

## $l$ -Diversidade e $\delta$ -Presença

Prof. Javam Machado

LSBD/DC/UFC

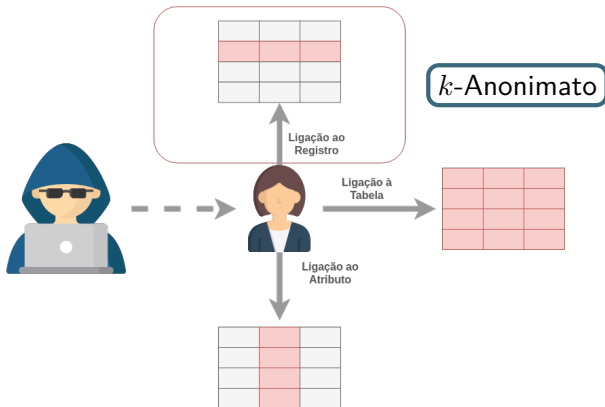
Setembro/2019

# $l$ -Diversidade

# Revisão

# Tipos de Ataques

## ■ Ataques de ligação ao registro



# Modelo $k$ -Anonimato

As classes de equivalência devem ter tamanho mínimo  $k$ .

# Exemplo

## ■ 2-Anonimato com Generalização

Profissão	Idade	Cidade	Doença
Engenheiro	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	Fortaleza	AIDS
Programador	32	Quixadá	AIDS
Dentista	35	Quixadá	AIDS

# Exemplo

## ■ 2-Anonimato com Generalização

Profissão	Idade	Cidade	Doença
Engenheiro	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	Fortaleza	AIDS
Programador	32	Quixadá	AIDS
Dentista	35	Quixadá	AIDS

# Exemplo

## ■ 2-Anonimato com Generalização

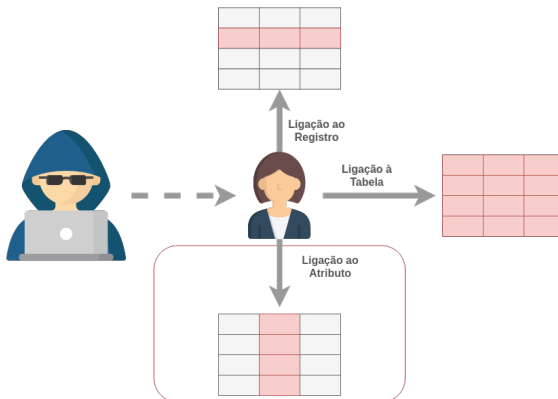
Profissão	Idade	Cidade	Doença
Profissional	[25-35]	CE	Chikungunya
Médico	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	CE	AIDS
Programador	32	CE	AIDS
Profissional	[25-35]	CE	AIDS



# Motivação

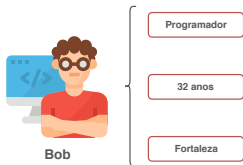
# Tipos de Ataques

## ■ Ataques de ligação ao atributo



# Modelo $k$ -Anonimato

- É suficiente?
  - Qual o registro de Bob no dado abaixo?



Profissão	Idade	Cidade	Doença
Profissional	[25-35]	CE	Chikungunya
Médico	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	CE	AIDS
Programador	32	CE	AIDS
Profissional	[25-35]	CE	AIDS

# Modelo $k$ -Anonimato

- É suficiente?
  - Qual o registro de Bob no dado abaixo?



Bob

Programador

32 anos

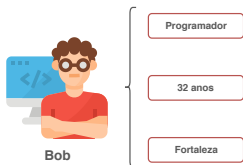
Fortaleza

Profissão	Idade	Cidade	Doença
Profissional	[25-35]	CE	Chikungunya
Médico	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	CE	AIDS
Programador	32	CE	AIDS
Profissional	[25-35]	CE	AIDS

✓ Satisfaz 2-Anonimato

# Modelo $k$ -Anonimato

- É suficiente?
  - Qual o registro de Bob no dado abaixo?



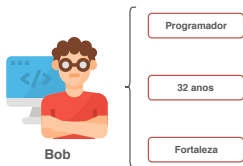
Profissão	Idade	Cidade	Doença
Profissional	[25-35]	CE	Chikungunya
Médico	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	CE	AIDS
Programador	32	CE	AIDS
Profissional	[25-35]	CE	AIDS

✓ Satisfaz 2-Anonimato

✓ Protege contra ataque de ligação ao **registro**

# Modelo $k$ -Anonimato

- É suficiente?
  - Qual o registro de Bob no dado abaixo?



