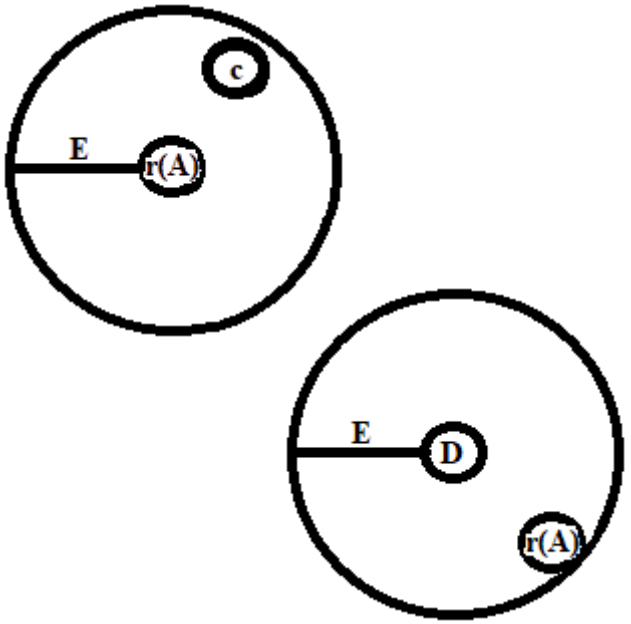


- An algorithm is blatantly non-private if an adversary can construct a database $c \in \{0, 1\}^n$ such that it matches the true database D in all but o(n) entries.
  - f(x)=o(g(x)): For *every* constant k > 0, the inequality $0 \le f(x) \le k \cdot g(x)$ holds asymptotically. The error could be arbitrarily small.
- Auditor output a noised version of A(S), such that $|r(S) - A(S)| < E$.
- If the analyst is allowed to ask 2^n subset queries, and the curator adds noise with some bound E, then the adversary can reconstruct the database in all but $4 \cdot E$ positions.
  - Adding 5% error to each answer means 80% reconstruction rate.



Proof

- The querier asks all 2^n subsets queries.
- For each candidate database $c \in \{0, 1\}^n$, if there exists a query set S such that $|\sum_S c_i - r(S)| > E$, then rule out c, otherwise, output c.
  - The true database D satisfies this condition, thus, there will be some output databases.
- Let I_0 be the set of indices where the secret bit is 0, i.e., $I_0 = \{i \mid b_i = 0\}$.
- Due to the adversary's strategy, $|\sum_{I_0} c_i - r(I_0)| \leq E$.
- Due to the auditor's strategy, $|\sum_{I_0} b_i - r(I_0)| \leq E$.
- Due to the triangle inequality, $|\sum_{I_0} b_i - c_i| \leq 2 \cdot E$.
- The same holds for I_1 where $I_1 = \{i \mid b_i = 1\}$, consequently $|\sum_{\{I_0\} \cup \{I_1\}} b_i - c_i| \leq 4 \cdot E$.



Stronger Attack

- The attack requires exponentially many queries.
- If the querier is allowed to ask O(n) random subset queries, and the auditor adds noise with some bound $E = O(\alpha \cdot \sqrt{n})$, then a computationally efficient adversary can reconstruct the database in all but $O(\alpha^2)$ positions.
  - f(x)=O(g(x)): There *exists* a constant k > 0 such that the inequality $0 \le f(x) \le k \cdot g(x)$ holds asymptotically. The error cannot be arbitrarily high.

- Also holds when the auditor makes a constant fraction of the responses with arbitrary noise magnitudes.
- If the queries are chosen randomly (i.e., each record is covered by a query with probability ½), then $O(n \cdot \log^2 n)$ queries are needed for almost full disclosure.

Diffix Attack

- In 2017, Aircloak released Diffix where an unlimited numbers of queries are allowed while preserving user privacy.
  - Provided the first bounty program for anonymized data reidentification. An effective reconstruction attack would be awarded with $5,000.
- It limit the ability of an adversary to use the full power of SQL to access the database.
- It has a collection of data-dependent ad-hoc methods to prevent leaking information about individuals or very small subsets of users.

