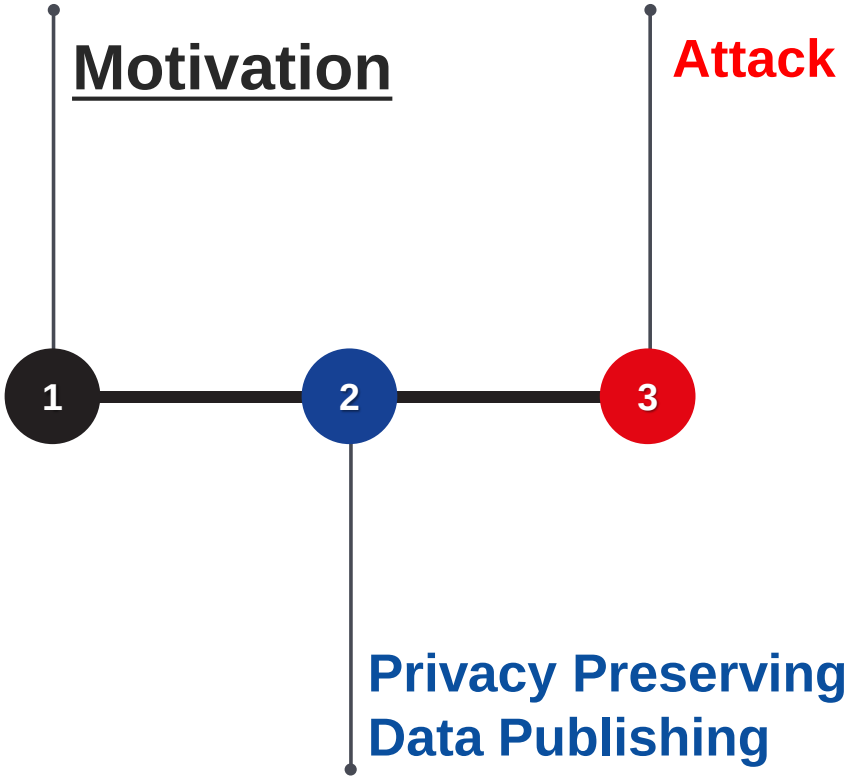
NANYANG
THE INNOVATIVE POLYTECHNIC

Data Privacy and Protection

Topic 2

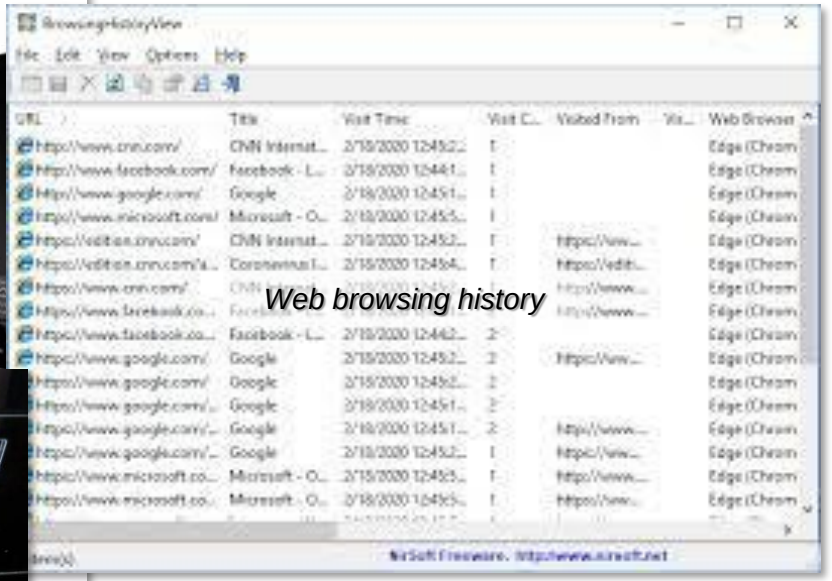
Attack Models

Contents

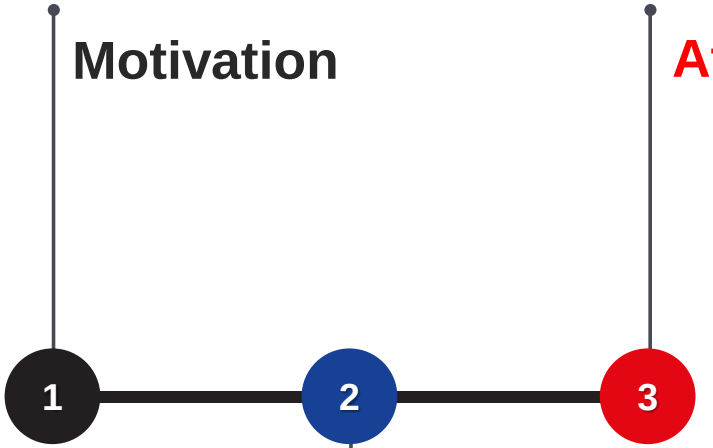

Motivation
Attack Models
1
2
3
Privacy Preserving
Data Publishing

Motivation

Intelligent Fridge
FOOD MANAGER
Right Door
Add New Food
Food Management
Information of contents in the fridge

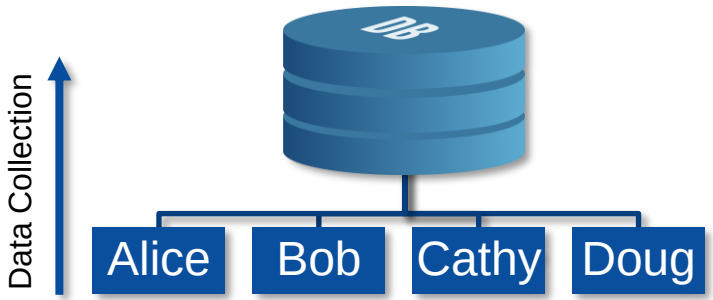

BrowsingHistoryView
File Edit View Options Help
URL
Title
Visit Time
Visited From
Web Browser
Web browsing history

