University Passau
Faculty of Computer Science and Mathematics

Bachelor Thesis

Substitution of Privacy Models for the Layered Privacy Language

Thesis Prepared for the Degree of
Bachelor of Science (B.Sc.)

at the Chair of Distributed Information Systems
of the Faculty of Computer Science and Mathematics
of the University Passau

Name:	Fabian Pfeil
Matriculation Number:	77560
Subject Area:	Computer Science
Course of studies:	Bachelor Internet Computing
Study Year:	2015
Examiner:	Prof. Dr. Kosch
Second Examiner:	Prof. Dr. Granitzer

Contents

List of Figures

List of Tables

Abstract

Zitiertest[13].
K-anonymity[19].
L-diversity[16].
T-closeness[14].
d-presence[17]
delta-disclosure[6]
basic-beta-likeness[7]
k-map1[11]
population uniqueness[8]
profitability 1[20]
profitability 2[18]
arx[1]
diff[10]
diff2[5]
diff3[15]

1 Introduction

1.1 Motivation

Figure 1.1: Describe this picture.

1.2 Layered Privacy Language

2 Privacy Model Classification

In this section, we will take a look at the Privacy Models implemented by Arx. Therefore, these Privacy Models will be explained in terms of requirements for the data sets to which they can be applied, the attacks they mitigate, their use cases, advantages and disadvantages.

2.1 Definitions

2.2 K-Anonymity

The first Privacy Model, we will take a look at is K-Anonymity. This Privacy Model was released in 2002 by Latanya Sweeney with the goal to prevent re-identification of data subjects of certain data sets[19]. K-Anonymity is defined as follows:

> "Let RT(A_1,...,A_n) be a table and QI_{RT} be the quasi-identifier associated with it. RT is said to satisfy k-anonymity if and only if each sequence of values in RT[QI_{RT}] appears with at least k occurrences in RT[QI_{RT}]."[19]

To clarify what this means, figure 2.1 shows the example used in Sweeney's work[19]. The figure shown here, is an example for a table that satisfies the k-anonymity criterion. The quasi-identifier for this particular case is QI_T= {Race, Birth, Gender, ZIP} and k = 2. This means that every tuple of quasi-identifying attributes appears at least in two records in T.
As k-Anonymity is defined, we will now take a look at the attacks, which can be mitigated with the property of k-anonymity applied. As this issue was already addressed by Fung et al., their classification will be used here. In the work of Fung et al. it is stated that k-anonymity, applied to a data set, will prevent only Record Linkage attacks. Consequently, other attacks like Attribute Linkage, Table Linkage or a Probabilistic Attack can not be mitigated by k-Anonymity[12].

	Race	Birth	Gender	ZIP	Problem
t1	Black	1965	m	0214*	short breath
t2	Black	1965	m	0214*	chest pain
t3	Black	1965	f	0213*	hypertension
t4	Black	1965	f	0213*	hypertension
t5	Black	1964	f	0213*	obesity
t6	Black	1964	f	0213*	chest pain
t7	White	1964	m	0213*	chest pain
t8	White	1964	m	0213*	obesity
t9	White	1964	m	0213*	short breath
t10	White	1967	m	0213*	chest pain
t11	White	1967	m	0213*	chest pain

Figure 2.1: Example of a 2-anonymous table [19]

2.2.1 K-Map

The Privacy Model of K-Map is directly related to k-Anonymity. El Emam et al. show that extending k-Anonymity to k-Map can however reduce the loss of information due to over-anonymization of the dataset[11]. Damien Desfontaines describes k-Map like this: "Your data satisfies k-map if every combination of values for the quasi-identifiers appears at least k times in the reidentification dataset."[9]
The similarity to k-Anonymity is obvious, as the main criterion in both Privacy Models is the same, the only difference is the data set on which they are based. In the case of k-Map it is not the data set of the data custodian, it is the reidentification dataset[9]. Consequently, K-Map mitigates the same attacks (Record Linkage Attacks) as k-Anonymity. Other attacks can not be prevented by this Privacy Model.
However, the work of El Emam et al. states, that even though k-Map reduces information loss in comparison to k-Anonymity, the model of k-Map is not used in practices because one can assume that a data custodian does not have access to a reidentification dataset, while an attacker does[11].

2.3 l-Diversity

The next Privacy Model we will take a look at is l-Diversity. L-Diversity was proposed by Machanavajjhala et al. to overcome the weaknesses of k-Anonymity, namely the attribute linkage attacks[16]. To understand why l-Diversity mitigates attribute linkage, we will take a short look at the definition. In the work of Machanavajjhala et al. the principle of l-Diverity is defined as follows:

> "A q^*-block is l-diverse if it contains at least l well-represented values for the sensitive attribute S. A table is l-diverse if every q^*-block is l-diverse." [16]

This means that every group of quasi-identifiers has to at least contain l different values for the sensitive attribute S. It is because of this property that l-diversity can prevent attribute linkage attacks. However, table linkage attacks can not be mitigated through l-diversity[12]. Machanavajjhala et al. use the table shown in figure 2.2 as an example for l-diversity. In this example, one can see that every group of quasi-identifiers contains at least three different values for the sensitive attribute.

	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Condition
1	1305*	≤ 40	*	Heart Disease
4	1305*	≤ 40	*	Viral Infection
9	1305*	≤ 40	*	Cancer
10	1305*	≤ 40	*	Cancer
5	1485*	> 40	*	Cancer
6	1485*	> 40	*	Heart Disease
7	1485*	> 40	*	Viral Infection
8	1485*	> 40	*	Viral Infection
2	1306*	≤ 40	*	Heart Disease
3	1306*	≤ 40	*	Viral Infection
11	1306*	≤ 40	*	Cancer
12	1306*	≤ 40	*	Cancer

Figure 2.2: Example of a 3-diverse table [16]

There are three instantiations of l-diversity implemented in Arx. Consequently, we will only take a look at these.

2.3.1 Distinct-l-Diversity

Distinct-l-Diversity is arguably the simplest of the three instantiations of l-diversity. It is also known as p-sensitive k-anonymity. Distinct-l-diversity uses the plain definition of l-diversity mentioned above. Therefore, no advanced calculation of the l-factor has to take place. One can simply take the number of different values of the sensitive attribute as the factor l[12].
Lets get back to our example in figure 2.2. With distinct-l-diversity the table shown here is, as simple as it sounds, 3-diverse because every group of quasi-identifiers has three different values. Consequently, this table is diverse.
With this form of l-diversity, probabilistic inference attacks cannot be mitigated, because of the fact that some values of sensitive attributes are more

frequent than others. Therefore, two stronger instantiations of l-diversity have been created[12].

2.3.2 Entropy-l-Diversity

The next instantiation of l-Diversity is Entropy-l-Diversity. It is defined as follows:

> "A table is Entropy l-Diverse if, for every q^*-block,
>
> $$-\sum_{s \in S} p(q^*, s')log(p(q^*, s')) \geq log(l)$$
>
> where
>
> $$p(q^*, s) = \frac{n(q^*, s)}{\sum_{s' \in S} n(q^*, s')}$$
>
> is the fraction of tuples in the q^*-block with sensitive attribute value equal to s."[16]

With this calculation of the l-factor for our example in figure 2.2, our table is actually 2.8-diverse. Consequently, every group of quasi-identifiers has at least 2.8 different values for the sensitive attribute. And as there can not be 2.8 different values there are in our case three existing values.

2.3.3 Recursive (c, l)-Diversity

The third and final instantiation of l-Diversity we will take a look at is the recursive (c, l)-Diversity. This privacy model makes sure, that the most frequent values of sensitive attributes do not appear to often in the table. Furthermore it makes the uncommon values appear not too rarely[12]. Machanavajjhala et al. define recursive (c, l)-Diversity like this:

> "In a given q^*-block, let r_i denote the number of times the i^{th} most-frequent sensitive value appears in that q^*-block. Given a constant c, the q^*-block satisfies recursive (c, l)-diversity if r1 > c(r_l + r_{l+1} + ... + r_m). A table T^* satisfies recursive (c, l)-diversity, if every q^*-block satisfies recursive l-diversity. We say that 1-diversity is always satisfied."[16]

Both of these definitions for l-diversity, entropy and recursive, may be too restrictive. For entropy-l-diversity this can bee seen in the fact that, if entropy l-diversity should be applied, the entropy of whole table has to be log(l) or higher. This may be very difficult to achieve when, for example, a value for a

sensitive attribute is too common.
The same is the case for recursive (c, l)-Diversity. When for example a sensitive attribute value is present 90% of the time and the factor c is chosen < 9, it is impossible to achieve recursive (c, l)-Diversity[16].
This is not the only Problem with l-Diversity. Sensitive information can be leaked, despite the fact that l-diversity is applied, because l-diversity only guarantees the diversity of sensitive attribute values, but does not take the semantical relations of these values into account. Consequently, l-Diversity is insufficient to prevent attribute linkage in some cases where the distribution of a sensitive attribute is skewed[14].

2.4 t-Closeness

The next privacy model is t-Closeness. It was proposed by Li et al. to prevent attribute linkage through skewness attacks, mentioned earlier in section 2.3.3 [14]. In the work of Li et al. the property of t-Closeness is defined as follows:

> "An equivalence class is said to have t-closeness if the distance between the distribution of a sensitive attribute in this class and the distribution of the attribute in the whole table is no more than a threshold t. A table is said to have t-closeness if all equivalence classes have t-closeness."[14]

The factor t can be used to manage a trade off between privacy and utility. To measure the distance between those distributions and to compute the t-factor, Li et al. use the Earth Mover's Distance (EMD). In the work of Li et al, certain formulas were derived to calculate the EMD for the cases we will consider in the following subsections of t-closeness[14].

2.4.1 Ordered Distance t-Closeness

This type of t-closeness is used for numerical attributes, as these can be ordered. The distance between the distributions (in this case P and Q, with $r_i = p_i - q_i, (i = 1, 2, ...m)$) is calculated as follows [14]:

$$D[P,Q] = \frac{1}{m-1}(|r_1| + |r_1 + r_2| + ... + |r_1 + r_2 + ...r_{m-1}|) = \frac{1}{m-1}\sum_{i=1}^{i=m}\left|\sum_{j=1}^{j=i} r_j\right|$$

2.4.2 Equal Distance t-Closeness

Some types of attributes can not be orered like numerical values. Therefore, the next two computations of the EMD are used for categorical attributes, as they can not be ordered.
With the equal distance approach, all distances between any two values of categorical attributes are considered to be 1. Consequently, the following formula to compute the distance is the result[14]:

$$D[P,Q] = \frac{1}{2}\sum_{i=1}^{m}|p_i - q_i| = \sum_{p_i \geq q_i}(p_i - q_i) = -\sum_{p_i < q_i}(p_i - q_i)$$

2.4.3 Hierarchical Distance t-Closeness

Another way to compute the distance between distributions of categorical values is the hierarchical distance approach. This technique bases the distances of values on the minimum level to which these two values can be generalized to, based on a domain hierarchy[14].
Therefore, the so called extra of a leaf of this hierarchy is defined as:

$$extra(N) = \begin{cases} p_i - q_i & \text{if N is a leaf} \\ \sum_{C \in Child(N)} extra(C) & \text{otherwise} \end{cases}$$

where Child(N) is the portion of leaf nodes below N. Furthermore, two additional functions for internal Nodes are defined:

$$pos\_extra(N) = \sum_{C \in Child(N) \wedge extra(C) > 0} |extra(C)|$$

$$neg\_extra(N) = \sum_{C \in Child(N) \wedge extra(C) < 0} |extra(C)|$$

This leads to the cost-function, which describes the cost of moving between some leaf's (N) children branches.

$$cost(N) = \frac{height(N)}{H} min(pos\_extra(N), neg\_extra(N))$$

Finally, the EMD can be seen as follows:

$$D[P,Q] = \sum_{N} cost(N)$$

Now that we have defined the special cases for t-closeness, we will take a look at a short example used by Li et al. in their work[14].

	ZIP Code	Age	Salary	Disease
1	4767*	≤ 40	3K	gastric ulcer
3	4767*	≤ 40	5K	stomach cancer
8	4767*	≤ 40	9K	pneumonia
4	4790*	≥ 40	6K	gastritis
5	4790*	≥ 40	11K	flu
6	4790*	≥ 40	8K	bronchitis
2	4760*	≤ 40	4K	gastritis
7	4760*	≤ 40	7K	bronchitis
9	4760*	≤ 40	10K	stomach cancer

Figure 2.3: Table that has 0.167-closeness w.r.t. Salary and 0.278-closeness w.r.t. Disease [14]

Figure 2.3, shows a Table which has 0.167-closeness in relation to the sensitive attribute Salary, which can be calculated using the ordered distance formula for numerical values, and 0.278-closeness in relation to the attribute Disease. This can be easily computed using the given domain hierarchy shown in figure 2.4.

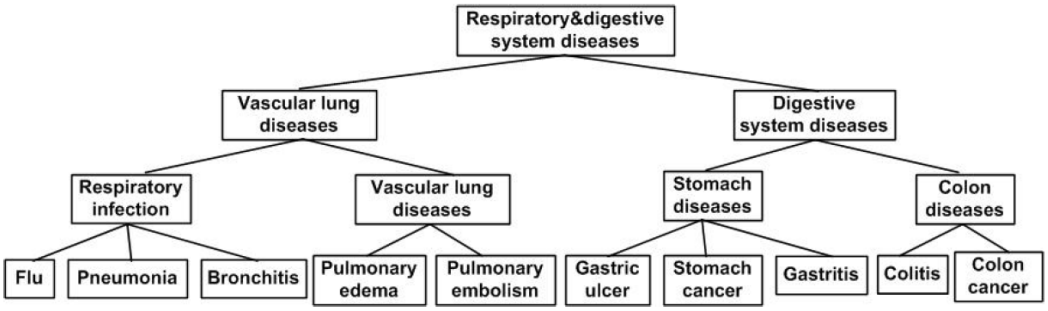


Figure 2.4: Hierarchy for categorical attributes Disease [14]

2.5 delta-Disclosure privacy

The next privacy model is delta-disclosure privacy. It also can be used to protect data sets from attribute linkage attacks, due to the fact that it is related to the t-closeness model. It also enforces restrictions on the distances between the sensitive attribute distributions. In contrast to the model proposed by Li et al., delta-disclosure privacy uses a multiplicative approach, which makes it stricter than t-closeness[6].
Delta-disclosure privacy was proposed by Brickell et al. and is defined as follows:

"We say that an equivalence class $\langle t \rangle$ is δ-disclosure-private with regard to the sensitive attribute S if, for all s $\in$ S

$$A_{quot}(\langle t \rangle) = \left| log \frac{p(\langle t \rangle, s)}{p(T, s)} \right| < \delta$$

A table T is δ-disclosure-private if for every t $\in E_Q$, $\langle t \rangle$ is δ-disclosure private."[6]

In easier words, one can say that a table T is delta-disclosure private if the distributions of sensitive attributes in the equivalence classes and in the overall table are approximately the same[6].

2.6 beta-Likeness

The privacy model beta-Likeness tries to overcome the limitations of the previous two models, as the EMD in t-closeness does not provide a clear privacy guarantee and the delta-disclosure privacy requires that every value of a sensitive attribute of the overall table occurs in every equivalence class. This makes delta-disclosure privacy unnecessarily strict.
To avoid these disadvantages, Cao et al. propose a privacy model called beta-Likeness. This model assumes that the distribution of the sensitive attributes in a table are known to the public and bases it's privacy constraint on information gain, which is denoted as a difference function between the distribution of a sensitive attribute in the overall table and in an equivalence class[7].
In the work of Cao et al., two instantiations of beta-Likeness are introduced.