- It adds a data-dependent noise term with a constant variance.
- It adds a query-dependent noise term whose variance depends on the complexity of the query.
  - Creating subset queries via accumulated conditions such as “NAME = John S. OR Jack D.” would introduce too much noise.

Aloni & Nissim

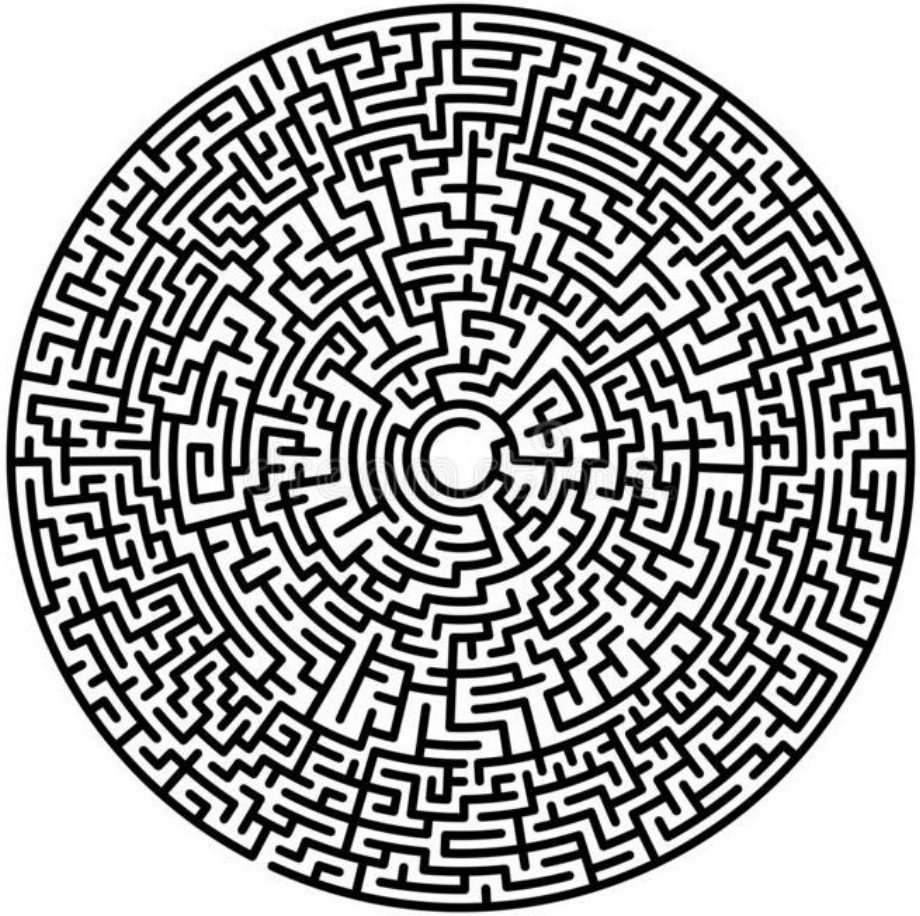
- Instead of choosing a subset and then producing conditions to specify it, they created a query involving a low number of conditions which specifies a “random” set.

- Each user in the dataset has a unique client ID.
- They use functions specified by four variables.
  - mult, exp, d, pred
- Does the d-th digit of (mult · client-id)^exp satisfy pred?
  - Mult=17,exp=0.5, d=3, pred=Is the digit even?
    Client ID=1 → $(17 \cdot 1)^{0.5} = 4.123\ldots$, so it is included
    Client ID=2 → $(17 \cdot 2)^{0.5} = 5.830\ldots$, so it is excluded
    …

Dick–Joseph–Schutzman

- As a response, Aircloak further restrict the queries allowed by Diffix.
  - Columns like clientId, where most of the values correspond to a single user, are tagged as ‘isolating’, and mathematical functions can no longer be used on such columns.
- The hope was that this modification would prevent the extraction of entropy from an identifying column via hashing.
- Without uniquely identifying column from the database, another way is needed to single out rows.
  - Recap: almost 90 percent of Americans could be identified with only a date of birth, ZIP code, and gender.