Contents



Privacy Preserving Data Publishing

Data Collection and Data Publishing

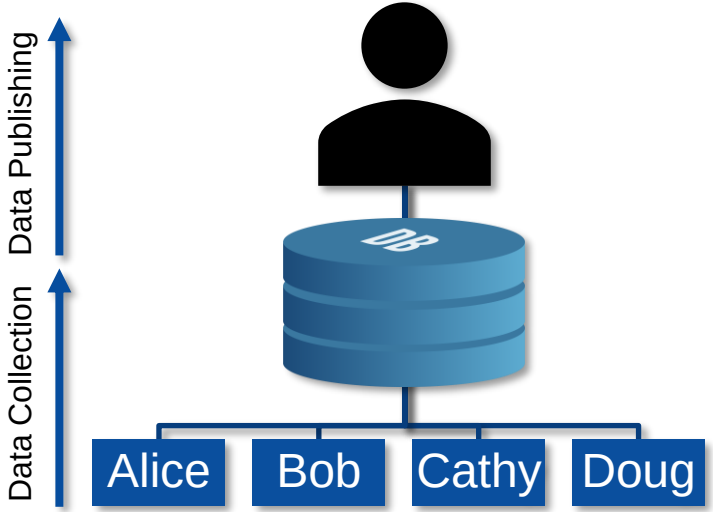


Role

Data Publisher

Record Owners

Data Collection and Data Publishing



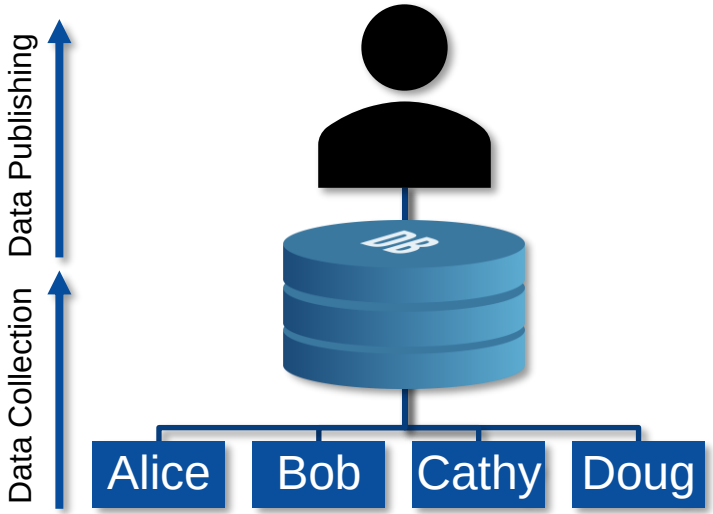
Role

Data Recipient

Data Publisher

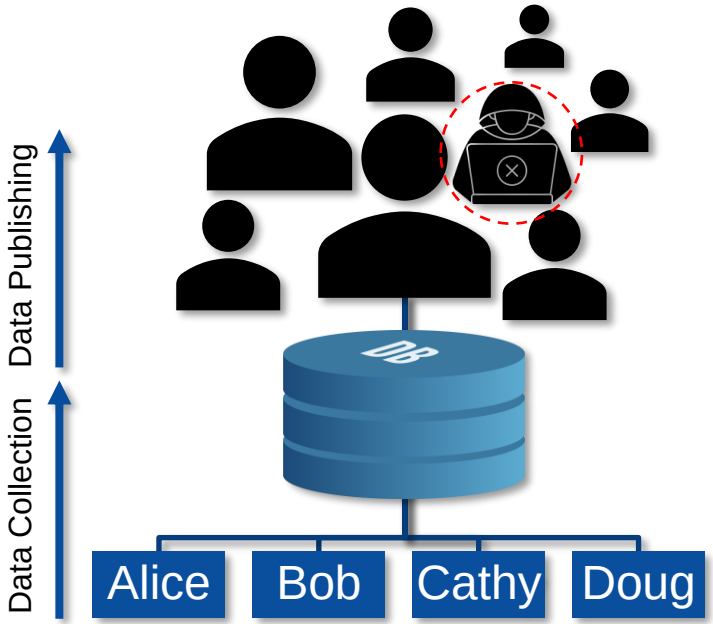
Record Owners

Data Collection and Data Publishing



Role	Example
Data Recipient	Medical Center (data mining)
Data Publisher	Hospital (data anonymization)
Record Owners	Patients

Privacy-Preserving Data Publishing (PPDP)



Role	Example
Data Recipient	***Pharmaceutical Company*** (data mining)
Data Publisher	Hospital (data anonymization)
Record Owners	Patients

Data Attributes

Data Attributes	Description
Explicit Identifier	Data attributes that **explicitly identifies** record owners, e.g., name, identity card number, mobile phone number.
Quasi Identifier (QID)	Data attributes that could **potentially identify** record owners, e.g., postal code, age, gender.
Sensitive Attributes	Data attributes that are **sensitive person-specific information**, e.g., salary, disease, disability status
Non-Sensitive Attributes	Data attributes that do ***not*** fall into all of the other categories.