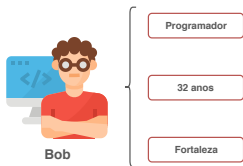
Profissão	Idade	Cidade	Doença
Profissional	[25-35]	CE	Chikungunya
Médico	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	CE	AIDS
Programador	32	CE	AIDS
Profissional	[25-35]	CE	AIDS



Ataque de ligação ao **atributo**

# Modelo $k$ -Anonimato

- É suficiente?
  - Qual o registro de Bob no dado abaixo?



Profissão	Idade	Cidade	Doença
Profissional	[25-35]	CE	Chikungunya
Médico	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	CE	AIDS
Programador	32	CE	AIDS
Profissional	[25-35]	CE	AIDS



Ataque de ligação ao **atributo**



Não associa-se o indivíduo ao seu registro, porém se descobre a informação sensível

# Modelo $l$ -Diversidade

- Atua como complemento do  $k$ -Anonimato
- Protege contra ataques de ligação ao **atributo**
- $k$ -Anonimato +  $l$ -Diversidade
  - Proteção contra ataques de ligação ao **registro** e ligação ao **atributo**



# Modelo $l$ -Diversidade

Cada classe de equivalência deve possuir pelo menos  $l$  valores distintos para o atributo sensível.

# Exemplo

■  $l = 2$

Profissão	Idade	Cidade	Doença
Profissional	[25-35]	CE	Chikungunya
Médico	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	CE	AIDS
Programador	32	CE	AIDS
Profissional	[25-35]	CE	AIDS

✓ 2-Anônimo

# Exemplo

■  $l = 2$

Profissão	Idade	Cidade	Doença
Profissional	[25-35]	CE	Chikungunya
Médico	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	CE	AIDS
Programador	32	CE	AIDS
Profissional	[25-35]	CE	AIDS

✓ 2-Diverso

# Exemplo

■  $l = 2$

Profissão	Idade	Cidade	Doença
Profissional	[25-35]	CE	Chikungunya
Médico	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	CE	AIDS
Programador	32	CE	AIDS
Profissional	[25-35]	CE	AIDS

✓ 2-Diverso

# Exemplo

■  $l = 2$

Profissão	Idade	Cidade	Doença
Profissional	[25-35]	CE	Chikungunya
Médico	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	CE	AIDS
Programador	32	CE	AIDS
Profissional	[25-35]	CE	AIDS



Problema

# Exemplo

■  $l = 2$

Profissão	Idade	Cidade	Doença
Profissional	[25-35]	CE	Chikungunya
Médico	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	CE	AIDS
Programador	32	CE	<b>Gripe</b>
Profissional	[25-35]	CE	AIDS

✓ 2-Diverso

# Exemplo

- Anonimizar a seguinte tabela para atender 2-Anonimato e 2-Diversidade

<i>semi-identificadores</i>			<i>sensível</i>
Situação	Idade	Gênero	Valor
Aprovado	34	M	550
Aprovado	31	M	2800
Aprovado	45	F	8000
Negado	42	M	8000
Aprovado	18	F	350
Negado	20	F	1000

# Exemplo

- Anonimizar a seguinte tabela para atender 2-Anonimato e 2-Diversidade

Situação	Idade	Gênero	Valor
Aprovado	34	M	550
Aprovado	31	M	2800
Aprovado	45	F	8000
Negado	42	M	8000
Aprovado	18	F	350
Negado	20	F	1000



# Exemplo

- Anonimizar a seguinte tabela para atender 2-Anonimato e 2-Diversidade

Situação	Idade	Gênero	Valor
Aprovado	34	M	550
Aprovado	31	M	2800
Aprovado	45	F	8000
Negado	42	M	8000
*	[15-20]	F	350
*	[15-20]	F	1000

# Exemplo

- Anonimizar a seguinte tabela para atender 2-Anonimato e 2-Diversidade

Situação	Idade	Gênero	Valor
Aprovado	34	M	550
Aprovado	31	M	2800
Aprovado	45	F	8000
Negado	42	M	8000
*	[15-20]	F	350
*	[15-20]	F	1000

# Exemplo

- Anonimizar a seguinte tabela para atender 2-Anonimato e 2-Diversidade

Situação	Idade	Gênero	Valor
Aprovado	[31-35]	M	550
Aprovado	[31-35]	M	2800
Aprovado	45	F	8000
Negado	42	M	8000
*	[15-20]	F	350
*	[15-20]	F	1000

# Exemplo

- Anonimizar a seguinte tabela para atender 2-Anonimato e 2-Diversidade

Situação	Idade	Gênero	Valor
Aprovado	[31-35]	M	550
Aprovado	[31-35]	M	2800
Aprovado	45	F	8000
Negado	42	M	8000
*	[15-20]	F	350
*	[15-20]	F	1000