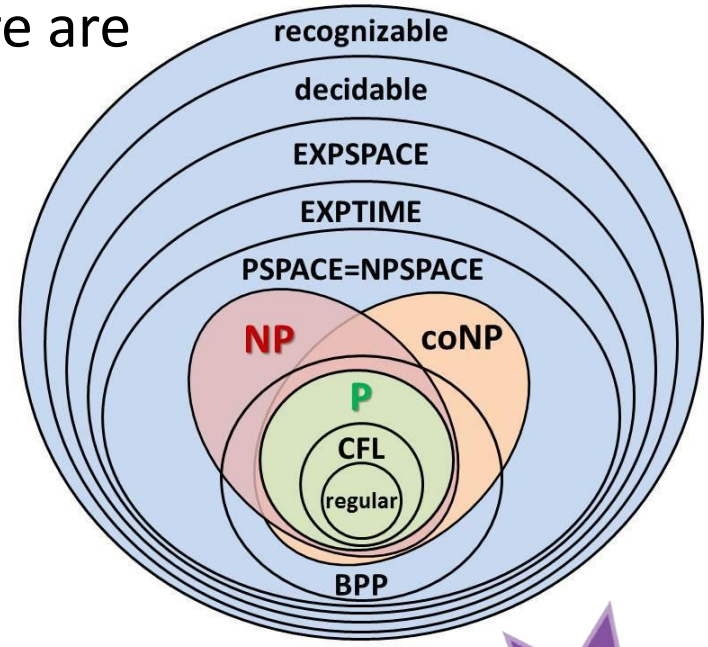
- Use the same idea by choosing multiple non-isolating columns which, when taken together, can isolate rows in the database.



SAT

Auditing Results

- If the attribute values are real, then there are efficient polynomial auditors for {SUM}, {MEDIAN}, {AVG}, and {MAX, MIN}.
- If the attribute values are boolean, then the problem for {SUM} is in coNP.
  - If the queries are 1-dimensional, then there is an efficient polynomial auditor.
    1D: the total number of HIV-positive people in various age groups.
    2D: the total number of HIV-positive depending on age and blood sugar.
- There is no polynomial time full-disclosure auditing algorithm for {SUM, MAX} queries unless P=NP.



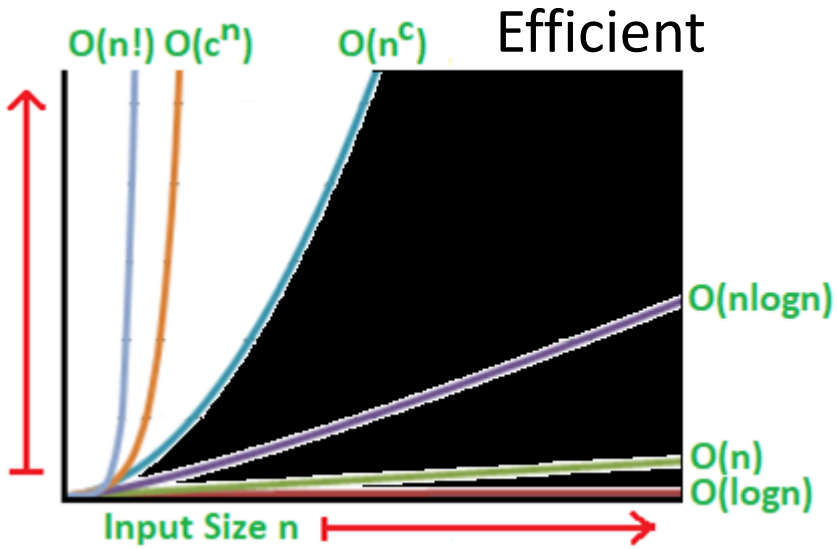
DeID 5

Results

Complexity

- P: Solvable in polynomial time.
- NP: “YES” answers checkable in polynomial time by a non-deterministic Turing machine.
  - coNP: same with “NO”.
- NP-complete: Every problem in NP could be reduced to it.

- Examples
  - P: decide if a number is prime.
  - NP: decide if X edge jumps are enough to visit in a given graph all desired nodes.
  - NP-complete: decide if a logical formula could be evaluated to True.



Boolean Satisfiability Problem

- For a given Boolean formula, replacing each variable with either true or false can make the formula evaluate to true?
  - $A \wedge B$ is satisfiable.
  - $A \wedge \neg A$ is unsatisfiable.