Data Attributes – Example

Name	DOB	Gender	Postal Code	Disease
Alice	13/4/86	Female	153715	Flu
Bob	21/1/79	Male	153715	Heart Disease
Cathy	28/2/76	Female	183210	Hepatitis
Doug	21/1/79	Male	183210	Broken Leg
Eliza	28/2/90	Female	191264	Asthma

Explicit Identifier

Data Attributes – Example

Name	DOB	Gender	Postal Code	Disease
Alice	13/4/86	Female	153715	Flu
Bob	21/1/79	Male	153715	Heart Disease
Cathy	28/2/76	Female	183210	Hepatitis
Doug	21/1/79	Male	183210	Broken Leg
Eliza	28/2/90	Female	191264	Asthma

Explicit Identifier

Quasi Identifier

Data Attributes – Example

Name	DOB	Gender	Postal Code	Disease
Alice	13/4/86	Female	153715	Flu
Bob	21/1/79	Male	153715	Heart Disease
Cathy	28/2/76	Female	183210	Hepatitis
Doug	21/1/79	Male	183210	Broken Leg
Eliza	28/2/90	Female	191264	Asthma

Explicit Identifier

Quasi Identifier

Sensitive Attribute

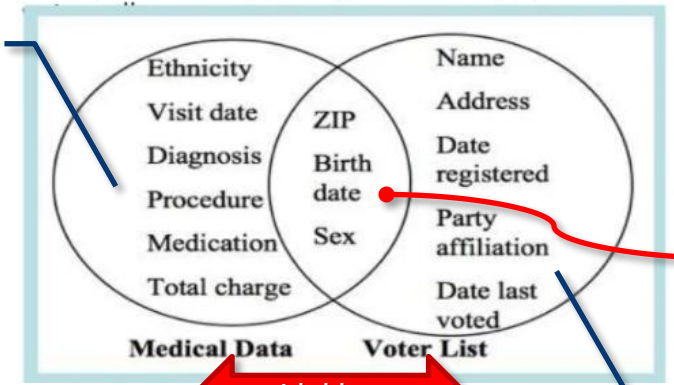
Quasi Identifier

William Weld vs Latanya Sweeney

Massachusetts Group Insurance Commission (1997):
Anonymized medical history of state employees (all hospital visits, diagnosis, prescriptions)

Latanya Sweeney (MIT grad student): $20 – Cambridge

Published medical database – privacy is preserved by removing *explicit identifiers*, leaving only *quasi identifiers* and *sensitive attributes*

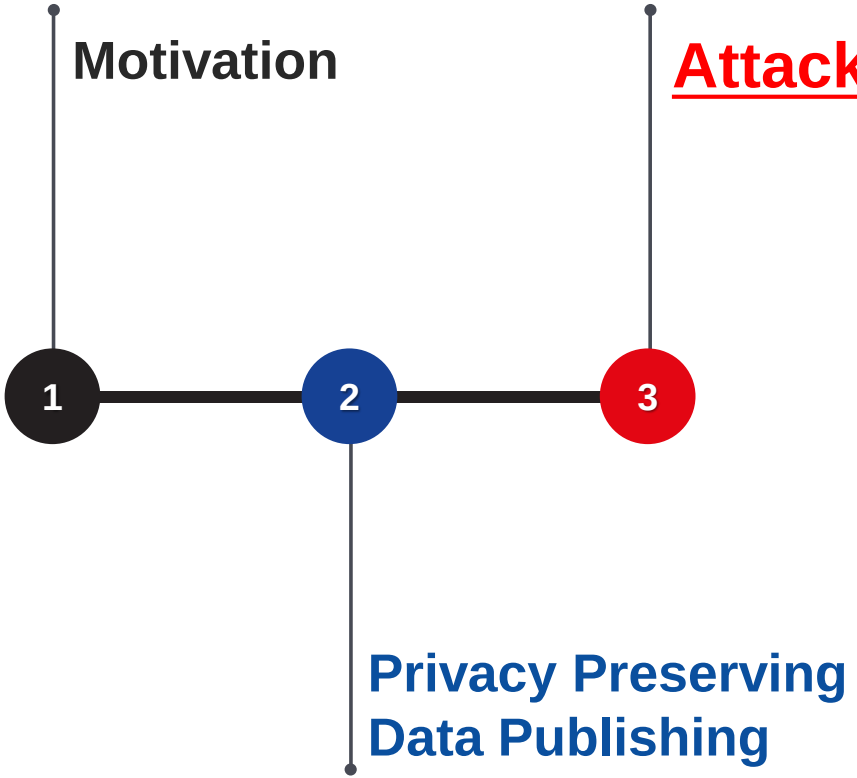


Re-identified
- ❑ Resident of 02138 (Cambridge, Massachusetts)
- ❑ Born July 31, 1945
- ❑ Male