2.6.1 basic beta-Likeness

The base approach for beta-likeness is defined as follows:

"Given table DB with sensitive attribute SA, let V = $\{v_1$, ... , $v_m\}$ be the SA domain, and P = (p_1, ... , p_m) the overall SA distribution in DB. An EC G with SA distribution Q = (q_1, . . . , q_m) is said to satisfy basic β-likeness, if and only if $max\{D(p_i, q_i) | p_i \in P, p_i < q_i\} \leq \beta$, where $\beta > 0$ is a threshold."[7]

This means, that every distance between the distributions of sensitive attributes in the overall table and in the equivalence class has to be lower or equal to a certain threshold β.

Furthermore, Cao et al. state that every equivalence class of an anonymized table satisfies beta-likeness, the whole table obeys beta-likeness.
The distance function for this privacy model is defined as $D(p_i, q_i) = \frac{q_i - p_i}{p_i}$ as Cao et al. opt for a relative difference instead of a absolute difference, because it does not suite their purposes. This relative distance function pays attention to less frequent values of sensitive attributes. Consequently, sensitive attribute values with a large frequency were not put into consideration. To mitigate the privacy threat caused by this, Cao et al. provide a stronger definition of beta-Likeness[7].

2.6.2 Enhanced beta-Likeness

This instantiation of beta-Likeness is stronger than basic beta-Likeness, as the name suggests. Enhanced beta-Likeness is defined as follows:

> "For table DB with sensitive attribute SA, let V = $\{v_1$, ... , $v_m\}$ be the SA domain, and P = (p_1, ... , p_m) the overall SA distribution in DB. An EC G with SA distribution Q = (q_1, ... , q_m) is said to satisfy enhanced β-likeness, if and only if $\forall q_i, D(p_i, q_i) = \frac{q_i - p_i}{p_i} \leq min\{\beta, -\ln p_i\}$, where $\beta > 0$ is a threshold and ln p_i is the natural logarithm of p_i."[7]

The properties that come with this definition protects the privacy for all sensitive attribute values. Values with rare occurrence receive sufficient attention, while values that occur more often can not approach frequency values of 1. As it is more robust than basic beta-likeness in terms of privacy, it should be used instead of it's predecessor[7].
Due to these traits and the fact that it is related to t-closeness and delta-disclosure privacy, the privacy model of beta-likeness can protect data sets against attribute linkage attacks and probabilistic attacks.

2.7 delta-Presence

The next privacy model that we take into consideration is delta-Presence. This model was proposed by Nergiz et al. to prevent attacks from identifying that a certain suspect is part of a dataset (Table Linkage), as this can pose a serious privacy threat in certain cases[17].

The privacy model of delta-presence is defined as follows:

> "Given an external public table P, and a private table T , we say that δ-presence holds for a generalization T* of T, with $\delta = (\delta_{min}, \delta_{max})$ if
>
> $$\delta_{min} \leq P(t \in T|T*) \leq \delta_{max} \quad \forall t \in P"[17]$$

In a dataset which applies this privacy criterion, every tuple t is called δ-present within the range of $\delta = (\delta_{min}, \delta_{max})$. To clarify what this means, we will take a short look at the example used in the work of Nergiz et al.
In Figure 2.5 the tables $P*_3$ and $T*_3$ are given. To get the probabilities $\delta_{min}, \delta_{max}$, the following calculations are done. $P(a \in T|T*_3) = \frac{|b,c,f|}{|a,b,c,d,e,f|} = \frac{1}{2}$. The same is done for tuples b,c,d,e and f. The probability for the tuples g,h and i is calculated with $\frac{|h,i|}{|g,h,i|} = \frac{2}{3}$. Consequently, the probability that a tuple from Table $P*_3$ is also in $T*_3$ lies between $\frac{1}{2}$ and $\frac{2}{3}$.
Like in most privacy models the δ-factor can be used as a trade-off between privacy and utility. Therefore this factor has to be chosen carefully.

P_3^*

	Public Dataset			Sen.
	Zip	Age	Nationality	
a	47*	*	America	0
b	47*	*	America	1
c	47*	*	America	1
d	47*	*	America	0
e	47*	*	America	0
f	47*	*	America	1
g	48*	*	Europe	0
h	48*	*	Europe	1
i	48*	*	Europe	1

T_3^*

	Research Subset		
	Zip	Age	Nationality
b	47*	*	America
c	47*	*	America
f	47*	*	America
h	48*	*	Europe
i	48*	*	Europe

Figure 2.5: Table $T*_3$ that is $(\frac{1}{2}, \frac{2}{3})$-present [17]

2.7.1 Inclusion

This privacy model is only mentioned in the java code of Arx. It is stated that this privacy model is an extension of delta-Presence, but does not enforce any privacy guarantees. It allows the user to specify a data subset[3].

2.8 Profitability

Now we will take a look at a group of privacy models that are part of the so called game theoretic approach, which was proposed by Wan et al in 2015.

This approach tries to find the best possible de-identification strategy to maximize the data publishers monetary gain[20].
In the implementation of arx, four privacy models are present which represent two scenarios. In the *Prosecutor* model, we assume that the attacker knows that a certain record is present in our data base. Consequently groups have to be formed. This is realized by the k-anonymity privacy model, mentioned in section 2.2.
In the second scenario, the *Journalist* model, the assumption that a attacker knows about an individual's membership is not made. Therefore groups can be made by generalizing the population table. This corresponds to the k-map privacy model mentioned in section 2.2.1[18].
For every scenario mentioned here, two variants are implemented by arx. The first one is the SH-Friendly variant which uses generalization hierarchies and appropriate levels of minimal generalization to always satisfy the Safe Harbor policy of HIPAA. The second one is the No-Attack variant. This model guarantees that the adversary never has any urge to even try to attack the published data. Therefore an attacker will only attack, if his monetary gain for attacking a record is greater than the costs he has to encounter. This can be described as:

$$SP(r) \cdot gain > cost$$

Consequently the cost of the attack always has to be greater than the attackers estimated monetary gain.[18]
It is because of the use of k-anonymity and k-map, that the privacy models of Profitability can be used to protect data from Record Linkage while also maximizing the users monetary gain.
The specifications above result in four privacy models:

- ProfitabilityProsecutor
- ProfitabilityJournalist
- ProfitabilityProsecutor No Attack
- ProfitabilityJournalist No Attack

2.9 Differential privacy

The next privacy model is the model of differential privacy, first proposed by Cynthia Dwork in 2006, does not focus on the output dataset on its own but on the data processing method [10, 2]. This model ensures that a disclosure by any attacker is just as likely whether or not one individual does participate in a data base[10].

The privacy model implemented in arx is the model of (ϵ, δ)-Differential Privacy, which is an extension of ϵ-Differential Privacy. ϵ-Differential Privacy can be defined as follows:

> "A randomized function K gives ϵ-differential privacy if for all data sets D_1 and D_2 differing on at most one element, and all S $\subseteq$ Range(K),
>
> $$Pr\left[K(D_1) \in S\right] \leq exp(\epsilon) \times Pr\left[K(D_2) \in S\right]\text{"}[10]$$

As this definition of ϵ-Differential Privacy can be too strict in practice, a commonly used relaxation was introduced, namely (ϵ, δ)-Differential Privacy. This version of differential privacy allows a small error probability δ and is defined as follows[15]:

> "A randomized algorithm A satisfies (ϵ, δ)-differential privacy, if for any pair of neighboring datasets D and D' and for any O $\subseteq$ Range(A):
>
> $$Pr\left[A(D) \in O\right] \leq e^{\epsilon} \times Pr\left[A(D') \in O\right] + \delta\text{"}[15]$$

Due to the fact that (ϵ, δ)-Differential Privacy allows a small error probability, a higher data quality is possible in comparison to the traditional ϵ-Differential Privacy[5]. With this privacy model applied it becomes very difficult for adversaries to gain any information about specific individuals. Although Arx argues that this model protects from record linkage, attribute linkage and table linkage attacks, the work of Fung et al. states that (ϵ, δ)-differential privacy protects only against table linkage and the probabilistic attack[2, 12]. As we move forward the classification of Fung et al. is used.

2.10 Average risk

The next privacy model is the model of Average Risk. This privacy model makes sure that the average risk that a record gets identified is below a certain threshold. Consequently this model protects datasets from record linkage attacks in the marketer model[2].

2.11 Population uniqueness

This privacy model is another way to protect datasets from record linkage attacks in the marketer attack model. It assures that the fraction of record

which are unique within the population falls below a given threshold[2].
The realization of arx implements a decision rule constructed by Dankar et al. which decides between certain population uniqueness estimators. The estimators taken into consideration are the Pitman, slide negative binomial and Zayatz estimators. As no single estimator performs well across all conditions and data sets, this decision rule chooses the most suitable option and thus delivers all around the best results[8].

2.12 Sample uniqueness

For this privacy model mentioned by arx, little to no literature is available. But based on the java code and the short explanation of arx, this privacy model makes sure that the fraction of records which are in a equivalence class of the size one falls below a given threshold[2, 4]. Consequently the privacy model of sample uniqueness protect datasets from record linkage in the marketer model.

2.13 Classification Table

The results from chapter 2 are shown in the following classification table. The attacks mitigated by each privacy model are displayed here in table 2.1.

Privacy Models	Attacks Model			
	Record Linkage	Attribute Linkage	Table Linkage	Probabilistic Attack
k-Anonymity	X			
k-Map	X			
Distinct-l-Diversity	X	X		
Entropy-l-Diversity	X	X		
Recursive (c, l)-Diversity	X	X		
Ordered Distance t-Closeness		X		X
Equal Distance t-Closeness		X		X
Hierarchical Distance t-Closeness		X		X
delta-Disclosure Privacy		X		X
Basic beta-Likeness		X		X
Enhanced beta-Likeness		X		X
delta-Presence			X	
Inclusion				
Profitability Prosecutor	X			
Profitability Journalist	X			
Profitability Prosecutor No Attack	X			
Profitability Journalist No Attack	X			
(e, d)-differential Privacy			X	X
Average Risk	X			
Population Uniqueness	X			
Sample Uniqueness	X			

Table 2.1: Attacks mitigated by each privacy model

3 Evaluation of Privacy Models

Now that the privacy models are classified by the attacks they mitigate, we will now focus on their performance. Therefore, we will measure their performance when applied to different data sets.

3.1 Experimental Setup

For the desired performance evaluation, a privacy benchmark for LPL was used. This benchmark can run different configurations regarding the needed measures, the desired data sets to which the specified privacy models should be applied, the benchmark repetitions and other metrics.
For the support of all 21 privacy models of arx, the benchmark had to be extended in a variety of ways. First of all, due to the fact that only three of the 21 privacy models were supported, the remaining models and their attributes had to be added to the privacy model enumeration and the privacy model attribute enumeration. To give the privacy model attributes fitting values, the most common values for them were added to the default privacy model configuration file. The next step was to extend the 'BenchmarkUtil' class, which has to recognize every privacy model given from the test configuration, to add them to a list of distinct privacy models.
Furthermore, the LPLGenerator class, that generates the wanted Layered Privacy Policy, had to be modified. In this class two methods had to be changed, to support more privacy models. The first method is the '_generatePrivacyModel' method, which as the name suggests creates the specified privacy model. It checks the given privacy model enumeration, and therefore adds the needed privacy model attributes with the default value to the attribute list. How this was done can be seen in Figure 3.1. Besides that, the method '_generateData', which creates the data element had to be extended. This method checks, wether a privacy model requires that a sensitive attribute has to be specified. This is the case for example for the privacy models of l-Diversity. The next class that had to be modified is the class 'ArxAdapter'. Here the method 'applyPrivacyModels' had to be extended. This method reduces the list of all given privacy models and applies them. Once again, as this method only supported three privacy models, the remaining 18 had to be added. For

```
if (name == PrivacyModelEnum.K_ANONYMITY) {
   PrivacyModelAttribute attributeK = new PrivacyModelAttribute();
   attributeK.setKey(PrivacyModelAttributeEnum.K_ANONYMITY_K);
   attributeK.setValue(new
       Integer(privacyModelConfig.k).toString());
   attributeList.add(attributeK);
}
```

Figure 3.1: Code excerpt from the 'generatePrivacyModel' method

every privacy model the corresponding attributes had to be fetched and with those, a new privacy model instance of the required model can be added to the ArxConfiguration of the benchmark. These instances were imported from the 'org.deidentifier.arx.criteria' package, which contains all 21 privacy models of arx. This is shown in Figure 3.2

```
switch (privacyModel.getName()) {
   case BASIC_B_LIKENESS:
      PrivacyModelAttribute privacyModelAttributeB =
          LPLUtils.getSpecificPrivacyModelAttribute(
      privacyModel.getAttributeList(),
          PrivacyModelAttributeEnum.BASIC_B_LIKENESS_B);
      privacyModelAttributeSensitive =
          LPLUtils.getSpecificPrivacyModelAttribute(
      privacyModel.getAttributeList(),
          PrivacyModelAttributeEnum.BASIC_B_LIKENESS_SENSITIVE);

      this.config.addPrivacyModel(new
          BasicBLikeness(privacyModelAttributeSensitive.getValue().toString(),
      Double.parseDouble(privacyModelAttributeB.getValue())));

      break;
}
```

Figure 3.2: Code excerpt from the 'applyPrivacyModels' method

Finally, to be able to run the performance tests with all 21 privacy models, the last class that had to be modified, was the 'ArxRun' class. This class is responsible for running the arx anonymization to be compared to LPL. Here the method '_generateDefaultPrivacyModel' had to be extended, to support every privacy model needed in our tests.
With these changes done, we are now able to run tests on all 21 privacy models of arx, which will deliver the desired information about the performance of those privacy models and privacy model combinations.

Dataset	Size	Distinct Values
Adult	30162	166
CUP	63441	14407
FARS	100937	238
ATUS	539253	305
IHIS	1193504	186

Table 3.1: Used Datasets for the evaluation

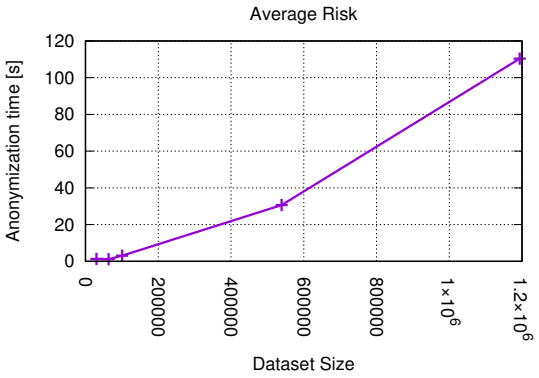
3.2 Tests

The goal of this evaluation is to get the optimal privacy model combination for each attacking scenario. That means the user specifies against which attacks he wants to protect his data and thereby receives the privacy model or privacy model combination that mitigates every attack mentioned with the best performance.
To get the needed data a variety of tests have been performed in order to get the performance data. Therefore five data sets were used to receive performance data for small data, medium data and big data. The data sets used can be seen in Figure 3.1. As we take a look at this table, we can see that the cup data set contains a very high number of distinct values (14407) compared to the others, whose number of values only ranges between approximately 150 to 300 distinct values. Consequently the performance on the cup data set can not be put into consideration for the evaluation. The three data sets chosen to represent small, medium and big data are the adult data set, the fars data set and the ihis data set.
To compare the performance of privacy models we will measure the LPL anonymization time of each privacy model when applied to every data set. This will deliver comparable data for a suitable comparison and evaluation.