- There is no known algorithm that efficiently solves each SAT problem, and it is generally believed that no such algorithm exists (P vs NP).
- However, heuristic SAT-algorithms can handle tens of thousands of variables and formulas consisting of millions of symbols.

- SAT is the first problem that was proven to be NP-complete (Cook-Levin theorem).
  - Many problems can be cast to SAT solving and hence use these highly optimized solvers.

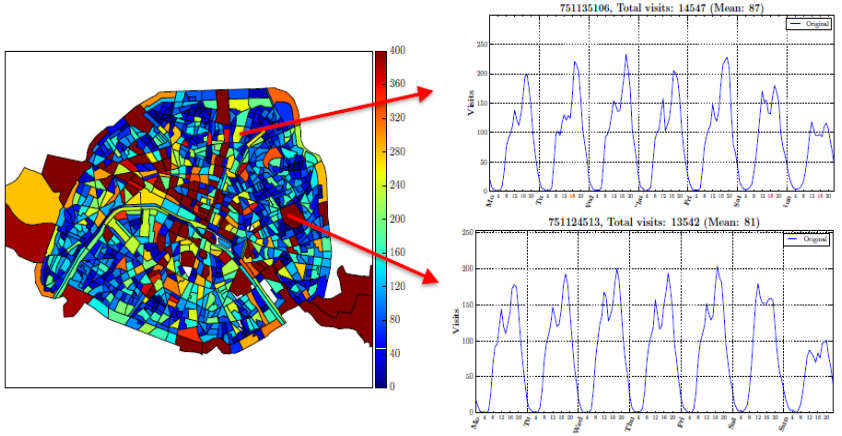
Location Data

Membership Test

- Telecom companies wish to publish spatio-temporal densities computed from call detail records (i.e., number of cell phones at each tower depending on the time).
  - To optimize traffic, identify the spots of new services, etc.

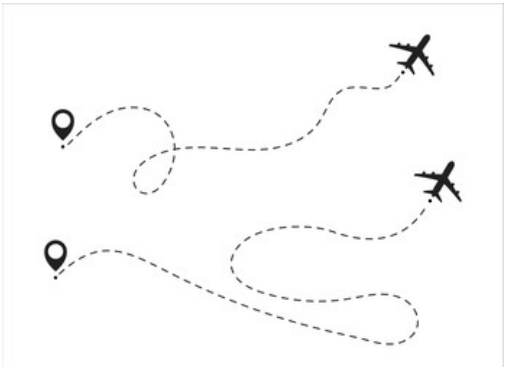
- x = {1, …, 168} denotes the hour of a week.
- T[x] is the number of people who visited the location in slot x.

- Can this aggregate leak information about a single person?