# Exemplo

- Anonimizar a seguinte tabela para atender 2-Anonimato e 2-Diversidade

Situação	Idade	Gênero	Valor
Aprovado	[31-35]	M	550
Aprovado	[31-35]	M	2800
*	[41-45]	*	8000
*	[41-45]	*	8000
*	[15-20]	F	350
*	[15-20]	F	1000

# Exemplo

- Anonimizar a seguinte tabela para atender 2-Anonimato e 2-Diversidade

Situação	Idade	Gênero	Valor
Aprovado	[31-35]	M	550
Aprovado	[31-35]	M	2800
*	[41-45]	*	8000
*	[41-45]	*	8000
*	[15-20]	F	350
*	[15-20]	F	1000

✓ 2-Diverso

# Exemplo

- Anonimizar a seguinte tabela para atender 2-Anonimato e 2-Diversidade

Situação	Idade	Gênero	Valor
Aprovado	[31-35]	M	550
Aprovado	[31-35]	M	2800
*	[41-45]	*	8000
*	[41-45]	*	8000
*	[15-20]	F	350
*	[15-20]	F	1000

✓ 2-Diverso

# Exemplo

- Anonimizar a seguinte tabela para atender 2-Anonimato e 2-Diversidade

Situação	Idade	Gênero	Valor
Aprovado	[31-35]	M	550
Aprovado	[31-35]	M	2800
*	[41-45]	*	8000
*	[41-45]	*	8000
*	[15-20]	F	350
*	[15-20]	F	1000



Problema



# Exemplo

- Anonimizar a seguinte tabela para atender 2-Anonimato e 2-Diversidade

Situação	Idade	Gênero	Valor
Aprovado	[31-35]	M	550
Aprovado	[31-35]	M	2800
*	[41-45]	*	8000
*	[41-45]	*	<b>5000</b>
*	[15-20]	F	350
*	[15-20]	F	1000

✓ 2-Diverso

## Para Pensar...

Situação	Idade	Gênero	Valor
Aprovado	[31-35]	M	550
Aprovado	[31-35]	M	2800
*	[41-45]	*	8000
*	[41-45]	*	<b>7900</b>
*	[15-20]	F	350
*	[15-20]	F	1000



Dessa forma estaria diversificado?

# Exemplo

- Anonimizar a seguinte tabela para atender 3-Anonimato e 2-Diversidade

<i>semi-identificadores</i>			<i>sensível</i>
<b>Gênero</b>	<b>Idade</b>	<b>CEP</b>	<b>Doença</b>
M	27	60999-430	Gripe
M	29	60999-441	Gripe
M	28	60999-501	Gripe
F	33	64888-331	Gripe
F	34	64888-350	Dengue
F	34	64888-351	Gripe
F	26	68777-610	Zika
F	29	68777-611	Zika
F	27	68777-611	Gripe

# Exemplo

- Anonimizar a seguinte tabela para atender 3-Anonimato e 2-Diversidade

Gênero	Idade	CEP	Doença
M	27	60999-430	Gripe
M	29	60999-441	Gripe
M	28	60999-501	Gripe
F	33	64888-331	Gripe
F	34	64888-350	Dengue
F	34	64888-351	Gripe
F	26	68777-610	Zika
F	29	68777-611	Zika
F	27	68777-611	Gripe

# Exemplo

- Anonimizar a seguinte tabela para atender 3-Anonimato e 2-Diversidade

Gênero	Idade	CEP	Doença
M	[26-30]	60999-430	Gripe
M	[26-30]	60999-441	Gripe
M	[26-30]	60999-501	Gripe
F	33	64888-331	Gripe
F	34	64888-350	Dengue
F	34	64888-351	Gripe
F	26	68777-610	Zika
F	29	68777-611	Zika
F	27	68777-611	Gripe

# Exemplo

- Anonimizar a seguinte tabela para atender 3-Anonimato e 2-Diversidade

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
F	33	64888-331	Gripe
F	34	64888-350	Dengue
F	34	64888-351	Gripe
F	26	68777-610	Zika
F	29	68777-611	Zika
F	27	68777-611	Gripe

# Exemplo

- Anonimizar a seguinte tabela para atender 3-Anonimato e 2-Diversidade

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
F	33	64888-331	Gripe
F	34	64888-350	Dengue
F	34	64888-351	Gripe
F	26	68777-610	Zika
F	29	68777-611	Zika
F	27	68777-611	Gripe

# Exemplo

- Anonimizar a seguinte tabela para atender 3-Anonimato e 2-Diversidade

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
F	[31-35]	64888-331	Gripe
F	[31-35]	64888-350	Dengue
F	[31-35]	64888-351	Gripe
F	26	68777-610	Zika
F	29	68777-611	Zika
F	27	68777-611	Gripe