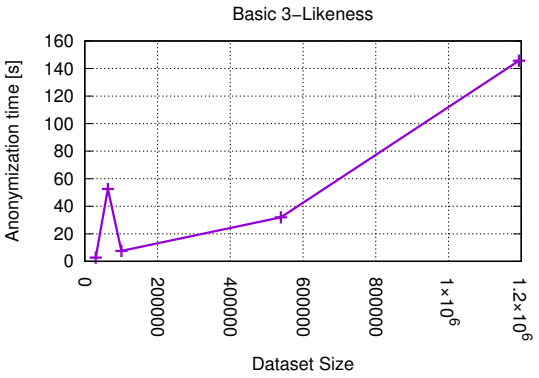
3.2.1 Single Privacy Model Test

The first big test that was made, took all 21 privacy models of arx. Every one of these privacy models has been applied to our five data sets. This procedure was repeated a total of ten times to get a suitable average value for the anonymization time of LPL. The results of this first test run can be seen in the following figures 3.3a through 3.3u.
The results show, that as expected the anonymization time with growing data set size is increasing as well. As mentioned before, the high number of distinct values of the data set cup, causes enormous deviations for some privacy models, for example the model of B-Likeness (Figure 3.3b) or d-Presence(Figure 3.3e). That is why, the values for this data set will not be used in the upcoming

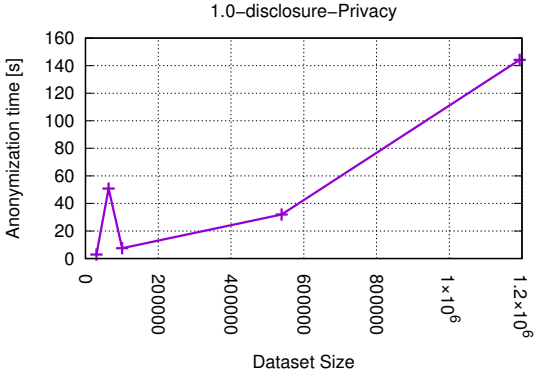
privacy model substitutions.



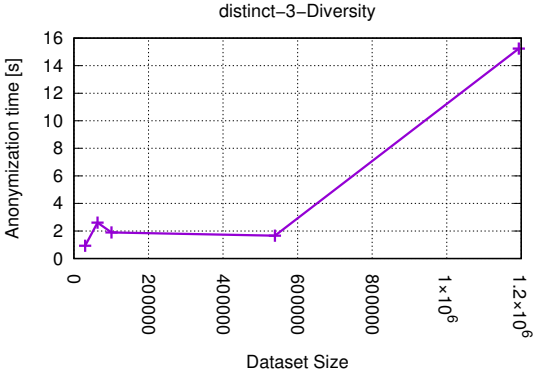
(a) Anonymization time for the Average Risk Privacy Model



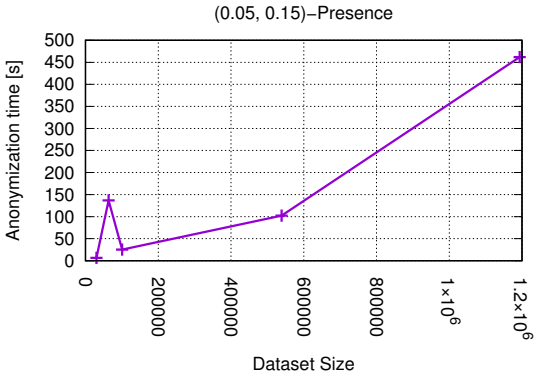
(b) Anonymization time for the Basic B Likeness Privacy Model



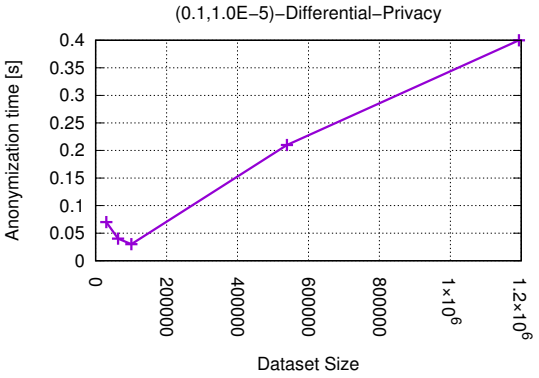
(c) Anonymization time for the d Disclosure Privacy Privacy Model



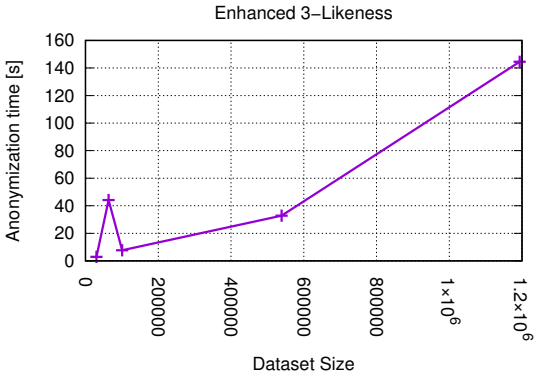
(d) Anonymization time for the Distinct L Diversity Privacy Model



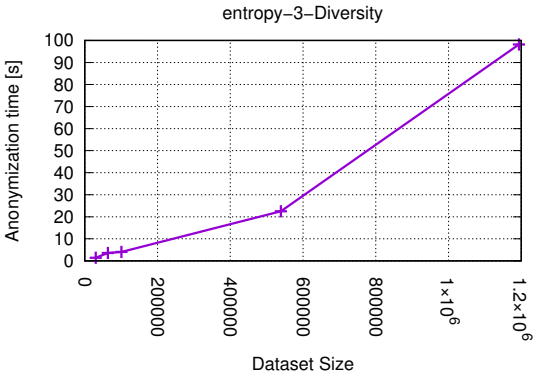
(e) Anonymization time for the d Presence Privacy Model



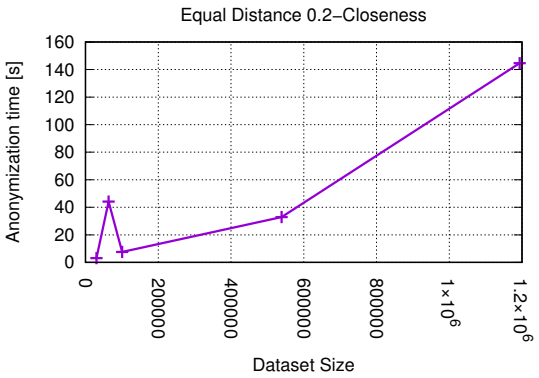
(f) Anonymization time for the (e,d) Differential Privacy Privacy Model



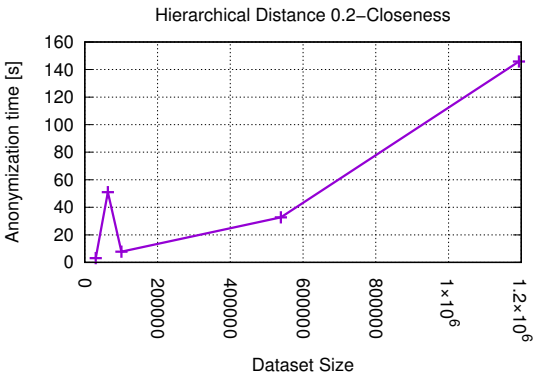
(g) Anonymization time for the Enhanced B Likeness Privacy Model



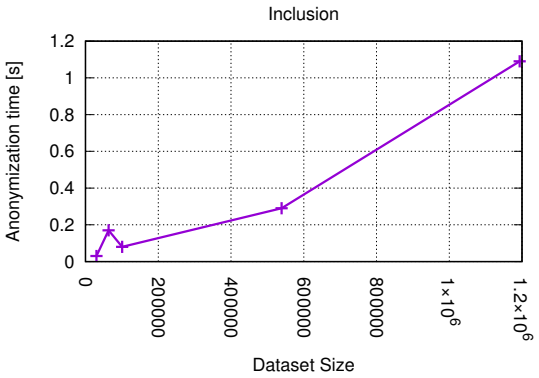
(h) Anonymization time for the Entropy L Diversity Privacy Model



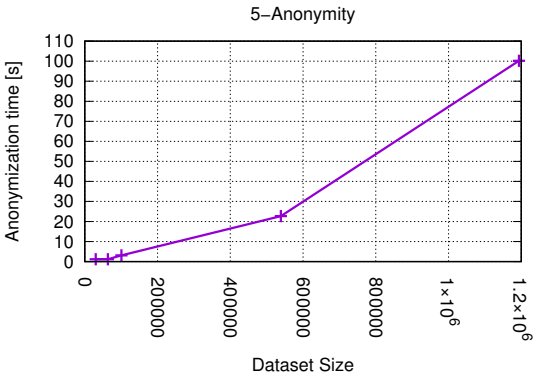
(i) Anonymization time for the Equal Distance T Closeness Privacy Model



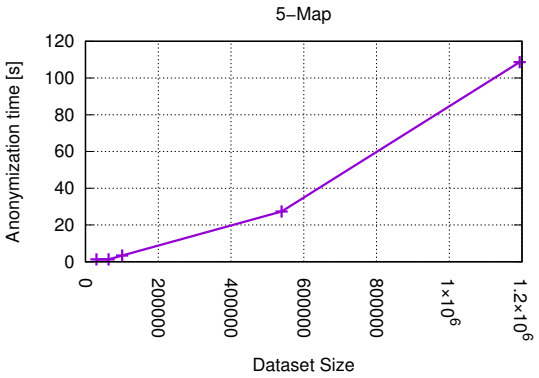
(j) Anonymization time for the Hierarchical Distance T Closeness Privacy Model



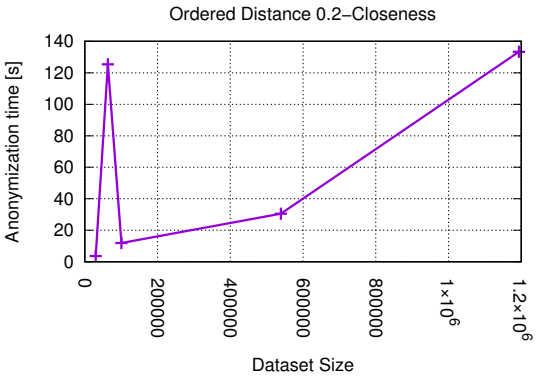
(k) Anonymization time for the Inclusion Privacy Model



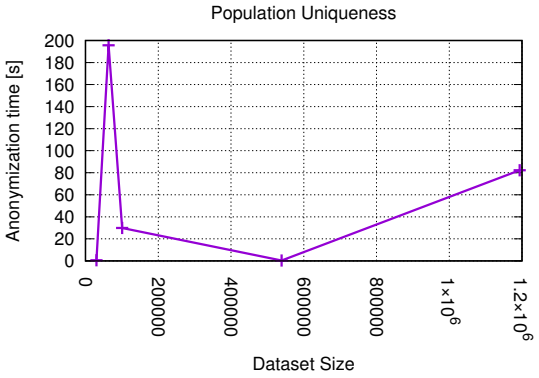
(l) Anonymization time for the k Anonymity Privacy Model



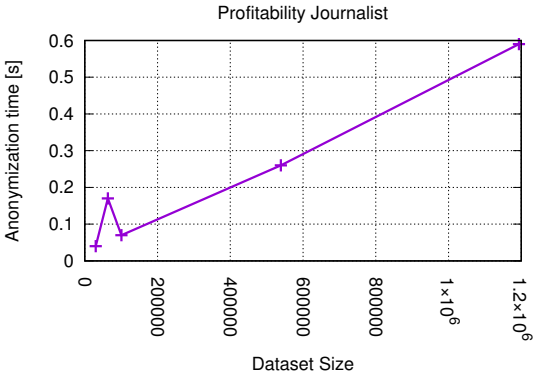
(m) Anonymization time for the k Map Privacy Model



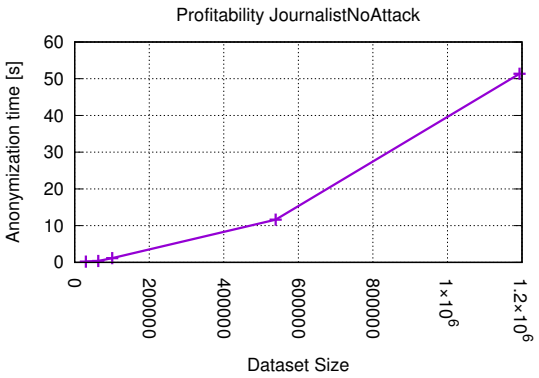
(n) Anonymization time for the Ordered Distance T Closeness Privacy Model



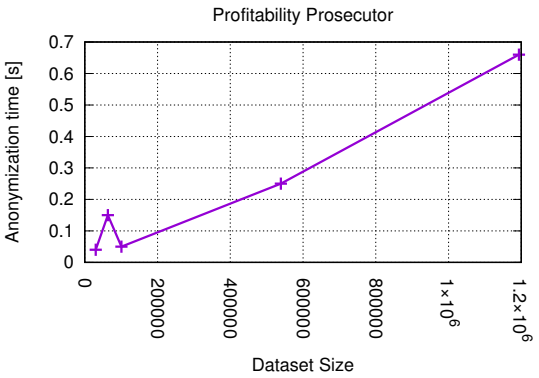
(o) Anonymization time for the Population Uniqueness Privacy Model



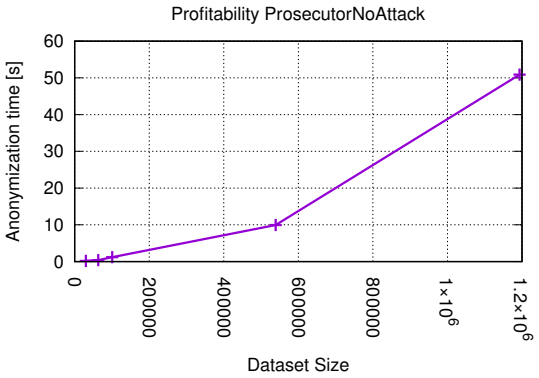
(p) Anonymization time for the Profitability Journalist Privacy Model



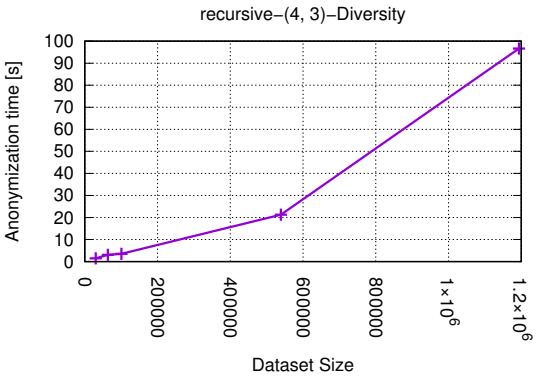
(q) Anonymization time for the Profitability Journalist No Attack Privacy Model



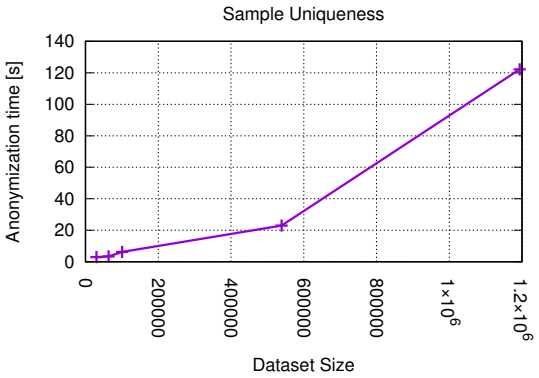
(r) Anonymization time for the Profitability Prosecutor Privacy Model