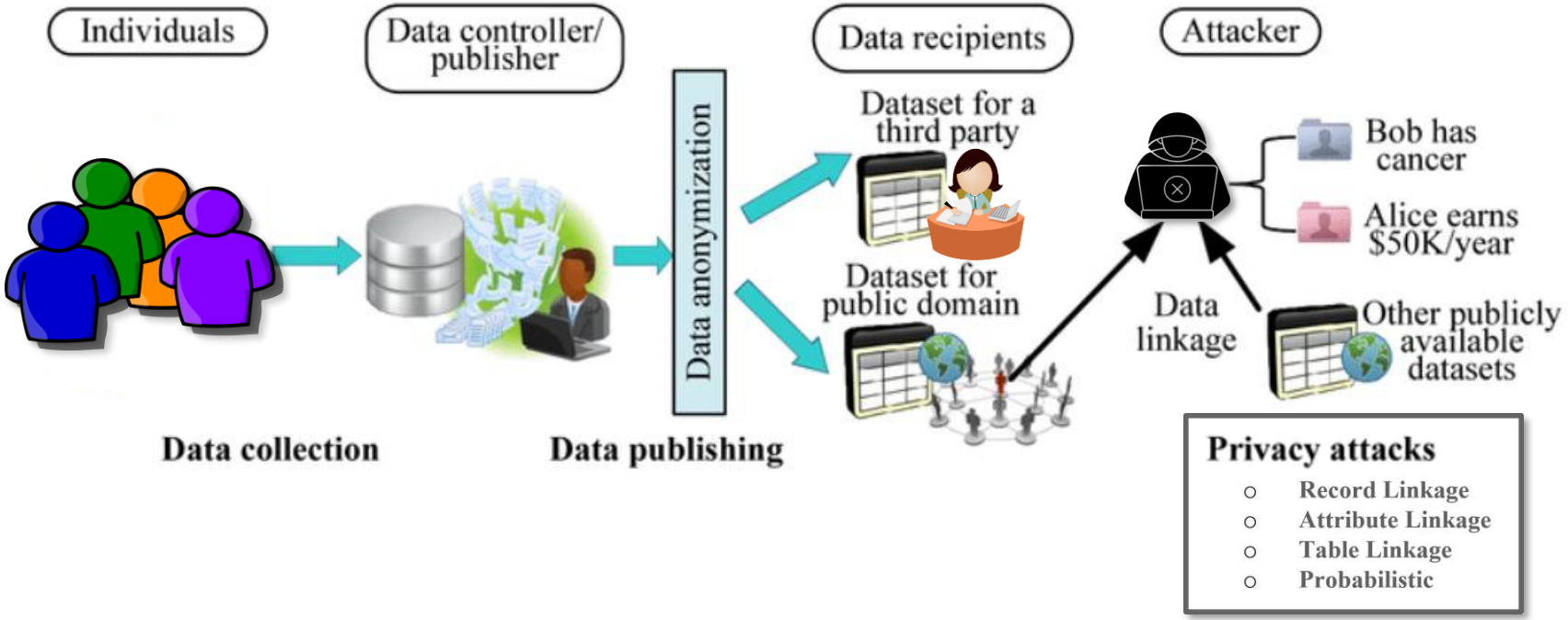
Searchable **public voter information** in the US

https://www.slideshare.net/KrishnaramKenthapadi/fairness-transparency-and-privacy-in-ai-linkedin