# Exemplo

- Anonimizar a seguinte tabela para atender 3-Anonimato e 2-Diversidade

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
F	[31-35]	64888-*	Gripe
F	[31-35]	64888-*	Dengue
F	[31-35]	64888-*	Gripe
F	26	68777-610	Zika
F	29	68777-611	Zika
F	27	68777-611	Gripe

# Exemplo

- Anonimizar a seguinte tabela para atender 3-Anonimato e 2-Diversidade

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
F	[31-35]	64888-*	Gripe
F	[31-35]	64888-*	Dengue
F	[31-35]	64888-*	Gripe
F	26	68777-610	Zika
F	29	68777-611	Zika
F	27	68777-611	Gripe

# Exemplo

- Anonimizar a seguinte tabela para atender 3-Anonimato e 2-Diversidade

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
F	[31-35]	64888-*	Gripe
F	[31-35]	64888-*	Dengue
F	[31-35]	64888-*	Gripe
F	[26-30]	68777-610	Zika
F	[26-30]	68777-611	Zika
F	[26-30]	68777-611	Gripe

# Exemplo

- Anonimizar a seguinte tabela para atender 3-Anonimato e 2-Diversidade

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
F	[31-35]	64888-*	Gripe
F	[31-35]	64888-*	Dengue
F	[31-35]	64888-*	Gripe
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Gripe

# Exemplo

- Anonimizar a seguinte tabela para atender 3-Anonimato e 2-Diversidade

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
F	[31-35]	64888-*	Gripe
F	[31-35]	64888-*	Dengue
F	[31-35]	64888-*	Gripe
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Gripe



Problema

# Exemplo

- Anonimizar a seguinte tabela para atender 3-Anonimato e 2-Diversidade

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	<b>Gastrite</b>
F	[31-35]	64888-*	Gripe
F	[31-35]	64888-*	Dengue
F	[31-35]	64888-*	Gripe
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Gripe

# Exemplo

- Anonimizar a seguinte tabela para atender 3-Anonimato e 2-Diversidade

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gastrite
F	[31-35]	64888-*	Gripe
F	[31-35]	64888-*	Dengue
F	[31-35]	64888-*	Gripe
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Gripe

# Exemplo

- Anonimizar a seguinte tabela para atender 3-Anonimato e 2-Diversidade

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gastrite
F	[31-35]	64888-*	Gripe
F	[31-35]	64888-*	Dengue
F	[31-35]	64888-*	Gripe
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Gripe



# Exemplo

- Anonimizar a seguinte tabela para atender 3-Anonimato e 2-Diversidade

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gastrite
F	[31-35]	64888-*	Gripe
F	[31-35]	64888-*	Dengue
F	[31-35]	64888-*	Gripe
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Gripe

✓ 3-Anônimo e 2-Diverso

# Skewness Attack

- Atacante sabe que uma pessoa **X** está doente
- Qual a chance de **X** ter **Zika**?

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gastrite
F	[31-35]	64888-*	Gripe
F	[31-35]	64888-*	Dengue
F	[31-35]	64888-*	Gripe
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Gripe

3-Anônimo e 2-Diverso

# Skewness Attack

- Atacante sabe que uma pessoa **X** está doente
- Qual a chance de **X** ter **Zika**?

$$\frac{2}{9} \approx 22\%$$

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gastrite
F	[31-35]	64888-*	Gripe
F	[31-35]	64888-*	Dengue
F	[31-35]	64888-*	Gripe
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Gripe

3-Anônimo e 2-Diverso

# Skewness Attack

- Atacante sabe que uma **mulher X** está doente
- Qual a chance de **X** ter **Zika**?

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gastrite
F	[31-35]	64888-*	Gripe
F	[31-35]	64888-*	Dengue
F	[31-35]	64888-*	Gripe
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Gripe

3-Anônimo e 2-Diverso

# Skewness Attack

- Atacante sabe que uma **mulher X** está doente
- Qual a chance de **X** ter **Zika**?

$$\frac{2}{6} \approx 33\%$$

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gastrite
F	[31-35]	64888-*	Gripe
F	[31-35]	64888-*	Dengue
F	[31-35]	64888-*	Gripe
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Gripe

3-Anônimo e 2-Diverso

# Skewness Attack

- Atacante sabe que uma **mulher X** está doente e que ela mora em um bairro cujo CEP é **68777-000**
- Qual a chance de **X** ter **Zika**?

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gastrite
F	[31-35]	64888-*	Gripe
F	[31-35]	64888-*	Dengue
F	[31-35]	64888-*	Gripe
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Gripe

3-Anônimo e 2-Diverso

# Skewness Attack

- Atacante sabe que uma **mulher X** está doente e que ela mora em um bairro cujo CEP é **68777-000**
- Qual a chance de **X** ter **Zika**?

$$\frac{2}{3} \approx 66\%$$

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gastrite
F	[31-35]	64888-*	Gripe
F	[31-35]	64888-*	Dengue
F	[31-35]	64888-*	Gripe
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Zika
F	[26-30]	68777-*	Gripe

3-Anônimo e 2-Diverso

# Similarity Attack

- Atacante sabe que uma **mulher X** está doente e que ela mora em um bairro cujo CEP é **68777-000**
- O que o atacante pode inferir sobre **X**?