(s) Anonymization time for the Profitability Prosecutor No Attack Privacy Model



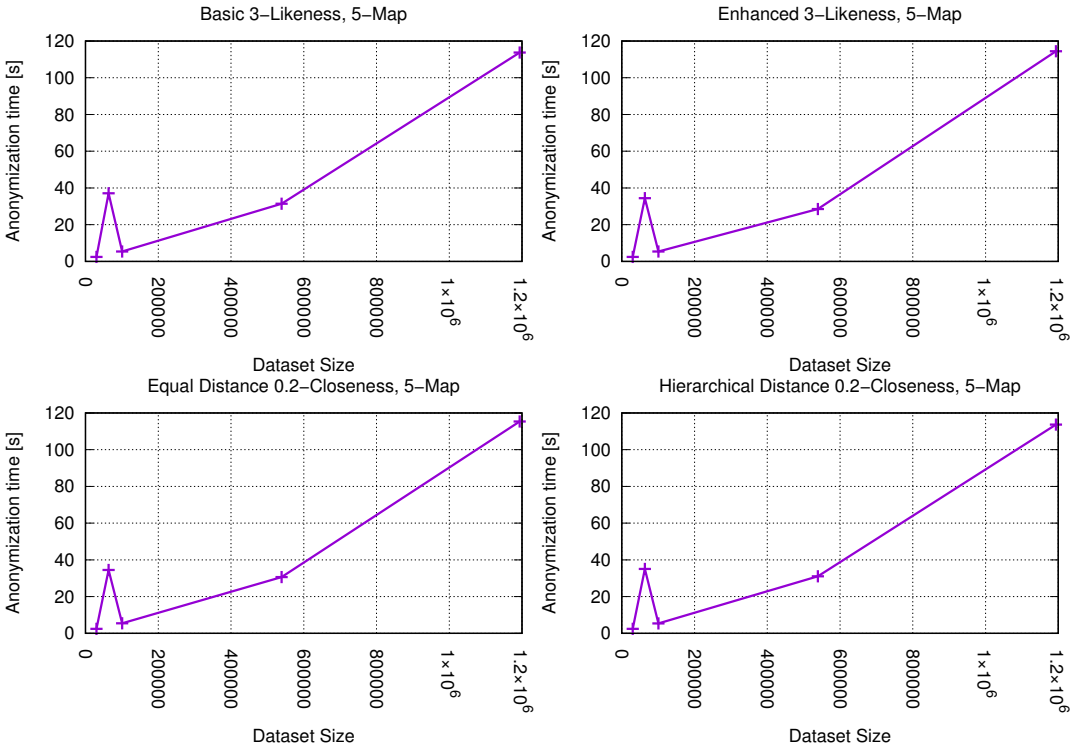
(t) Anonymization time for the Recursive (c,l) Diversity Privacy Model

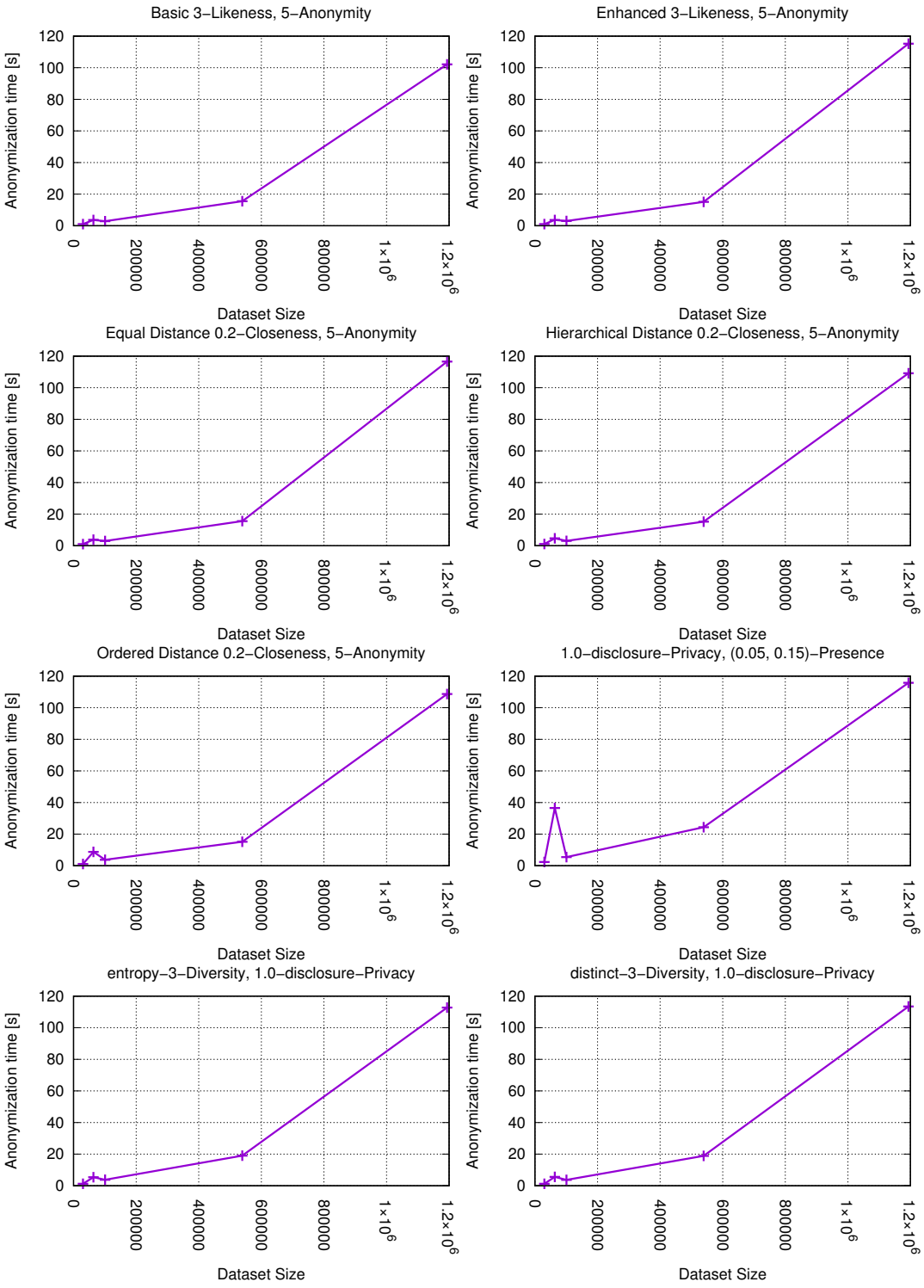


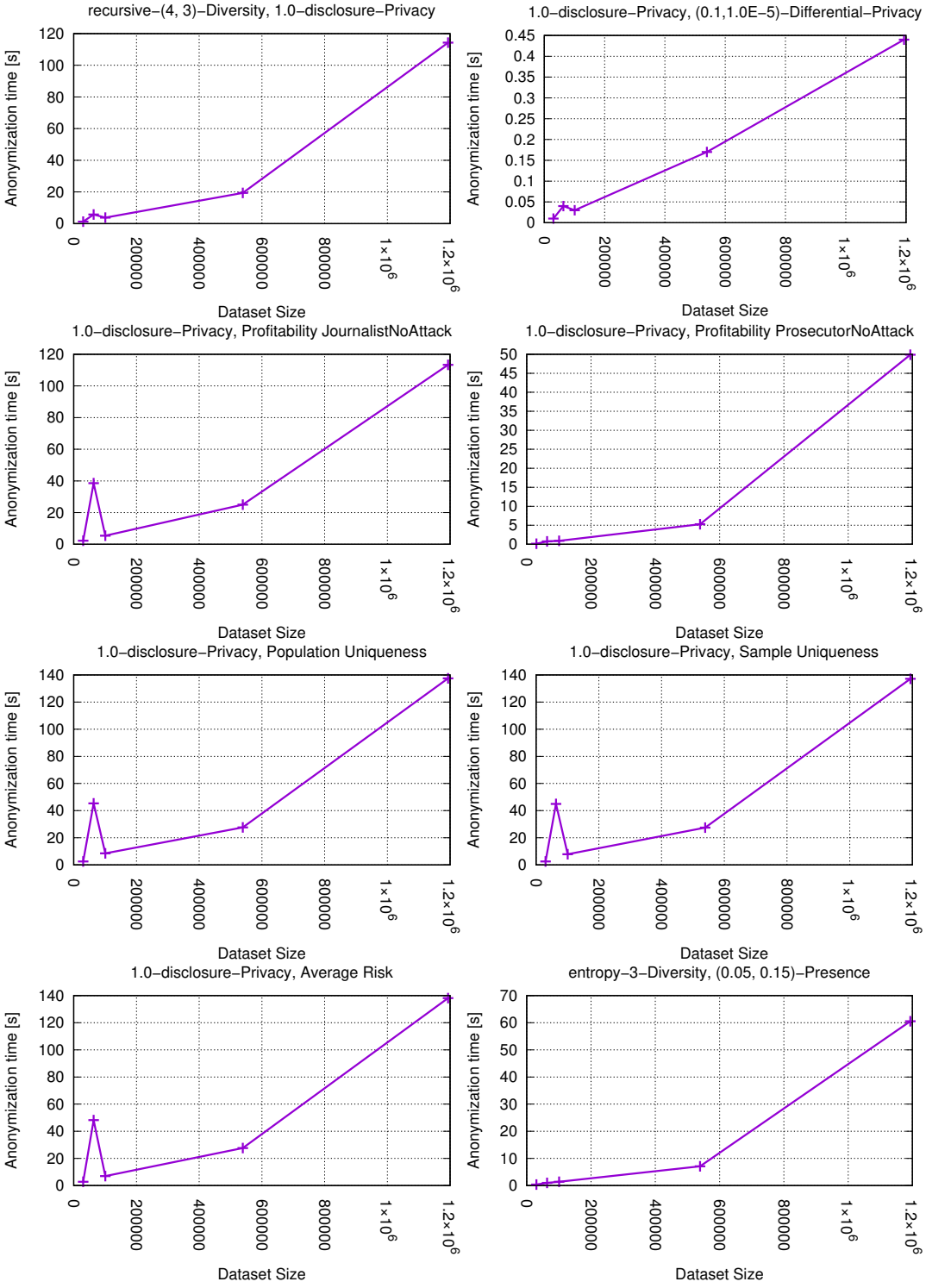
(u) Anonymization time for the Sample Uniqueness Privacy Model

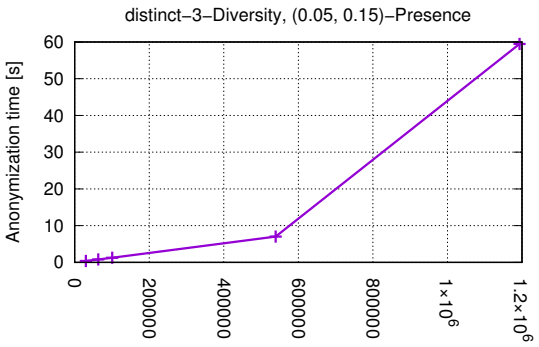
3.2.2 Privacy Model Combination Test

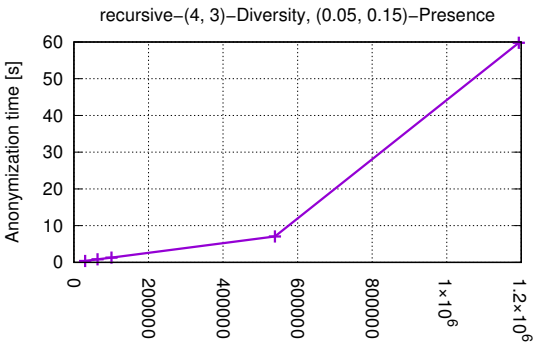
The second test run took 103 privacy model combinations and measured their performance. These combinations were not chosen randomly. They were chosen, because these are the privacy models that can not be substituted. That means, if a set of these privacy model were chosen, both of them would have to be applied to the data set.
Let us clarify that with an example. Given two privacy models (Basic-B-Likeness and k-Map), these privacy model can not be reduced to one, as they mitigate a different set of attacks. If we look at Table 2.1, we can see that Basic-B-Likeness protects against Attribute Linkage Attacks and the Probabilistic Attack, whereas k-Map mitigates only Record Linkage Attacks. Consequently, one can not reduce these two input models to one and both have to be applied. The results of this test run delivers the data about which privacy model combination performs the best for small, medium and big data as displayed in Figure 3.4.

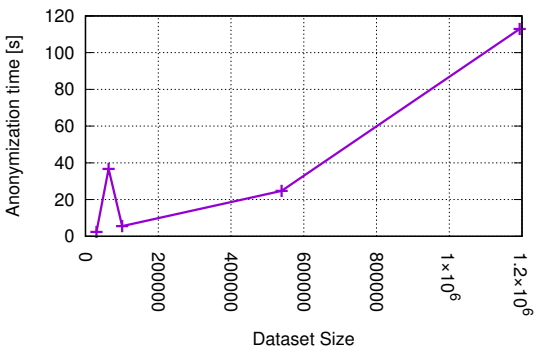


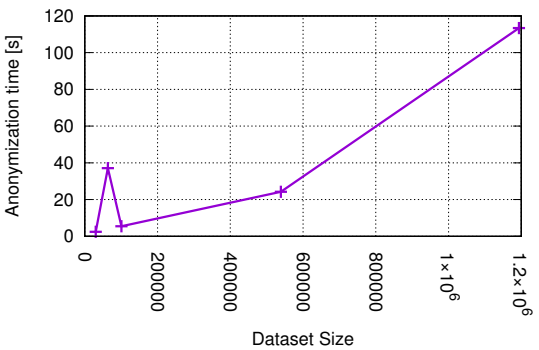

Basic 3−Likeness, 5−Anonymity
Enhanced 3−Likeness, 5−Anonymity
Equal Distance 0.2−Closeness, 5−Anonymity
Hierarchical Distance 0.2−Closeness, 5−Anonymity
Ordered Distance 0.2−Closeness, 5−Anonymity
1.0−disclosure−Privacy, (0.05, 0.15)−Presence
entropy−3−Diversity, 1.0−disclosure−Privacy
distinct−3−Diversity, 1.0−disclosure−Privacy
Anonymization time [s]
Dataset Size
0
20
40
60
80
100
120
200000
400000
600000
800000
1×10^6
1.2×10^6

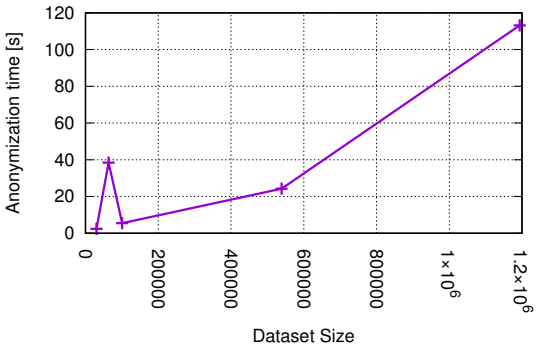

recursive−(4, 3)−Diversity, 1.0−disclosure−Privacy
1.0−disclosure−Privacy, (0.1,1.0E−5)−Differential−Privacy
1.0−disclosure−Privacy, Profitability JournalistNoAttack
1.0−disclosure−Privacy, Profitability ProsecutorNoAttack
1.0−disclosure−Privacy, Population Uniqueness
1.0−disclosure−Privacy, Sample Uniqueness
1.0−disclosure−Privacy, Average Risk
entropy−3−Diversity, (0.05, 0.15)−Presence
Anonymization time [s]
Dataset Size
0
200000
400000
600000
800000
1×10^6
1.2×10^6