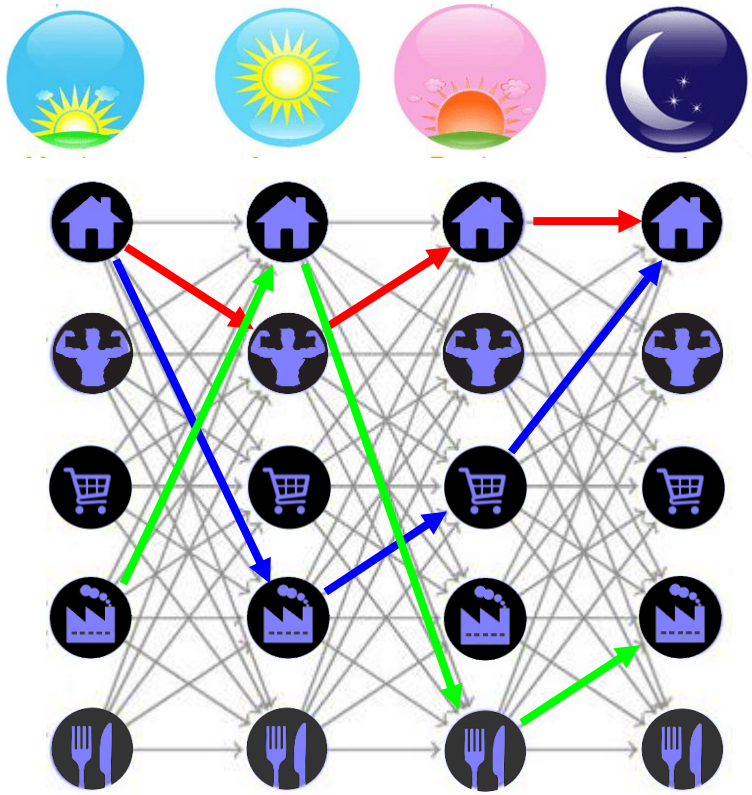
Location Trajectories

- Knowing only the number of visits per tower in each hour, it is possible to reconstruct all the towers of a person with more than 70% of accuracy.
  - Predictability: the next tower of a person is geographically close to the previous tower.
  - Regularity: every person visits the same (or very similar) towers each day because of daily routines.
  - Uniqueness: 4-5 towers visited by an individual are unique to that individual (there are no other individuals who visited these towers).



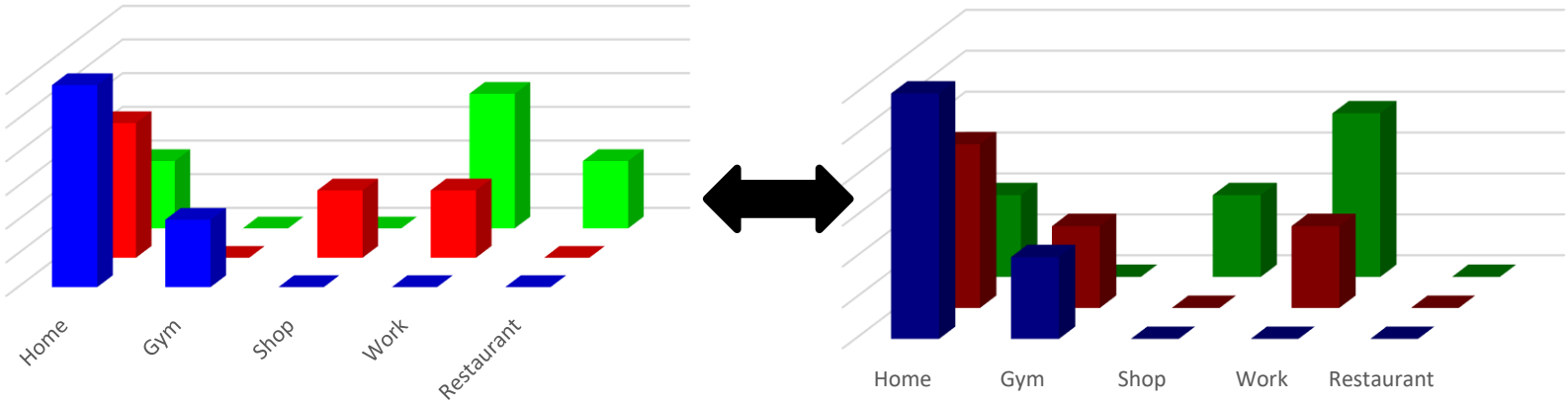
Trace Reconstruction

- Reconstruct traces within each day independently.
  - Exploit predictability.
- Assign weights between locations following geo-distance.
- Find an optimal assignment of visits between consecutive slots such that
  - Objective: the total weights of the assignment is minimized.
  - Constrain: at each time slot the given number of individuals are present in each location.
- Solving the reconstruction problem recovers the globally most likely moving patterns.

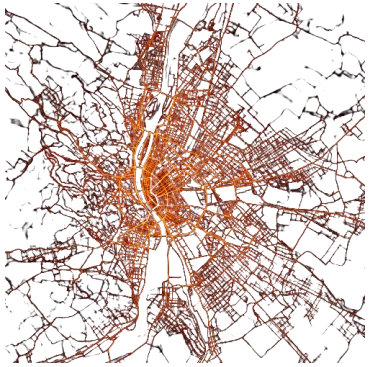


Linking Traces

- Link traces together across several days.
  - Exploit regularity.
- Restart the reconstruction at the beginning of each day.
- For every reconstructed trace within a day, compute the frequency of towers.
  - Get a distribution of towers per trace.
- Those traces belong to the same person, whose tower distribution is similar (according to some similarity measure).