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gastrite
F	[31-35]	64888-*	Gripe
F	[31-35]	64888-*	Dengue
F	[31-35]	64888-*	Gripe
F	[26-30]	68777-*	Rosácea
F	[26-30]	68777-*	Acne
F	[26-30]	68777-*	Vitiligo

3-Anônimo e 2-Diverso



# Similarity Attack

- Atacante sabe que uma **mulher X** está doente e que ela mora em um bairro cujo CEP é **68777-000**
- O que o atacante pode inferir sobre **X**?

**X** tem doença de pele

Gênero	Idade	CEP	Doença
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gripe
M	[26-30]	60999-*	Gastrite
F	[31-35]	64888-*	Gripe
F	[31-35]	64888-*	Dengue
F	[31-35]	64888-*	Gripe
F	[26-30]	68777-*	Rosácea
F	[26-30]	68777-*	Acne
F	[26-30]	68777-*	Vitiligo

3-Anônimo e 2-Diverso

# Outros Problemas

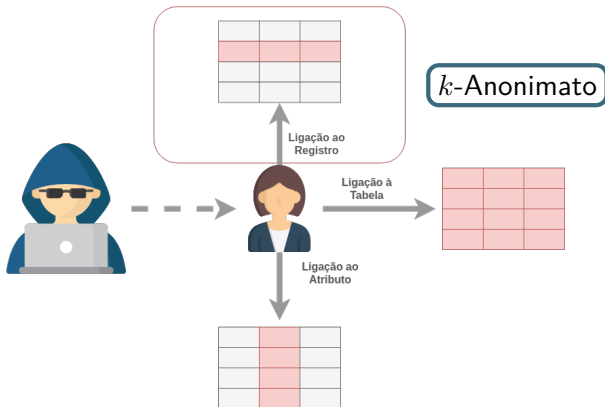
- Dados com muita repetição de um valor de atributo sensível e pouca dos demais
  - Grande distorção
  - Baixa utilidade
- Problema de semântica dos valores

# $\delta$ -Presença

# Revisão

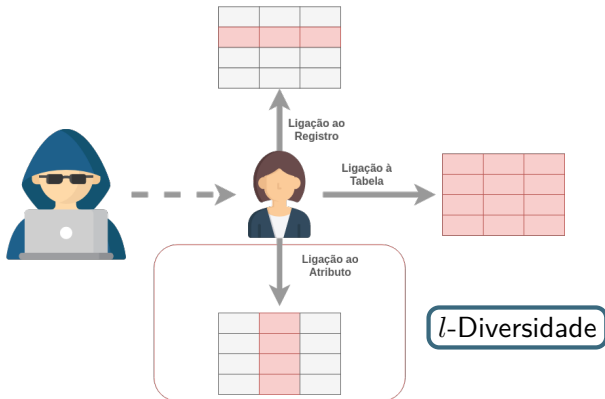
# Tipos de Ataques

## ■ Ataques de ligação ao registro



# Tipos de Ataques

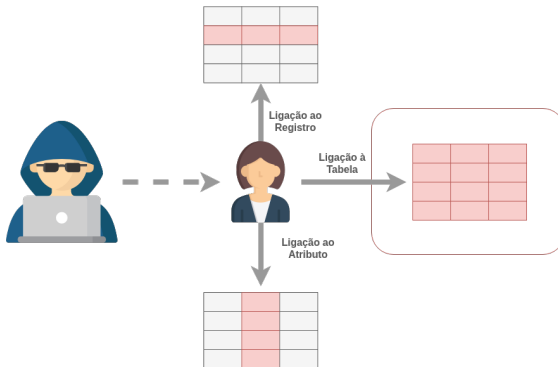
## ■ Ataques de ligação ao atributo



# Motivação

# Tipos de Ataques

## ■ Ataques de ligação à tabela





# Ligação à Tabela

- Dado o censo do estado do Ceará (CE) e os dados publicados de um hospital de Fortaleza
  - Quais ameaças de privacidade encontramos?

Dado anonimizado - Hospital

Profissão	Idade	Cidade	Doença
Engenheiro	[29-33]	Fortaleza	Zika
Médico	[29-33]	Fortaleza	Zika
Médico	[29-33]	Fortaleza	Hepatite C
Programador	[29-33]	Fortaleza	AIDS

Dado externo - Censo

Nome	Profissão	Idade	Estado
Davi	Programador	29	CE
Pedro	Médico	32	CE
Bob	Programador	32	CE
Maria	Dentista	35	CE
Caio	Engenheiro	29	CE
Alan	Médico	33	CE

# Ligação à Tabela

- Identificar que o indivíduo está no conjunto de dados é uma ameaça?