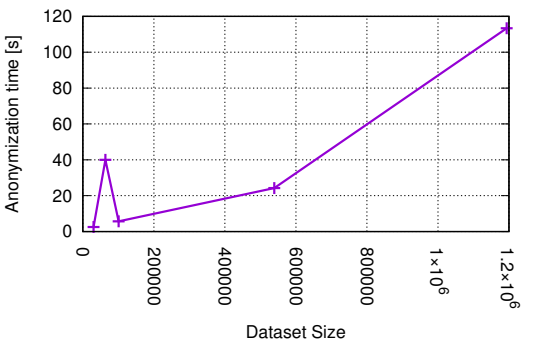

distinct−3−Diversity, (0.05, 0.15)−Presence
Anonymization time [s]
0
10
20
30
40
50
60
0
200000
400000
600000
800000
1×10^6
1.2×10^6
Dataset Size

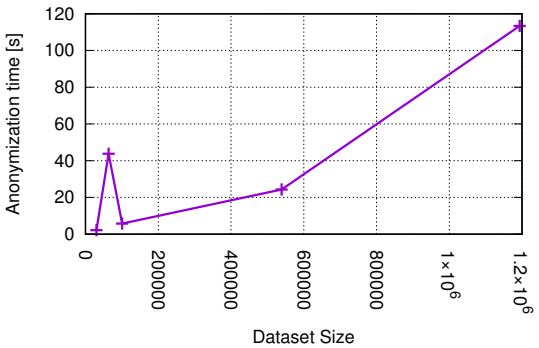

recursive−(4, 3)−Diversity, (0.05, 0.15)−Presence
Anonymization time [s]
0
10
20
30
40
50
60
0
200000
400000
600000
800000
1×10^6
1.2×10^6
Dataset Size

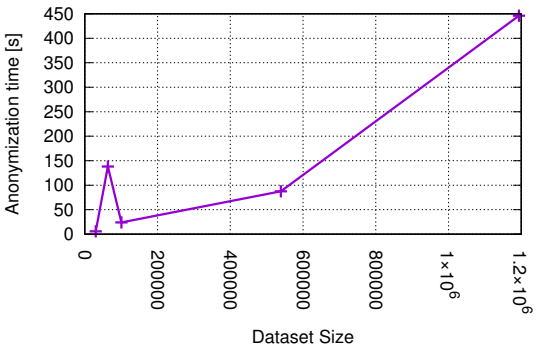

Basic 3−Likeness, (0.05, 0.15)−Presence
Anonymization time [s]
0
20
40
60
80
100
120
0
200000
400000
600000
800000
1×10^6
1.2×10^6
Dataset Size

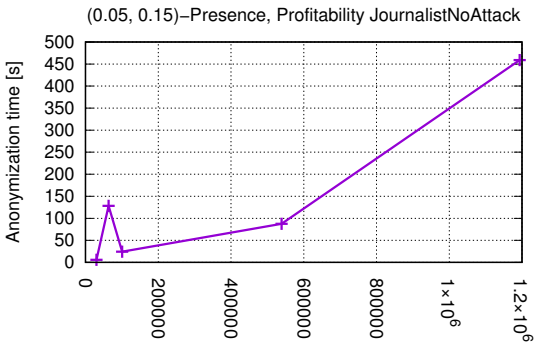

Enhanced 3−Likeness, (0.05, 0.15)−Presence
Anonymization time [s]
0
20
40
60
80
100
120
0
200000
400000
600000
800000
1×10^6
1.2×10^6
Dataset Size

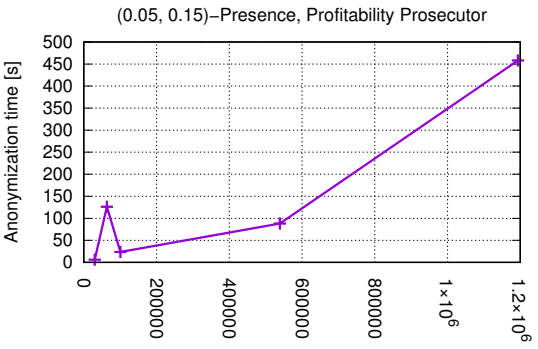

Equal Distance 0.2−Closeness, (0.05, 0.15)−Presence
Anonymization time [s]
0
20
40
60
80
100
120
0
200000
400000
600000
800000
1×10^6
1.2×10^6
Dataset Size

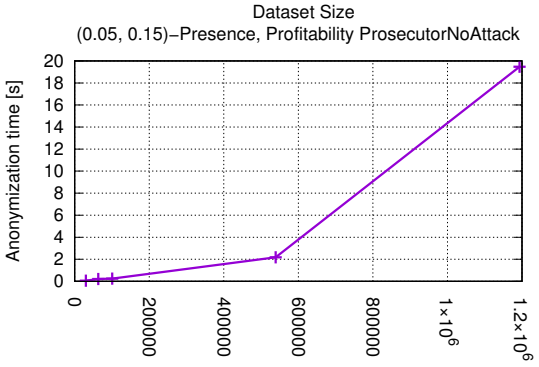

Hierarchical Distance 0.2−Closeness, (0.05, 0.15)−Presence
Anonymization time [s]
0
20
40
60
80
100
120
0
200000
400000
600000
800000
1×10^6
1.2×10^6
Dataset Size

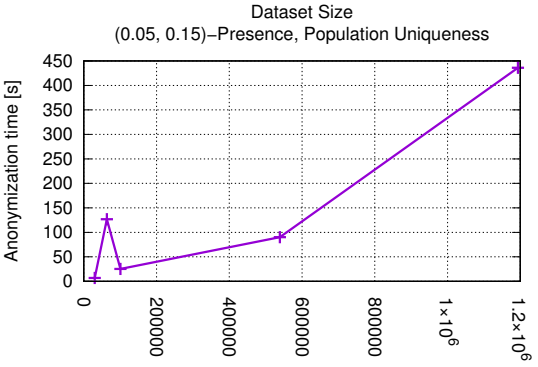

Ordered Distance 0.2−Closeness, (0.05, 0.15)−Presence
Anonymization time [s]
0
20
40
60
80
100
120
0
200000
400000
600000
800000
1×10^6
1.2×10^6
Dataset Size

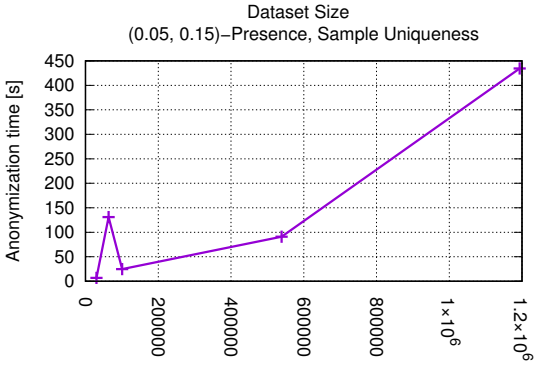

(0.05, 0.15)−Presence, Profitability Journalist
Anonymization time [s]
0
50
100
150
200
250
300
350
400
450
0
200000
400000
600000
800000
1×10^6
1.2×10^6
Dataset Size

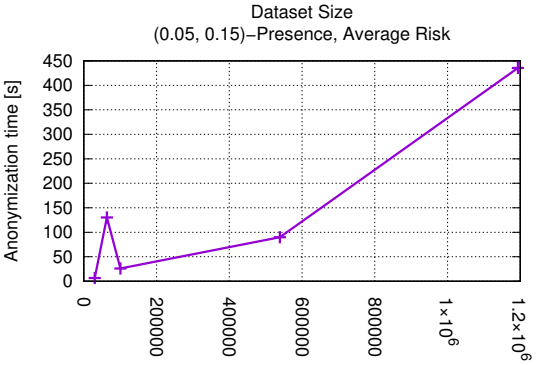

(0.05, 0.15)−Presence, Profitability JournalistNoAttack
Anonymization time [s]
Dataset Size

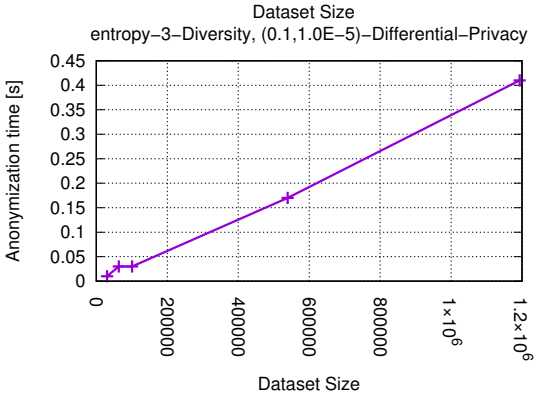

(0.05, 0.15)−Presence, Profitability Prosecutor
Anonymization time [s]
Dataset Size

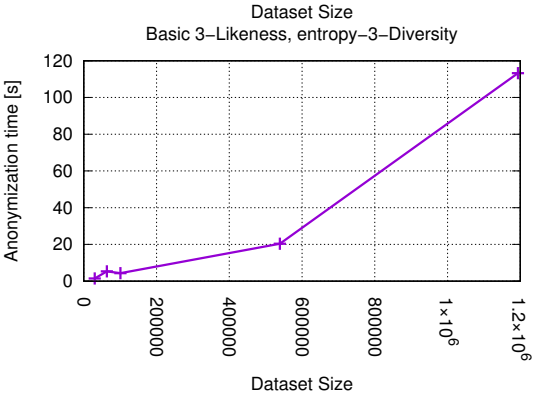

(0.05, 0.15)−Presence, Profitability ProsecutorNoAttack
Anonymization time [s]
Dataset Size

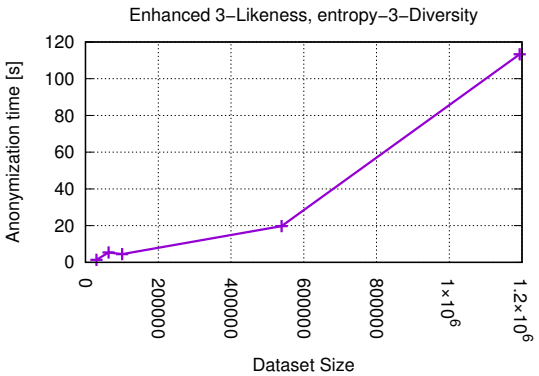

(0.05, 0.15)−Presence, Population Uniqueness
Anonymization time [s]
Dataset Size

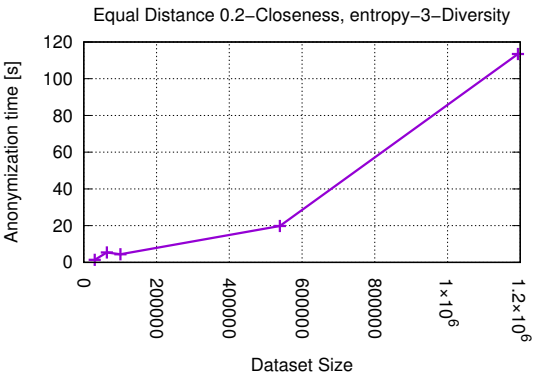

(0.05, 0.15)−Presence, Sample Uniqueness
Anonymization time [s]
Dataset Size

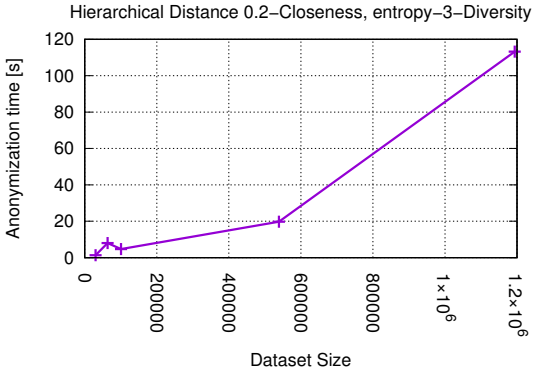

(0.05, 0.15)−Presence, Average Risk
Anonymization time [s]
Dataset Size


entropy−3−Diversity, (0.1,1.0E−5)−Differential−Privacy
Anonymization time [s]
Dataset Size

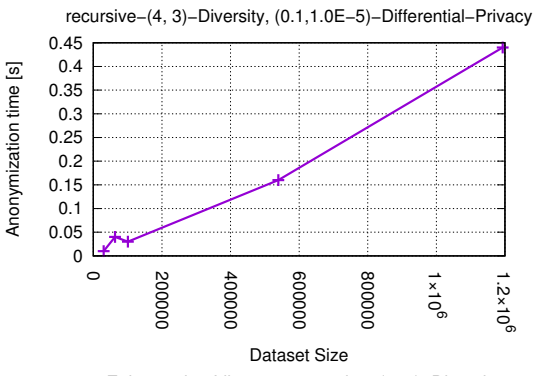

Basic 3−Likeness, entropy−3−Diversity
Anonymization time [s]
Dataset Size

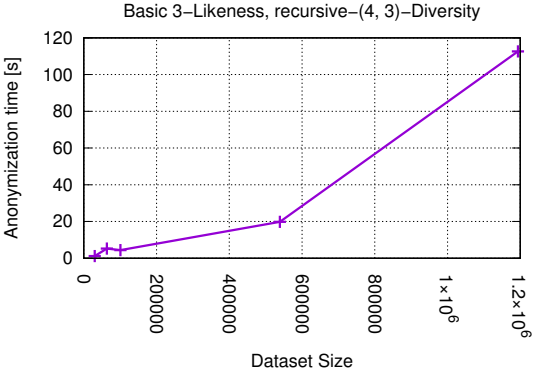

Enhanced 3−Likeness, entropy−3−Diversity
Anonymization time [s]
Dataset Size

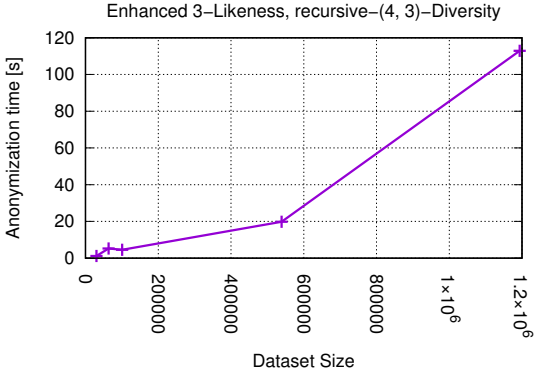

Equal Distance 0.2−Closeness, entropy−3−Diversity
Anonymization time [s]
Dataset Size

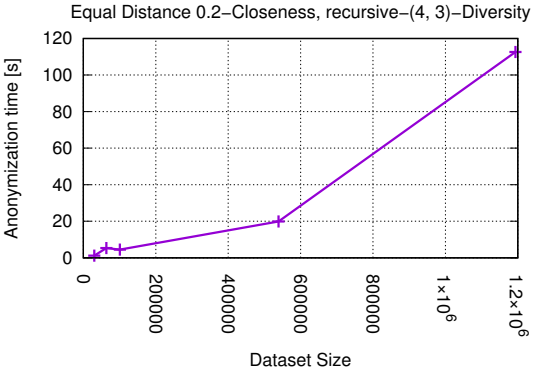

Hierarchical Distance 0.2−Closeness, entropy−3−Diversity
Anonymization time [s]
Dataset Size

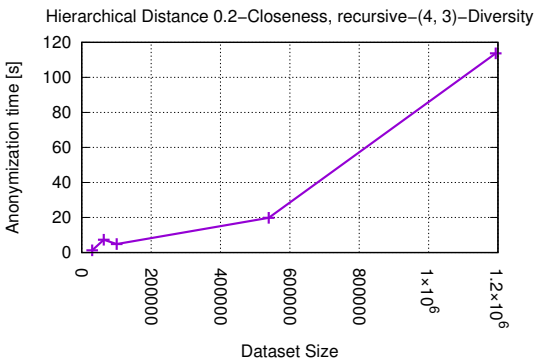

Ordered Distance 0.2−Closeness, entropy−3−Diversity
Anonymization time [s]
Dataset Size