Matching Individuals



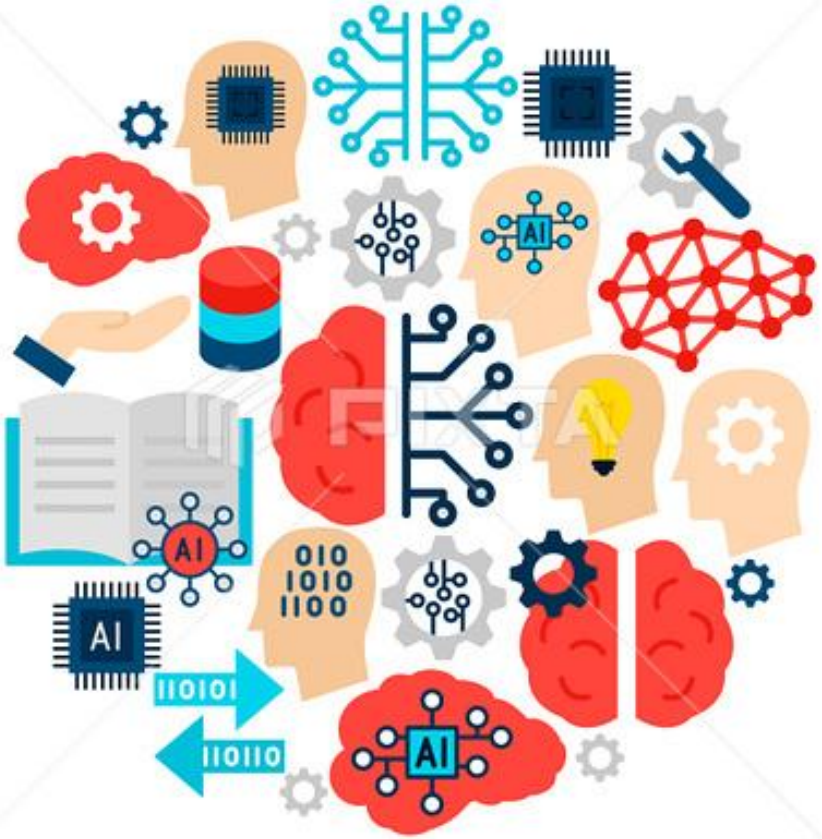
- Assign individuals to reconstructed traces.
  - Exploit uniqueness of location data.

- Do not know which trace belongs to whom.
- 4 location visits of a person is unique with a chance of 95% within a population of 1.5 million users.
  - If I know 4 locations and the time when (s)he were there, then (s)he is the only such person with these positions.

- Crawl Facebook for locations & time coordinates.
- Find the trace which has these 4 points.