Contents


Motivation
Attack Models
1
2
3
Privacy Preserving
Data Publishing

Data Privacy Attack Models



Record Linkage Model

- Similar Quasi Identifier (QID) values grouped into small number of records
- Victim’s QID matches and linked to this group
- Smaller number of possibilities in identifying the victim’s record
- Identifying the victim in this group, with additional information

Record Linkage Model – Example

Job	Sex	Age	Disease
Engineer	Male	35	Hepatitis
Engineer	Male	38	Hepatitis
Lawyer	Male	38	HIV
Writer	Female	30	Flu
Writer	Female	30	HIV
Dancer	Female	30	HIV
Dancer	Female	30	HIV

Table 1 Patient Data

Example: Hospital wants to publish the patient records in Table 1 to a research center

Record Linkage Model – Example

Job	Sex	Age	Disease
Engineer	Male	35	Hepatitis
Engineer	Male	38	Hepatitis
Lawyer	Male	38	HIV
Writer	Female	30	Flu
Writer	Female	30	HIV
Dancer	Female	30	HIV
Dancer	Female	30	HIV

Table 1 Patient Data

Name	Job	Sex	Age
Alice	Writer	Female	30
Bob	Engineer	Male	35
Cathy	Writer	Female	30
Doug	Lawyer	Male	38
Emily	Dancer	Female	30
Fred	Engineer	Male	38
Gladys	Dancer	Female	30
Henry	Lawyer	Male	39
Irene	Dancer	Female	32

Table 2 External Data

Example: Hospital wants to publish the patient records in Table 1 to a research center

- Research center has access to the external table, Table 2
- Research center knows that every person with a record in Table 2 has a record in Table 1

Record Linkage Model – Example

Job	Sex	Age	Disease
Engineer	Male	35	Hepatitis
Engineer	Male	38	Hepatitis
Lawyer	Male	38	HIV
Writer	Female	30	Flu
Writer	Female	30	HIV
Dancer	Female	30	HIV
Dancer	Female	30	HIV

Table 1 Patient Data

Name	Job	Sex	Age
Alice	Writer	Female	30
Bob	Engineer	Male	35
Cathy	Writer	Female	30
Doug	Lawyer	Male	38
Emily	Dancer	Female	30
Fred	Engineer	Male	38
Gladys	Dancer	Female	30
Henry	Lawyer	Male	39
Irene	Dancer	Female	32

Table 2 External Data

Example: Hospital wants to publish the patient records in Table 1 to a research center

- Research center has access to the external table, Table 2
- Research center knows that every person with a record in Table 2 has a record in Table 1
- Joining the two tables on the common attributes Job, Sex, and Age may link the identity of a person to his/her Disease
- Doug is identified by quasi identifier, Lawyer, Male, 38

Attribute Linkage Model

- Adversary may not precisely identify the record of the target victim
- Victim belongs to a group, based on a set of Sensitive Attributes
- Adversary could infer victim’s sensitive values from the published data

Attribute Linkage Model – Example

Job	Sex	Age	Disease
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	HIV
Artist	Female	[30-35)	Flu
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV

Table 3 Patient Data

Example: Hospital anonymizes the data, Job/Age, into a range, to reduce record linkage

Attribute Linkage Model – Example

Job	Sex	Age	Disease
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	HIV
Artist	Female	[30-35)	Flu
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV

Table 3 Patient Data

Example: Hospital anonymizes the data, Job/Age, into a range, to reduce record linkage

- Adversary has knowledge that the target victim Emily, is a dancer, is 30 years old, and has a record in the published data

Attribute Linkage Model – Example

Job	Sex	Age	Disease
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	HIV
Artist	Female	[30-35)	Flu
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV

Table 3 Patient Data

Example: Hospital anonymizes the data, Job/Age, into a range, to reduce record linkage

- Adversary has knowledge that the target victim Emily, is a dancer, is 30 years old, and has a record in the published data
- Adversary may infer that Emily has HIV with 75% confidence, because 3 out of 4 artists at age 30-35 have HIV