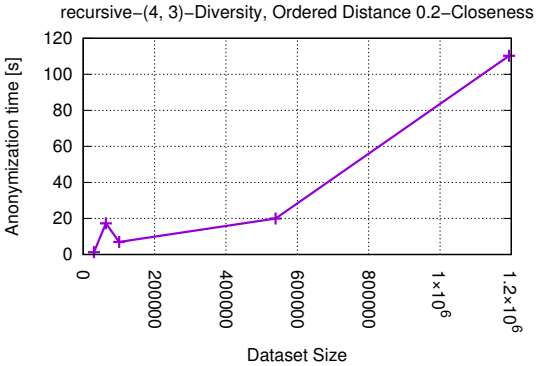

recursive−(4, 3)−Diversity, (0.1,1.0E−5)−Differential−Privacy
Anonymization time [s]
Dataset Size

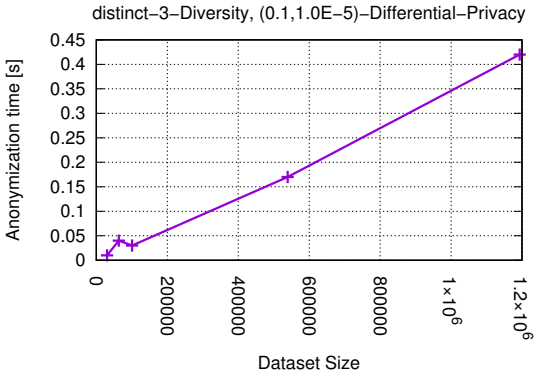

Basic 3−Likeness, recursive−(4, 3)−Diversity
Anonymization time [s]
Dataset Size

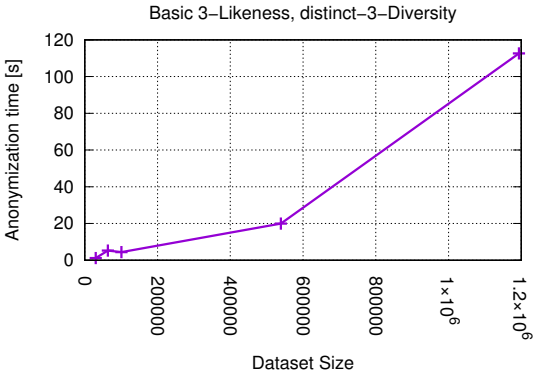

Enhanced 3−Likeness, recursive−(4, 3)−Diversity
Anonymization time [s]
Dataset Size

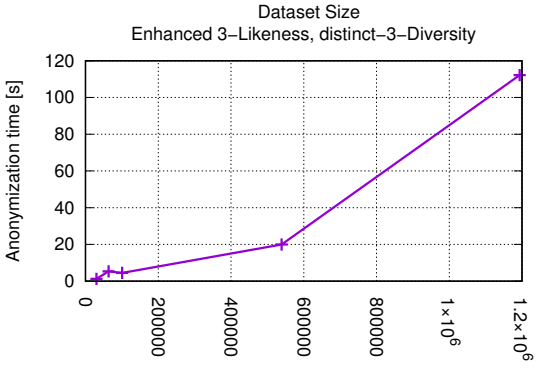

Equal Distance 0.2−Closeness, recursive−(4, 3)−Diversity
Anonymization time [s]
Dataset Size

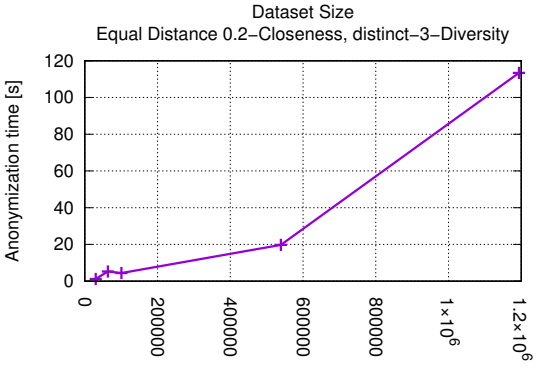

Hierarchical Distance 0.2−Closeness, recursive−(4, 3)−Diversity
Anonymization time [s]
Dataset Size

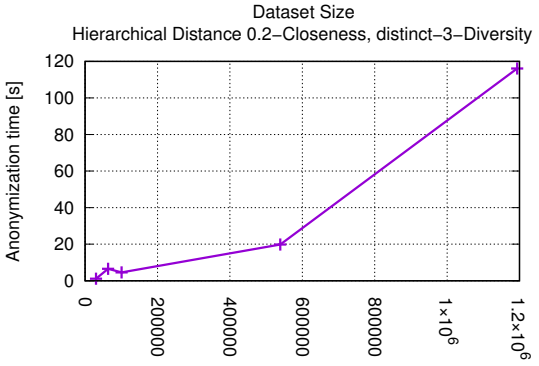

recursive−(4, 3)−Diversity, Ordered Distance 0.2−Closeness
Anonymization time [s]
Dataset Size

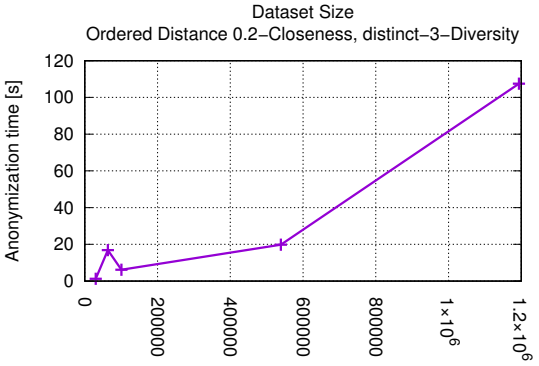

distinct−3−Diversity, (0.1,1.0E−5)−Differential−Privacy
Anonymization time [s]
Dataset Size

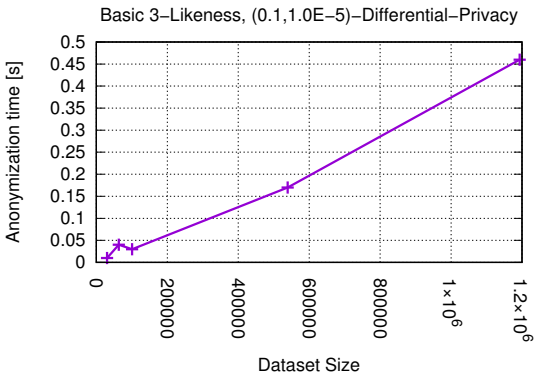

Basic 3−Likeness, distinct−3−Diversity
Anonymization time [s]
Dataset Size

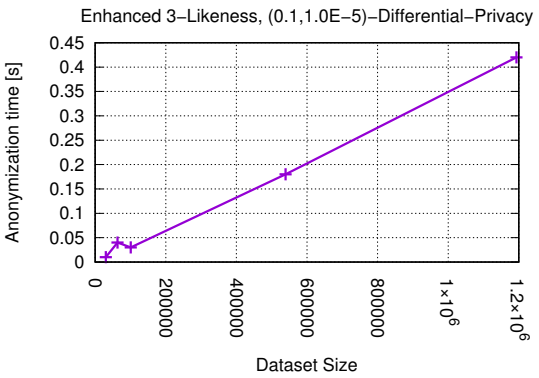

Enhanced 3−Likeness, distinct−3−Diversity
Anonymization time [s]
Dataset Size

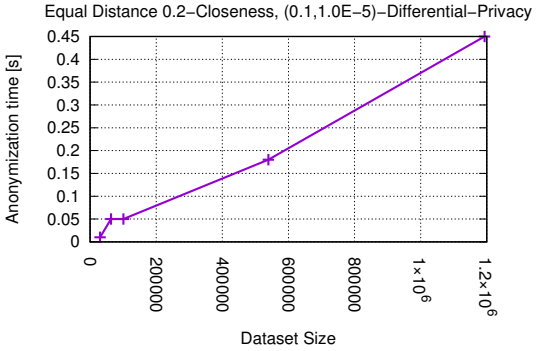

Equal Distance 0.2−Closeness, distinct−3−Diversity
Anonymization time [s]
Dataset Size

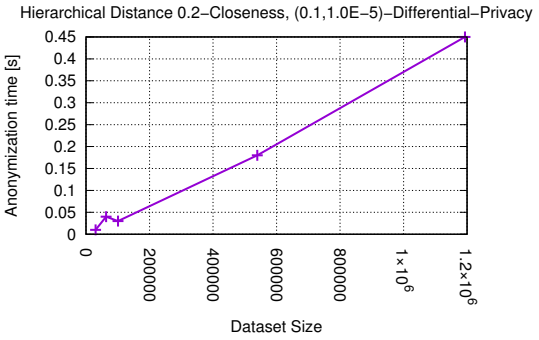

Hierarchical Distance 0.2−Closeness, distinct−3−Diversity
Anonymization time [s]
Dataset Size

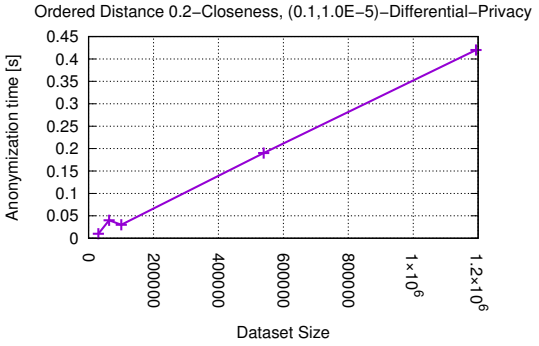

Ordered Distance 0.2−Closeness, distinct−3−Diversity
Anonymization time [s]
Dataset Size

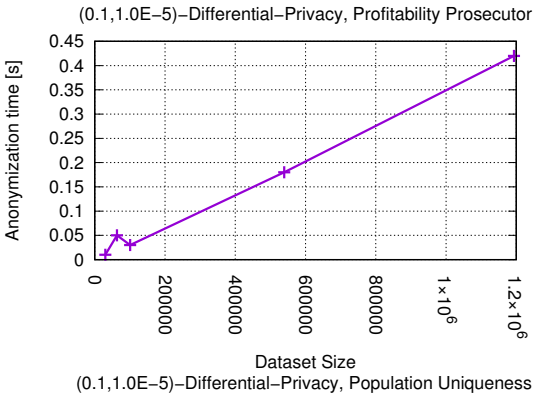

Basic 3−Likeness, (0.1,1.0E−5)−Differential−Privacy
Anonymization time [s]
Dataset Size

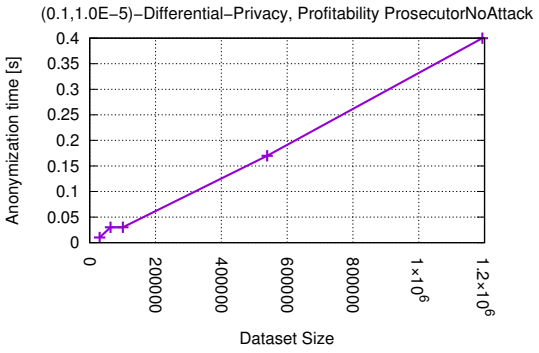

Enhanced 3−Likeness, (0.1,1.0E−5)−Differential−Privacy
Anonymization time [s]
Dataset Size

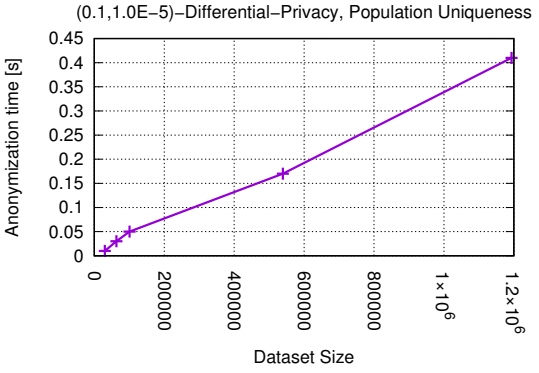

Equal Distance 0.2−Closeness, (0.1,1.0E−5)−Differential−Privacy
Anonymization time [s]
Dataset Size

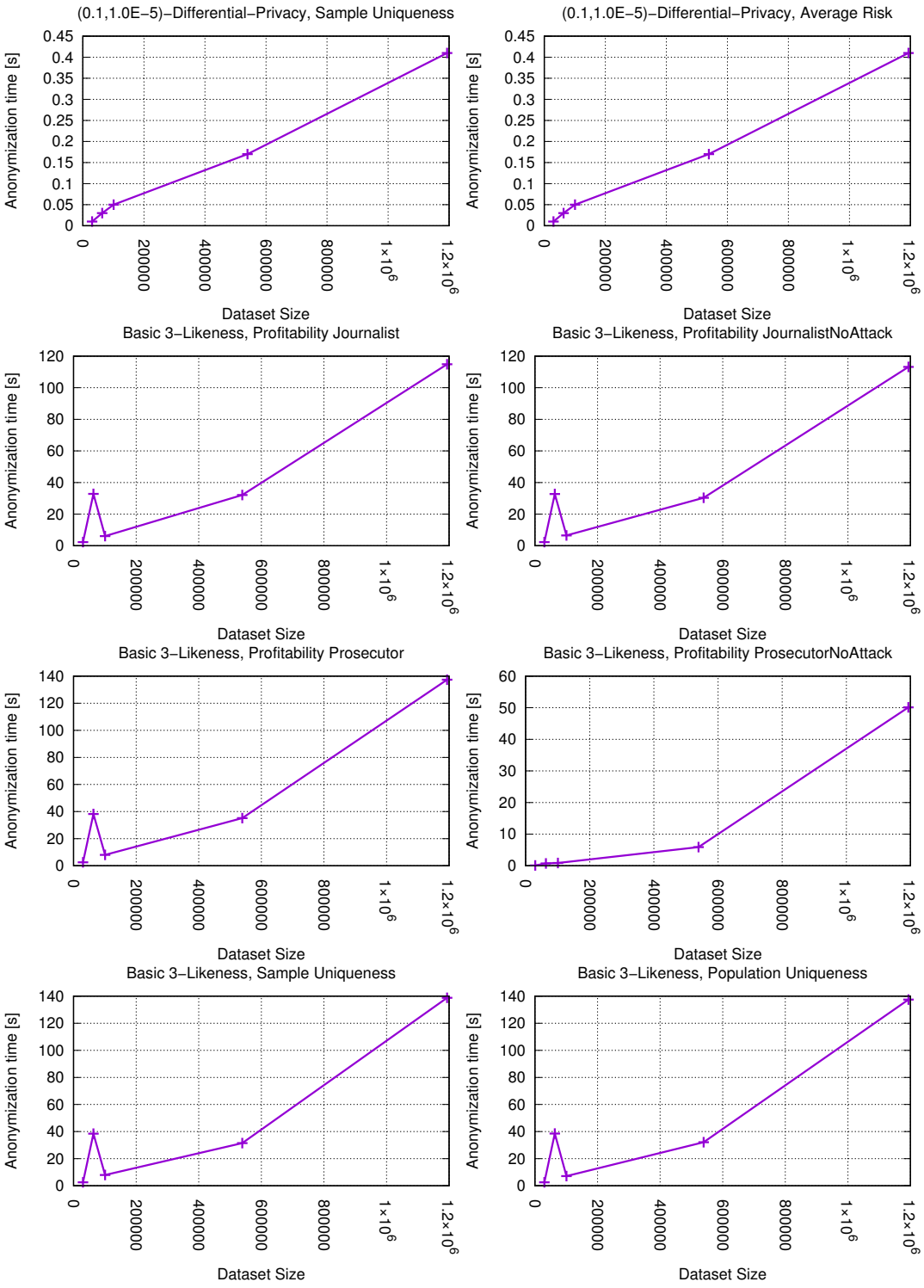

Hierarchical Distance 0.2−Closeness, (0.1,1.0E−5)−Differential−Privacy
Anonymization time [s]
Dataset Size