DeID 6

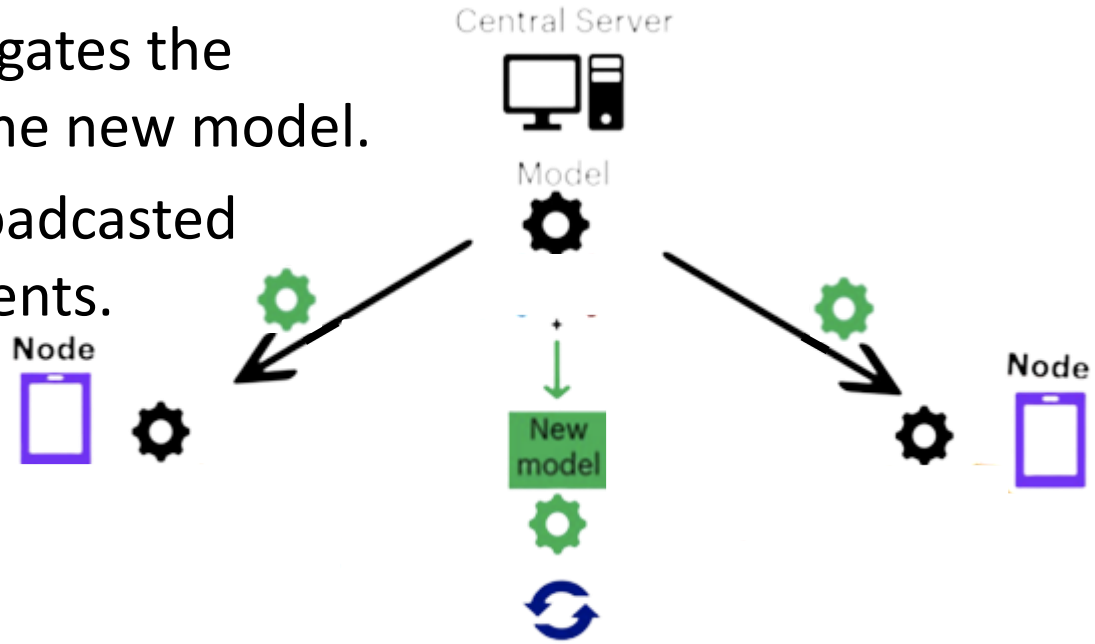
Example



Federated Learning

Federated Learning

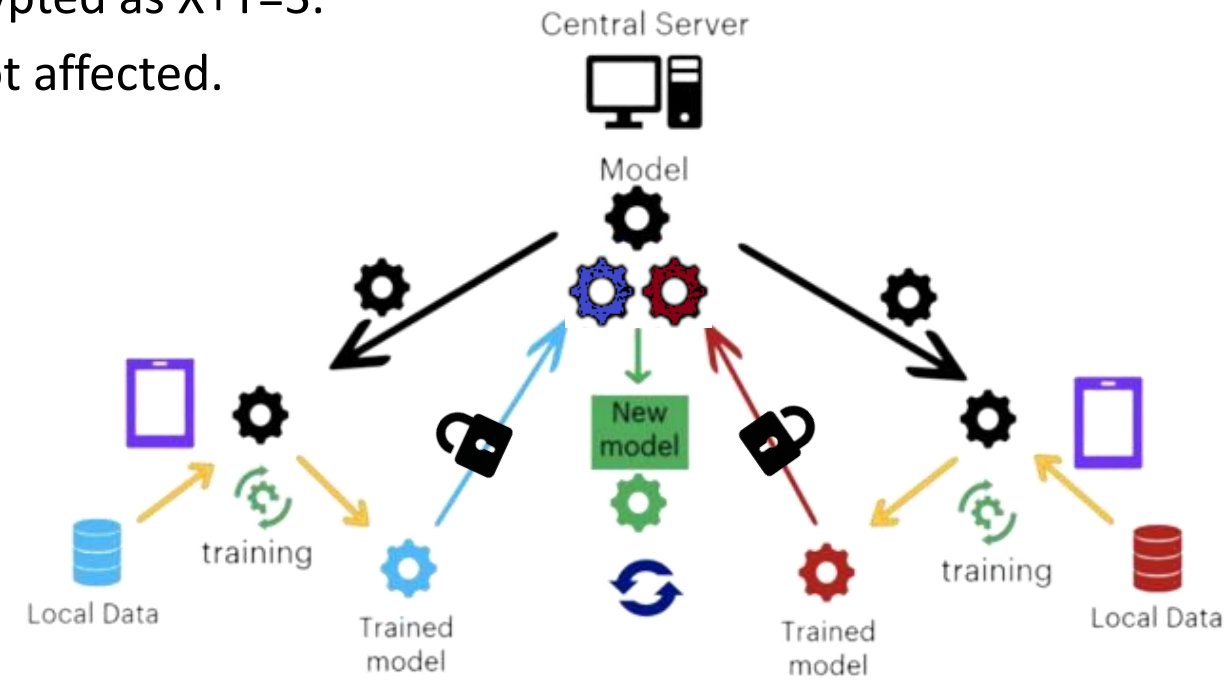
0) The aggregator server initializes the model, i.e., determine the hyperparameters.
1) The aggregator broadcast the model to some clients.
2) Those clients train that model on their local dataset and send the update to the aggregator.
3) The aggregator aggregates the model updates into the new model.
4) The final model is broadcasted to all participating clients.
5) Repeat from 1) until convergence is reached.



Secure Aggregation

- In FL the server learns the individual changes; hence, it can get information about underlying sensitive data.
- Secure Aggregation hides individual gradients with masks which cancel out after aggregation.
  - Relies on cryptography.
    E.g., 1+2=3 encrypted as X+Y=3.
  - Final model is not affected.