Table Linkage Model

- Adversary can confidently infer the presence, or the absence, of the victim's record in the published data
- If a hospital publishes data with a particular type of disease
- Inferring presence of the victim's record in the table is already damaging

Table Linkage Model – Example

Job	Sex	Age	Disease
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	HIV
Artist	Female	[30-35)	Flu
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV

Table 3 Patient Data

Example: Hospital publishes patient data in Table 3 – table linkage attack on target victim, Alice

Table Linkage Model – Example

Job	Sex	Age	Disease
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	HIV
Artist	Female	[30-35)	Flu
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV

Table 3 Patient Data

Name	Job	Sex	Age
Alice	Artist	Female	[30-35)
Bob	Professional	Male	[35-40)
Cathy	Artist	Female	[30-35)
Doug	Professional	Male	[35-40)
Emily	Artist	Female	[30-35)
Fred	Professional	Male	[35-40)
Gladys	Artist	Female	[30-35)
Henry	Professional	Male	[35-40)
Irene	Artist	Female	[30-35)

Table 4 External Data

Example: Hospital publishes patient data in Table 3 – table linkage attack on target victim, Alice

- Adversary is presumed to also have access to external public data in Table 4

Table Linkage Model – Example

Job	Sex	Age	Disease
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	HIV
Artist	Female	[30-35)	Flu
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV

Table 3 Patient Data

Name	Job	Sex	Age
Alice	Artist	Female	[30-35)
Bob	Professional	Male	[35-40)
Cathy	Artist	Female	[30-35)
Doug	Professional	Male	[35-40)
Emily	Artist	Female	[30-35)
Fred	Professional	Male	[35-40)
Gladys	Artist	Female	[30-35)
Henry	Professional	Male	[35-40)
Irene	Artist	Female	[30-35)

Table 4 External Data

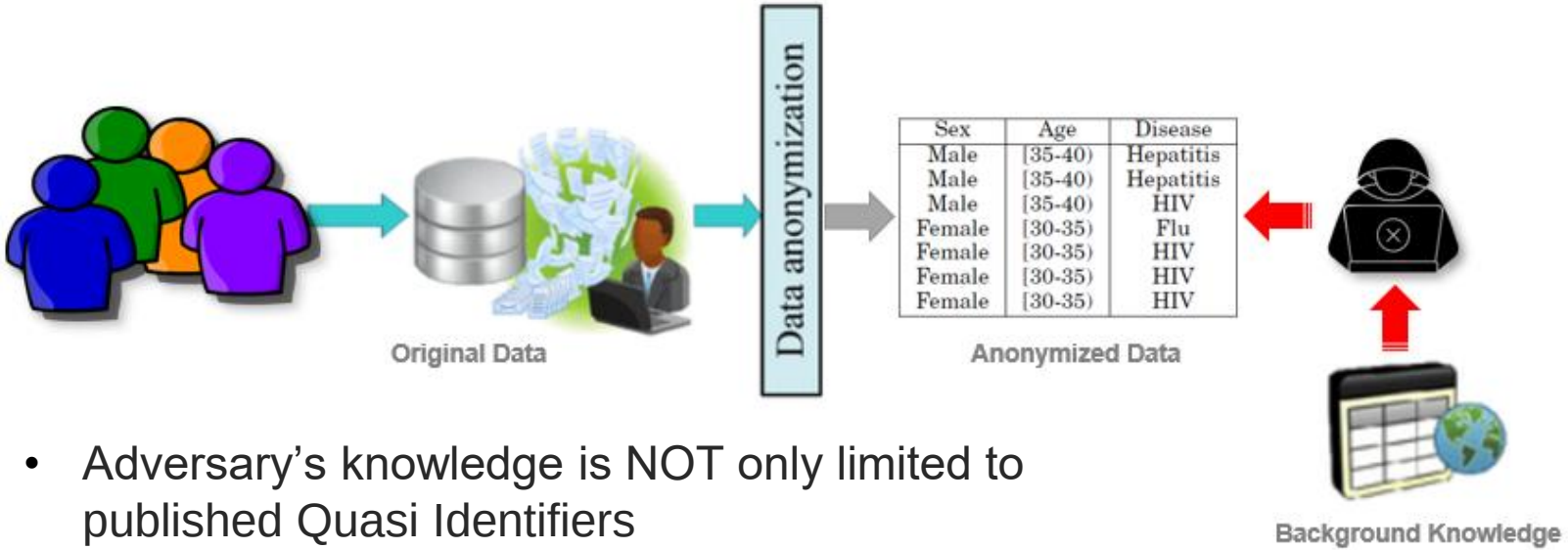
Example: Hospital publishes patient data in Table 3 – table linkage attack on target victim, Alice