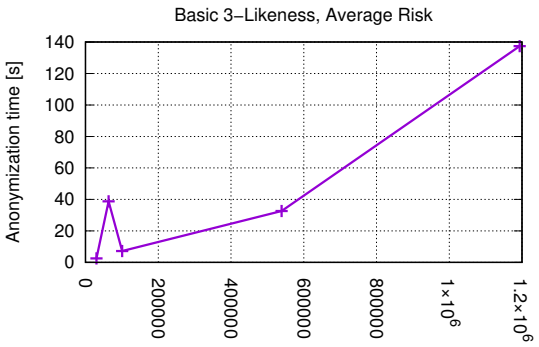

Ordered Distance 0.2−Closeness, (0.1,1.0E−5)−Differential−Privacy
Anonymization time [s]
Dataset Size

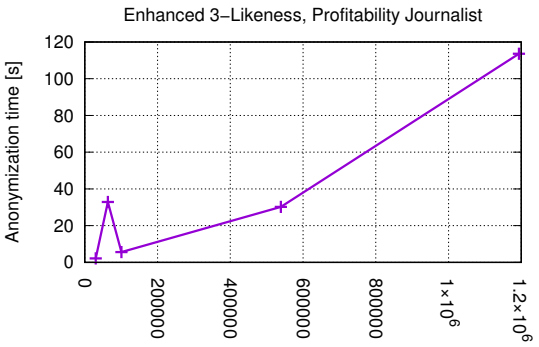

(0.1,1.0E−5)−Differential−Privacy, Profitability Prosecutor
Anonymization time [s]
Dataset Size

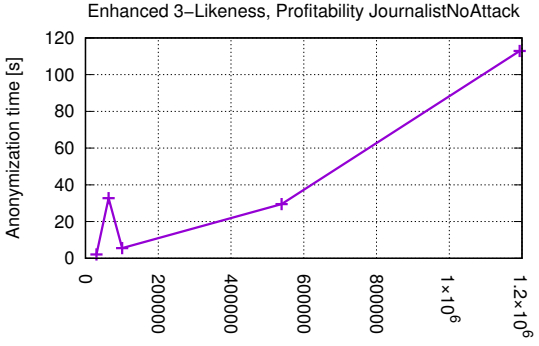

(0.1,1.0E−5)−Differential−Privacy, Profitability ProsecutorNoAttack
Anonymization time [s]
Dataset Size

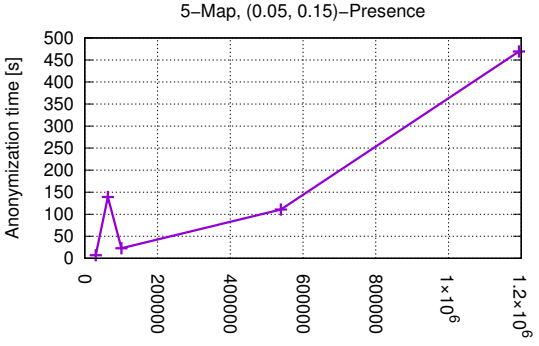

(0.1,1.0E−5)−Differential−Privacy, Population Uniqueness
Anonymization time [s]
Dataset Size

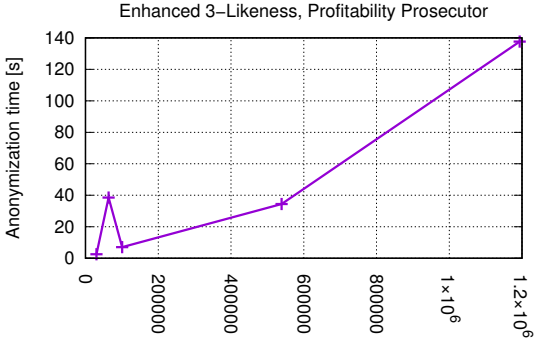

(0.1,1.0E−5)−Differential−Privacy, Sample Uniqueness
(0.1,1.0E−5)−Differential−Privacy, Average Risk
Basic 3−Likeness, Profitability Journalist
Basic 3−Likeness, Profitability JournalistNoAttack
Basic 3−Likeness, Profitability Prosecutor
Basic 3−Likeness, Profitability ProsecutorNoAttack
Basic 3−Likeness, Sample Uniqueness
Basic 3−Likeness, Population Uniqueness
Anonymization time [s]
Dataset Size
0
200000
400000
600000
800000
1×10^6
1.2×10^6

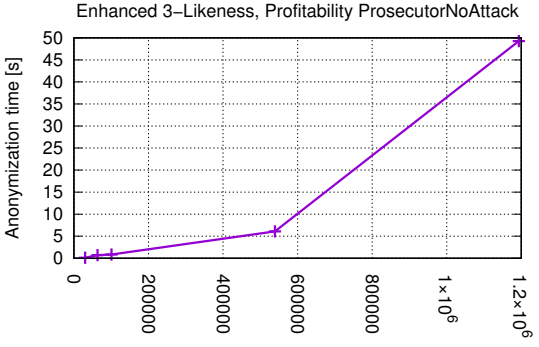

Basic 3−Likeness, Average Risk
Anonymization time [s]
0
20
40
60
80
100
120
140
0
200000
400000
600000
800000
1×10^6
1.2×10^6
Dataset Size

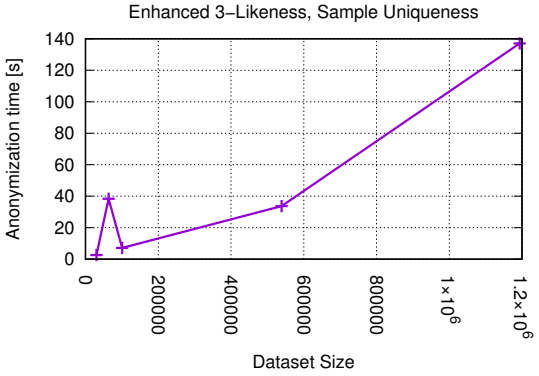

Enhanced 3−Likeness, Profitability Journalist
Anonymization time [s]
0
20
40
60
80
100
120
0
200000
400000
600000
800000
1×10^6
1.2×10^6
Dataset Size

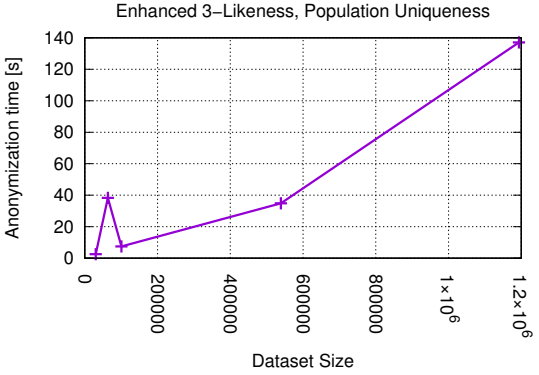

Enhanced 3−Likeness, Profitability JournalistNoAttack
Anonymization time [s]
0
20
40
60
80
100
120
0
200000
400000
600000
800000
1×10^6
1.2×10^6
Dataset Size

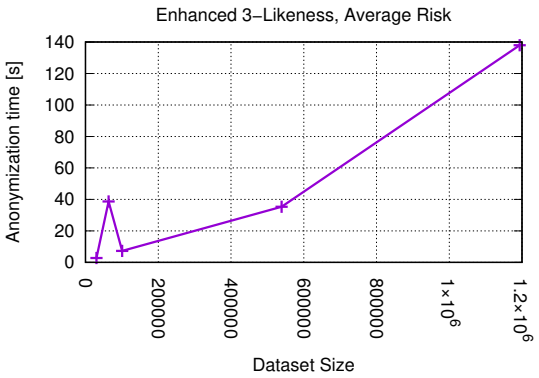

5−Map, (0.05, 0.15)−Presence
Anonymization time [s]
0
50
100
150
200
250
300
350
400
450
500
0
200000
400000
600000
800000
1×10^6
1.2×10^6
Dataset Size

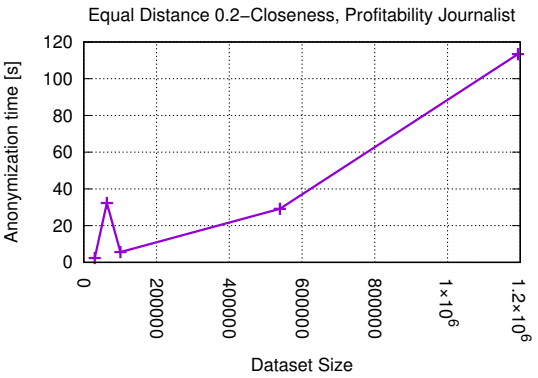

Enhanced 3−Likeness, Profitability Prosecutor
Anonymization time [s]
0
20
40
60
80
100
120
140
0
200000
400000
600000
800000
1×10^6
1.2×10^6
Dataset Size

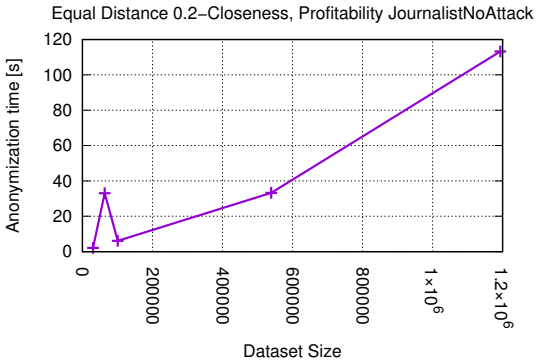

Enhanced 3−Likeness, Profitability ProsecutorNoAttack
Anonymization time [s]
0
5
10
15
20
25
30
35
40
45
50
0
200000
400000
600000
800000
1×10^6
1.2×10^6
Dataset Size

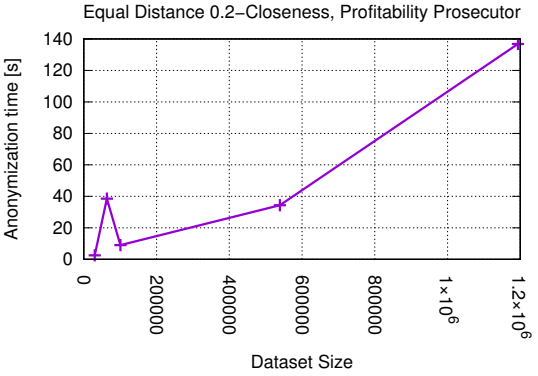

Enhanced 3−Likeness, Sample Uniqueness
Anonymization time [s]
0
20
40
60
80
100
120
140
0
200000
400000
600000
800000
1×10^6
1.2×10^6
Dataset Size

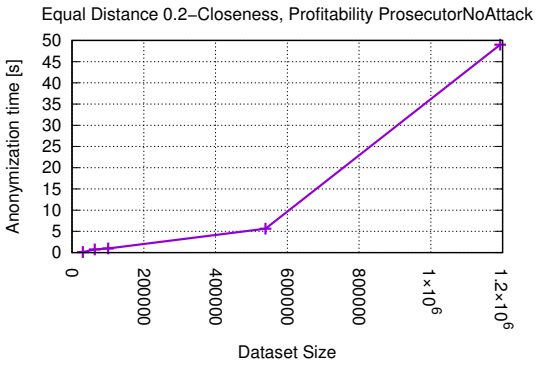

Enhanced 3−Likeness, Population Uniqueness
Anonymization time [s]
0
20
40
60
80
100
120
140
0
200000
400000
600000
800000
1×10^6
1.2×10^6
Dataset Size

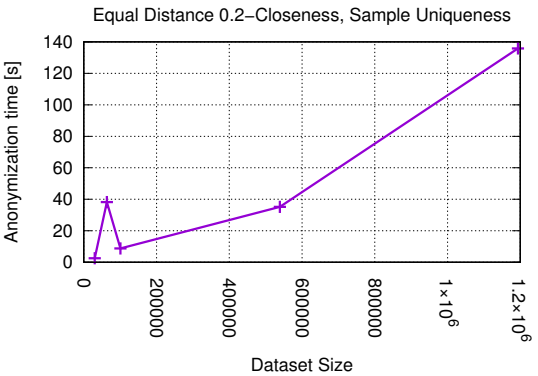

Enhanced 3−Likeness, Average Risk
Anonymization time [s]
Dataset Size

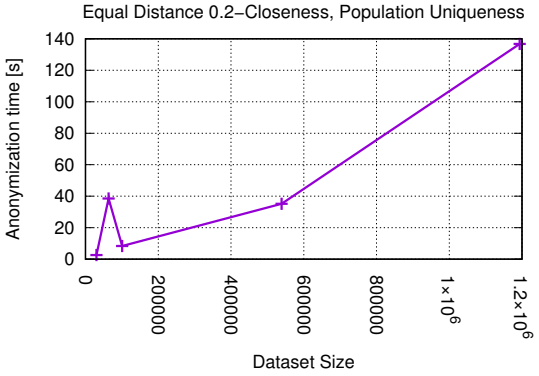

Equal Distance 0.2−Closeness, Profitability Journalist
Anonymization time [s]
Dataset Size

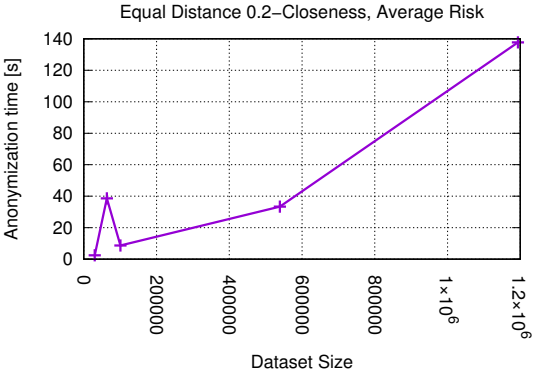

Equal Distance 0.2−Closeness, Profitability JournalistNoAttack
Anonymization time [s]
Dataset Size

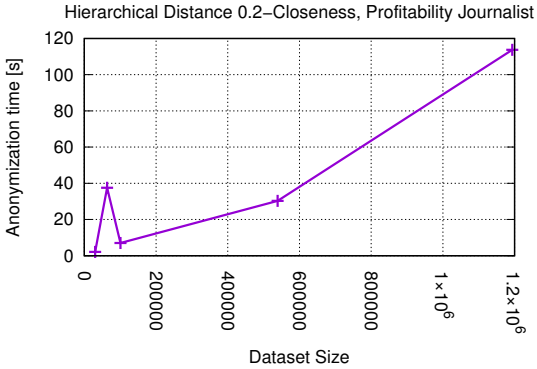

Equal Distance 0.2−Closeness, Profitability Prosecutor
Anonymization time [s]
Dataset Size