Dado anonimizado - Hospital

Profissão	Idade	Cidade	Doença
Engenheiro	[29-33]	Fortaleza	Zika
Médico	[29-33]	Fortaleza	Zika
Médico	[29-33]	Fortaleza	Hepatite C
Programador	[29-33]	Fortaleza	AIDS

Dado externo - Censo

Nome	Profissão	Idade	Estado
Davi	Programador	29	CE
Pedro	Médico	32	CE
Bob	Programador	32	CE
Maria	Dentista	35	CE
Caio	Engenheiro	29	CE
Alan	Médico	33	CE

# Ligação à Tabela

- **Bob** está no conjunto de dados publicado?

Dado anonimizado - Hospital

Profissão	Idade	Cidade	Doença
Engenheiro	[29-33]	Fortaleza	Zika
Médico	[29-33]	Fortaleza	Zika
Médico	[29-33]	Fortaleza	Hepatite C
Programador	[29-33]	Fortaleza	AIDS

Dado externo - Censo

Nome	Profissão	Idade	Estado
Davi	Programador	29	CE
Pedro	Médico	32	CE
Bob	Programador	32	CE
Maria	Dentista	35	CE
Caio	Engenheiro	29	CE
Alan	Médico	33	CE

# Ligação à Tabela

- **Bob** está no conjunto de dados publicado?

Dado anonimizado - Hospital

Profissão	Idade	Cidade	Doença
Engenheiro	[29-33]	Fortaleza	Zika
Médico	[29-33]	Fortaleza	Zika
Médico	[29-33]	Fortaleza	Hepatite C
Programador	[29-33]	Fortaleza	AIDS

Dado externo - Censo

Nome	Profissão	Idade	Estado
Davi	Programador	29	CE
Pedro	Médico	32	CE
Bob	Programador	32	CE
Maria	Dentista	35	CE
Caio	Engenheiro	29	CE
Alan	Médico	33	CE

# Ligação à Tabela

- **Bob** está no conjunto de dados publicado?

Dado anonimizado - Hospital

Profissão	Idade	Cidade	Doença
Engenheiro	[29-33]	Fortaleza	Zika
Médico	[29-33]	Fortaleza	Zika
Médico	[29-33]	Fortaleza	Hepatite C
Programador	[29-33]	Fortaleza	AIDS

Dado externo - Censo

Nome	Profissão	Idade	Estado
Davi	Programador	29	CE
Pedro	Médico	32	CE
Bob	Programador	32	CE
Maria	Dentista	35	CE
Caio	Engenheiro	29	CE
Alan	Médico	33	CE

# Ligação à Tabela

- **Bob** está no conjunto de dados publicado?

Dado anonimizado - Hospital

Profissão	Idade	Cidade	Doença
Engenheiro	[29-33]	Fortaleza	Zika
Médico	[29-33]	Fortaleza	Zika
Médico	[29-33]	Fortaleza	Hepatite C
Programador	[29-33]	Fortaleza	AIDS

Dado externo - Censo

Nome	Profissão	Idade	Estado
Davi	Programador	29	CE
Pedro	Médico	32	CE
Bob	Programador	32	CE
Maria	Dentista	35	CE
Caio	Engenheiro	29	CE
Alan	Médico	33	CE

# Ligação à Tabela

- **Bob** está no conjunto de dados publicado?

Dado anonimizado - Hospital

Profissão	Idade	Cidade	Doença
Engenheiro	[29-33]	Fortaleza	Zika
Médico	[29-33]	Fortaleza	Zika
Médico	[29-33]	Fortaleza	Hepatite C
Programador	[29-33]	Fortaleza	AIDS

Dado externo - Censo

Nome	Profissão	Idade	Estado
Davi	Programador	29	CE
Pedro	Médico	32	CE
Bob	Programador	32	CE
Maria	Dentista	35	CE
Caio	Engenheiro	29	CE
Alan	Médico	33	CE

50% de chance!

# Ligação à Tabela

- **Bob** está no conjunto de dados publicado?

Dado anonimizado - Hospital

Profissão	Idade	Cidade	Doença
Engenheiro	[29-33]	Fortaleza	Zika
Médico	[29-33]	Fortaleza	Zika
Médico	[29-33]	Fortaleza	Hepatite C
Programador	[29-33]	Fortaleza	AIDS

Dado externo - Censo

Nome	Profissão	Idade	Estado
Davi	Programador	29	CE
Pedro	Médico	32	CE
Bob	Programador	32	CE
Maria	Dentista	35	CE
Caio	Engenheiro	29	CE
Alan	Médico	33	CE

50% de chance de **Bob** ter AIDS!