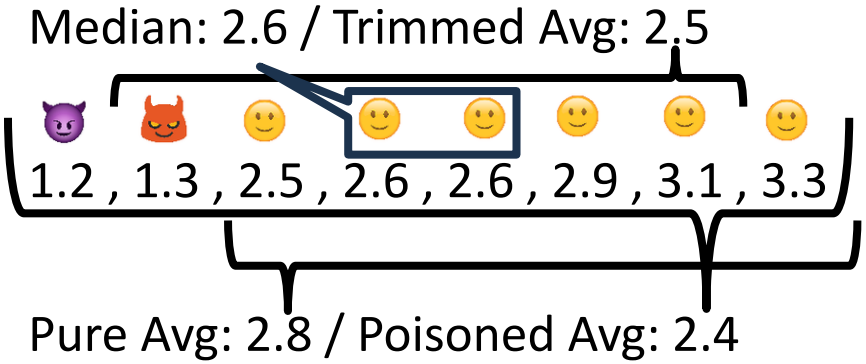
- Attribution is not possible without background knowledge.

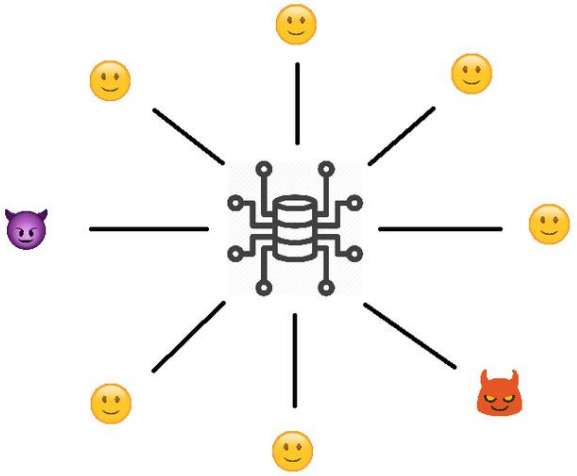


Byzantine Resilience

- The ability of the system to withstand malicious behaviors exhibited by participating devices (Byzantine faults).

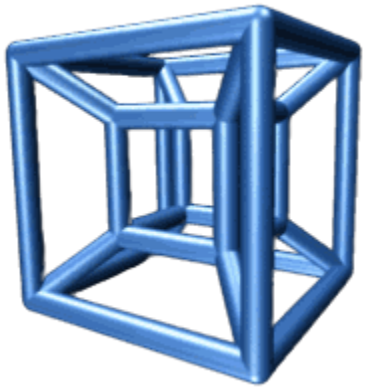
- It ensuring the integrity of the collective model updates.
- Usually achieved by employing robust algorithms that can tolerate and mitigate the impact of such adversarial actions.
  - Trimmed Mean
  - Median





Take Away

- Entropy decreases fast due to the curse of dimensionality.
  - More information decreases the anonymity set.
- Too many too accurately statistics could expose the entire database.
  - Reconstruction attacks are possible not just theoretically.
  - Possible defense is suppression or noise injection.



- Query auditing can be on-line (prevention) or off-line (detection).
  - For non-linear or arbitrary queries, one can use SAT solvers.

- A practical attack may recover exact location data using regularity, predictability, and uniqueness.
- Federated Learning is also vulnerable, Secure Aggregation and Byzantine Resilience are possible privacy and security defense mechanisms.

Control Questions

- What can you measure with entropy in relation to database reconstruction? What is the corresponding formula?
- What is the Fundamental Law of Information Recovery and when is an algorithm Blatantly Non-private?
- What are Federated Leaning, Secure Aggregation, and Byzantine Resilience?

References

- Tweetorial: Reconstruction-abetted re-identification attacks and other traditional vulnerabilities
- The algorithmic foundations of differential privacy
- Revealing information while preserving privacy
- Linear program reconstruction in practice
- A survey of Query auditing techniques for data privacy
- Knock Knock, Who's There? Membership Inference on Aggregate Location Data
- Where You Are Is Who You Are: User Identification by Matching Statistics
- Algorithms for Private Data Analysis: Reconstruction Attacks
- Diffix Cedar bounty prize