- Adversary is presumed to also have access to external public data in Table 4
- 4/5 or 80% probability that Alice has HIV

Probabilistic Model

- Does not focus on records, attributes, or tables that can be linked to a target victim
- Compare probability before and after access the published data
- Adversary believes that the probability of identifying the target victim's sensitive information, increases after accessing the published data, compared to the probability before

Background Knowledge

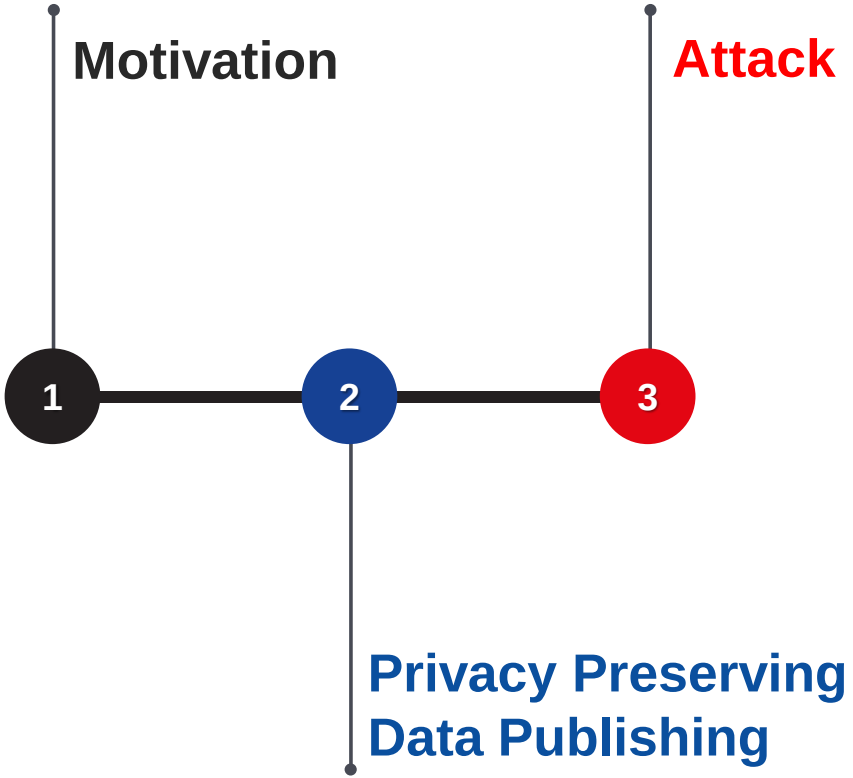


- Adversary's knowledge is NOT only limited to published Quasi Identifiers
- Privacy-preserving data publishing has to take additional Background Knowledge into consideration
- Includes, public statistical data, social networks like Facebook and LinkedIn, common sense, etc.

Can individuals be re-identified from anonymous data?

https://youtu.be/FohAm_nzx_w (CNBC Television)

Contents


Motivation
Attack Models
1
2
3
Privacy Preserving
Data Publishing