# Modelo $\delta$ -Presença

- Define  $\delta = (\delta_{min}, \delta_{max})$ 
  - Limites de probabilidade de inferir a presença de um indivíduo na tabela

# Exemplo

## Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
	47903	59	Canadá
	47906	42	EUA
	47633	63	Peru
	48972	47	Bulgária
	48970	52	França

# Exemplo

## Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
	47903	59	Canadá
	47906	42	EUA
	47633	63	Peru
	48972	47	Bulgária
	48970	52	França

# Exemplo

## Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
	47*	59	Canadá
	47*	42	EUA
	47*	63	Peru
	48972	47	Bulgária
	48970	52	França

# Exemplo

## Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
	47*	*	Canadá
	47*	*	EUA
	47*	*	Peru
	48972	47	Bulgária
	48970	52	França

# Exemplo

## Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
	47*	*	América
	47*	*	América
	47*	*	América
	48972	47	Bulgária
	48970	52	França

# Exemplo

## Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
	47*	*	América
	47*	*	América
	47*	*	América
	48972	47	Bulgária
	48970	52	França

# Exemplo

## Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
47*	*	*	América
47*	*	*	América
47*	*	*	América
48*	47	47	Bulgária
48*	52	52	França



# Exemplo

## Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
47*	*	*	América
47*	*	*	América
47*	*	*	América
48*	*	*	Bulgária
48*	*	*	França

# Exemplo

Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
	47*	*	América
	47*	*	América
	47*	*	América
	48*	*	Europa
	48*	*	Europa

$\subseteq$

Dado Público (E)

ID	Nome	CEP	Idade	País
a	Alice	47906	35	EUA
b	Bob	47903	59	Canadá
c	Christine	47906	42	EUA
d	Dirk	47630	18	Brasil
e	Eunice	47630	22	Brasil
f	Frank	47633	63	Peru
g	Gail	48973	33	Espanha
h	Harry	48972	47	Bulgária
i	Iris	48970	52	França

# Exemplo

Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
b	47*	*	América
c	47*	*	América
f	47*	*	América
h	48*	*	Europa
i	48*	*	Europa

$\subseteq$

Dado Público (E)

ID	Nome	CEP	Idade	País
a	Alice	47906	35	EUA
b	Bob	47903	59	Canadá
c	Christine	47906	42	EUA
d	Dirk	47630	18	Brasil
e	Eunice	47630	22	Brasil
f	Frank	47633	63	Peru
g	Gail	48973	33	Espanha
h	Harry	48972	47	Bulgária
i	Iris	48970	52	França

# Exemplo

Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
b	47*	*	América
c	47*	*	América
f	47*	*	América
h	48*	*	Europa
i	48*	*	Europa

$\subseteq$

Dado Público (E)

ID	Nome	CEP	Idade	País
a	Alice	47906	35	EUA
b	Bob	47903	59	Canadá
c	Christine	47906	42	EUA
d	Dirk	47630	18	Brasil
e	Eunice	47630	22	Brasil
f	Frank	47633	63	Peru
g	Gail	48973	33	Espanha
h	Harry	48972	47	Bulgária
i	Iris	48970	52	França

# Exemplo

Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
b	47*	*	América
c	47*	*	América
f	47*	*	América
h	48*	*	Europa
i	48*	*	Europa

$\subseteq$

Dado Público (E)

ID	Nome	CEP	Idade	País
a	Alice	47906	35	EUA
b	Bob	47903	59	Canadá
c	Christine	47906	42	EUA
d	Dirk	47630	18	Brasil
e	Eunice	47630	22	Brasil
f	Frank	47633	63	Peru
g	Gail	48973	33	Espanha
h	Harry	48972	47	Bulgária
i	Iris	48970	52	França

# Exemplo

Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
b	47*	*	América
c	47*	*	América
f	47*	*	América
h	48*	*	Europa
i	48*	*	Europa

$\subseteq$

Dado Público (E)

ID	Nome	CEP	Idade	País
a	Alice	47906	35	EUA
b	Bob	47903	59	Canadá
c	Christine	47906	42	EUA
d	Dirk	47630	18	Brasil
e	Eunice	47630	22	Brasil
f	Frank	47633	63	Peru
g	Gail	48973	33	Espanha
h	Harry	48972	47	Bulgária
i	Iris	48970	52	França

# Exemplo

Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
b	47*	*	América
c	47*	*	América
f	47*	*	América
h	48*	*	Europa
i	48*	*	Europa

$\subseteq$

Dado Público (E)

ID	Nome	CEP	Idade	País
a	Alice	47906	35	EUA
b	Bob	47903	59	Canadá
c	Christine	47906	42	EUA
d	Dirk	47630	18	Brasil
e	Eunice	47630	22	Brasil
f	Frank	47633	63	Peru
g	Gail	48973	33	Espanha
h	Harry	48972	47	Bulgária
i	Iris	48970	52	França

Qual a probabilidade de **Alice** estar no dado anonimizado?

# Exemplo

Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
b	47*	*	América
c	47*	*	América
f	47*	*	América
h	48*	*	Europa
i	48*	*	Europa

$\subseteq$

Dado Público (E)

ID	Nome	CEP	Idade	País
a	Alice	47906	35	EUA
b	Bob	47903	59	Canadá
c	Christine	47906	42	EUA
d	Dirk	47630	18	Brasil
e	Eunice	47630	22	Brasil
f	Frank	47633	63	Peru
g	Gail	48973	33	Espanha
h	Harry	48972	47	Bulgária
i	Iris	48970	52	França

$$Prob(Alice \in T|E) = ?$$



# Exemplo

Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
b	47*	*	América
c	47*	*	América
f	47*	*	América
h	48*	*	Europa
i	48*	*	Europa

$\subseteq$

Dado Público (E)

ID	Nome	CEP	Idade	País
a	Alice	47906	35	EUA
b	Bob	47903	59	Canadá
c	Christine	47906	42	EUA
d	Dirk	47630	18	Brasil
e	Eunice	47630	22	Brasil
f	Frank	47633	63	Peru
g	Gail	48973	33	Espanha
h	Harry	48972	47	Bulgária
i	Iris	48970	52	França

$$Prob(Alice \in T|E) = \frac{3}{6} = 0.50 = 50\%$$

# Exemplo

Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
b	47*	*	América
c	47*	*	América
f	47*	*	América
h	48*	*	Europa
i	48*	*	Europa

$\subseteq$

Dado Público (E)

ID	Nome	CEP	Idade	País
a	Alice	47906	35	EUA
b	Bob	47903	59	Canadá
c	Christine	47906	42	EUA
d	Dirk	47630	18	Brasil
e	Eunice	47630	22	Brasil
f	Frank	47633	63	Peru
g	Gail	48973	33	Espanha
h	Harry	48972	47	Bulgária
i	Iris	48970	52	França

Qual a probabilidade de **Harry** estar no dado anonimizado?

# Exemplo

Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
b	47*	*	América
c	47*	*	América
f	47*	*	América
h	48*	*	Europa
i	48*	*	Europa

$\subseteq$

Dado Público (E)

ID	Nome	CEP	Idade	País
a	Alice	47906	35	EUA
b	Bob	47903	59	Canadá
c	Christine	47906	42	EUA
d	Dirk	47630	18	Brasil
e	Eunice	47630	22	Brasil
f	Frank	47633	63	Peru
g	Gail	48973	33	Espanha
h	Harry	48972	47	Bulgária
i	Iris	48970	52	França

$$Prob(Harry \in T|E) = ?$$

# Exemplo

Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
b	47*	*	América
c	47*	*	América
f	47*	*	América
h	48*	*	Europa
i	48*	*	Europa

$\subseteq$

Dado Público (E)

ID	Nome	CEP	Idade	País
a	Alice	47906	35	EUA
b	Bob	47903	59	Canadá
c	Christine	47906	42	EUA
d	Dirk	47630	18	Brasil
e	Eunice	47630	22	Brasil
f	Frank	47633	63	Peru
g	Gail	48973	33	Espanha
h	Harry	48972	47	Bulgária
i	Iris	48970	52	França

$$Prob(Harry \in T|E) = \frac{2}{3} \approx 0.66 \approx 66\%$$

# Exemplo

Subconjunto de Pesquisa (T)

ID	CEP	Idade	País
b	47*	*	América
c	47*	*	América
f	47*	*	América
h	48*	*	Europa
i	48*	*	Europa

$\subseteq$

Dado Público (E)

ID	Nome	CEP	Idade	País
a	Alice	47906	35	EUA
b	Bob	47903	59	Canadá
c	Christine	47906	42	EUA
d	Dirk	47630	18	Brasil
e	Eunice	47630	22	Brasil
f	Frank	47633	63	Peru
g	Gail	48973	33	Espanha
h	Harry	48972	47	Bulgária
i	Iris	48970	52	França

$$\left. \begin{array}{l} \delta_{min} = 0.50 \\ \delta_{max} = 0.66 \end{array} \right\} = (0.50, 0.66)\text{-Presença}$$

# Conclusão

- Indiretamente previne ataque de ligação ao **registro** e ao **atributo**
- Ponto fraco
  - Assume que o *data holder* tem acesso ao mesmo dado externo que o adversário