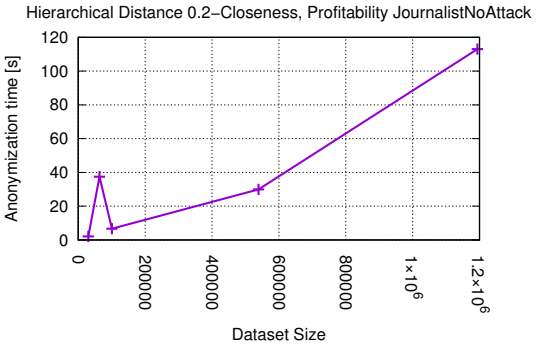

Equal Distance 0.2−Closeness, Profitability ProsecutorNoAttack
Anonymization time [s]
Dataset Size

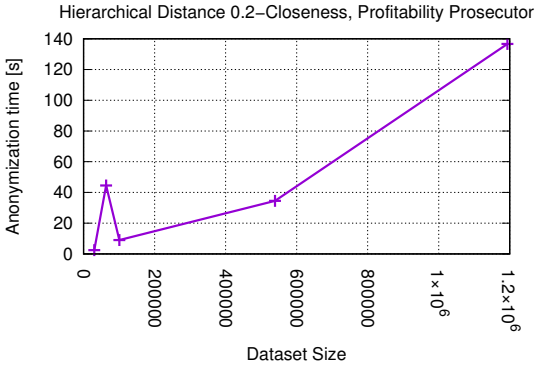

Equal Distance 0.2−Closeness, Sample Uniqueness
Anonymization time [s]
Dataset Size

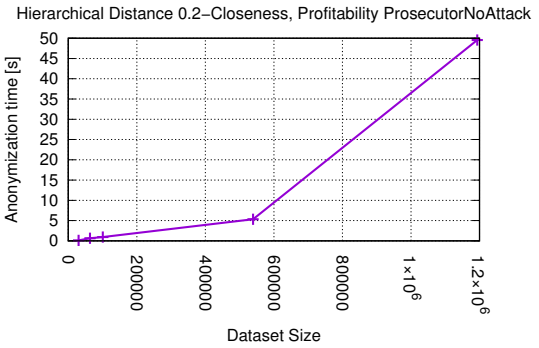

Equal Distance 0.2−Closeness, Population Uniqueness
Anonymization time [s]
Dataset Size

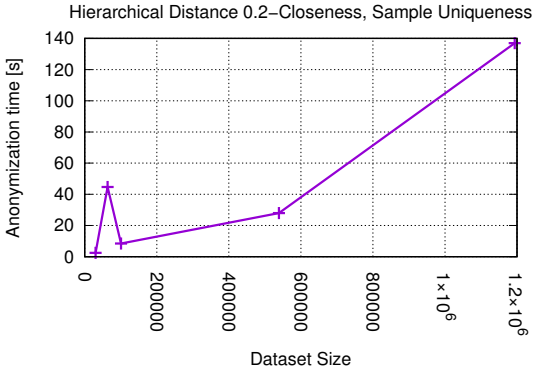

Equal Distance 0.2−Closeness, Average Risk
Anonymization time [s]
Dataset Size

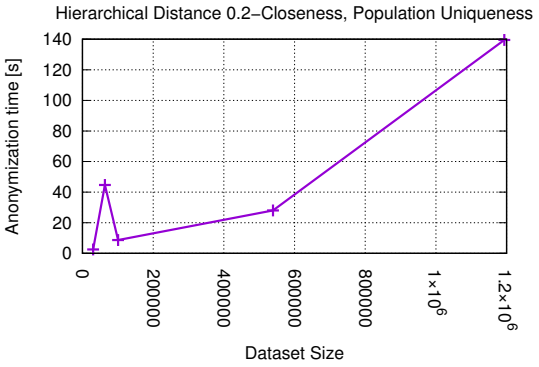

Hierarchical Distance 0.2−Closeness, Profitability Journalist
Anonymization time [s]
Dataset Size

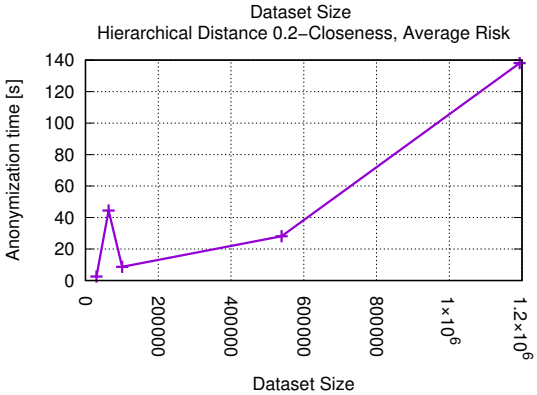

Hierarchical Distance 0.2−Closeness, Profitability JournalistNoAttack
Anonymization time [s]
Dataset Size

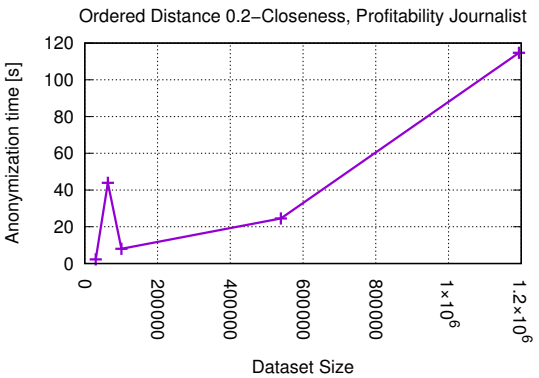

Hierarchical Distance 0.2−Closeness, Profitability Prosecutor
Anonymization time [s]
Dataset Size

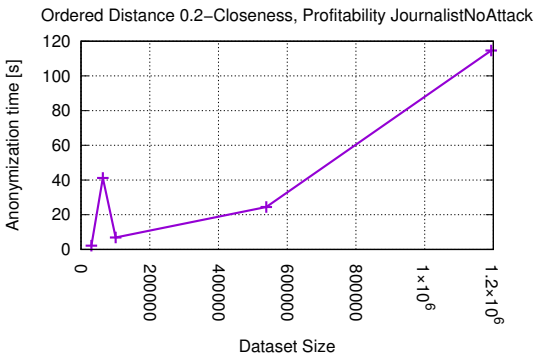

Hierarchical Distance 0.2−Closeness, Profitability ProsecutorNoAttack
Anonymization time [s]
Dataset Size

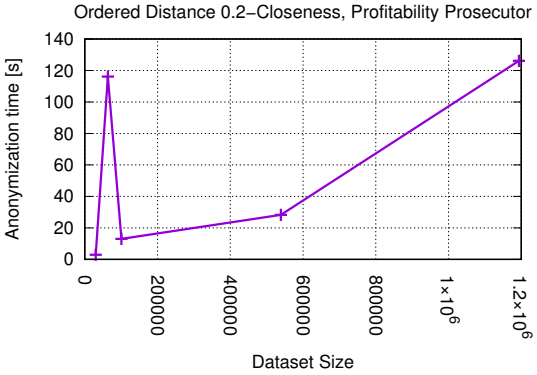

Hierarchical Distance 0.2−Closeness, Sample Uniqueness
Anonymization time [s]
Dataset Size

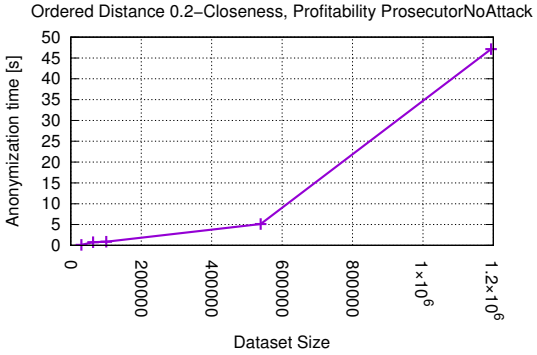

Hierarchical Distance 0.2−Closeness, Population Uniqueness
Anonymization time [s]
Dataset Size

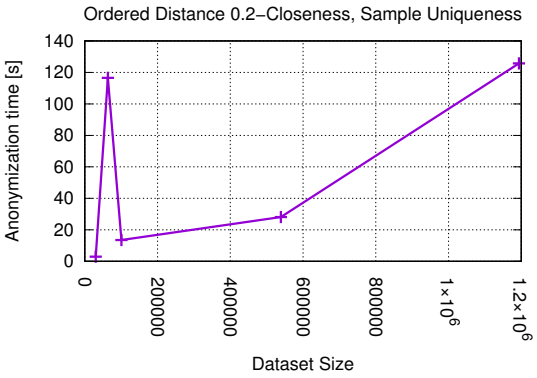

Hierarchical Distance 0.2−Closeness, Average Risk
Anonymization time [s]
Dataset Size

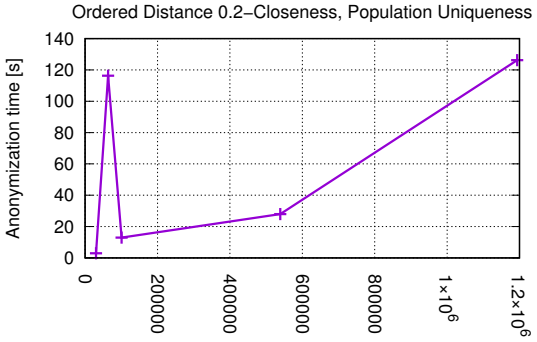

Ordered Distance 0.2−Closeness, Profitability Journalist
Anonymization time [s]
Dataset Size

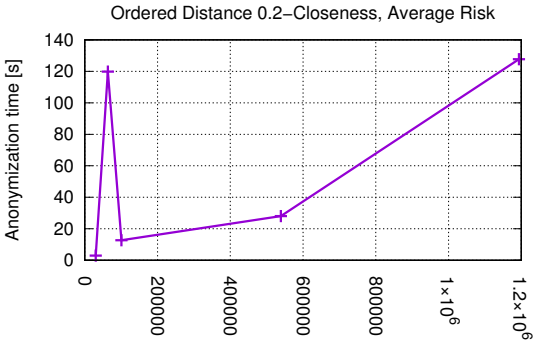

Ordered Distance 0.2−Closeness, Profitability JournalistNoAttack
Anonymization time [s]
Dataset Size

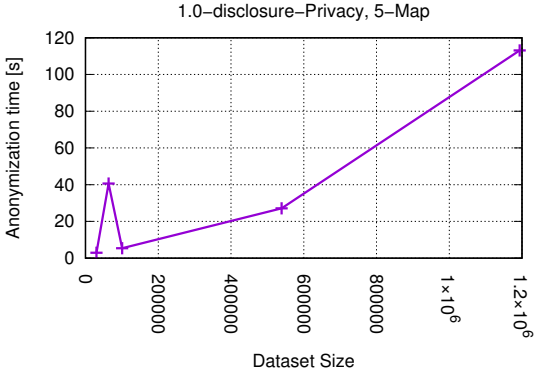

Ordered Distance 0.2−Closeness, Profitability Prosecutor
Anonymization time [s]
Dataset Size

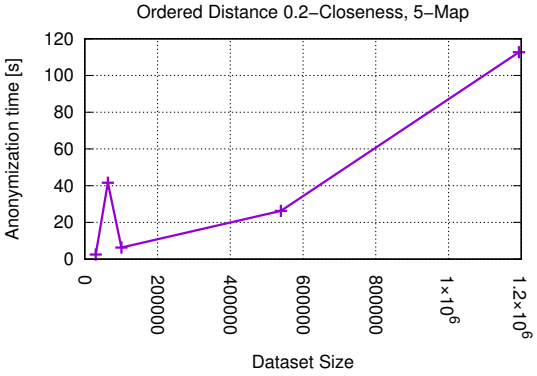

Ordered Distance 0.2−Closeness, Profitability ProsecutorNoAttack
Anonymization time [s]
Dataset Size


Ordered Distance 0.2−Closeness, Sample Uniqueness
Anonymization time [s]
Dataset Size


Ordered Distance 0.2−Closeness, Population Uniqueness
Anonymization time [s]
Dataset Size


Ordered Distance 0.2−Closeness, Average Risk
Anonymization time [s]
Dataset Size


1.0−disclosure−Privacy, 5−Map
Anonymization time [s]
Dataset Size


Ordered Distance 0.2−Closeness, 5−Map
Anonymization time [s]
Dataset Size

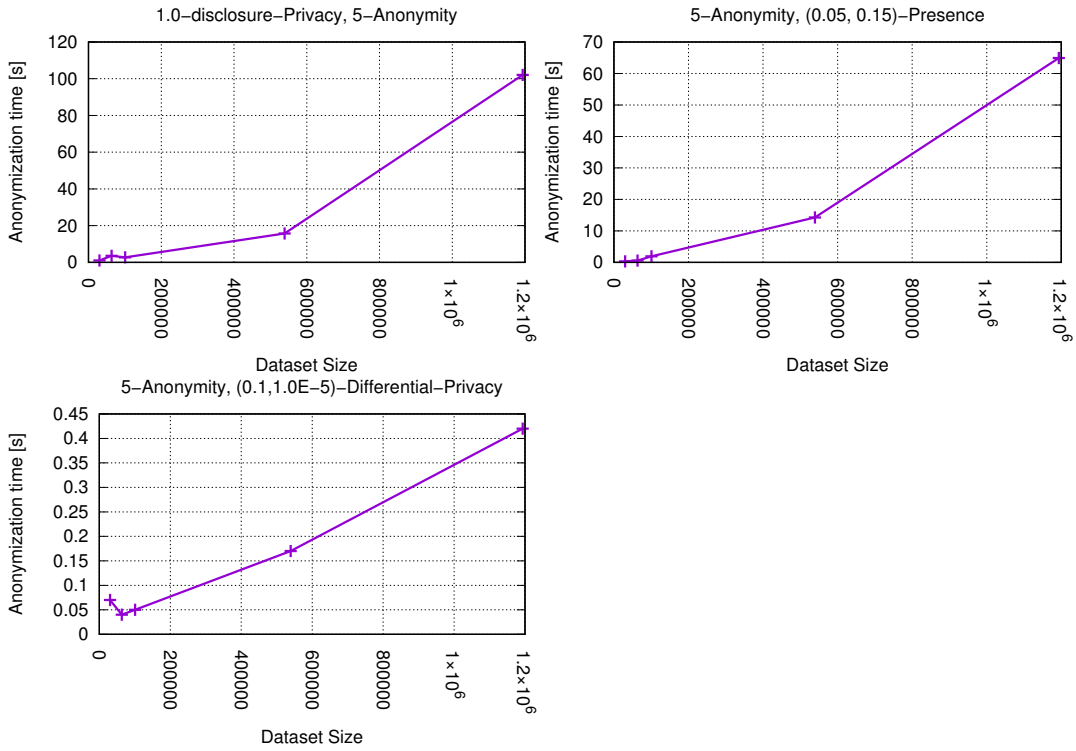


Figure 3.4: Anonymization times of LPL for PM Combinations that can not be substituted

3.3 Evaluation Results

3.3.1 Privacy Model Substitutions

In the work of Gerl et al., the *Privacy Model Substitution Table* is introduced. But only 3 out of the 21 privacy models from arx were taken into consideration[13]. Therefore the remaining substitutions can now be completed, with the resulting data from the first test run. Now we can not only substitute privacy model sets were one model covers all attacks of the other and more, but also decide for one privacy model, where the set of mitigated attacks is the same. Is this the case, we will decide for one privacy model based on their performance for small, medium and big data. In the following all substitutions will be described.

Some privacy model input sets can not be realized. For example, the privacy model of (e,d)-differential Privacy can not be combined with a subset. This prohibits us from combining it with the privacy models of d-Presence, Inclusion, Profitability Journalist, Profitability Journalist No Attack and k-Map. More inputs that are not possible are Ordered Distance t-Closeness with the other two form of t-Closeness as they differ in their semantic definition. Ordered Distance t-Closeness is used for numerical values, whereas the other two instances are used for categorical data only. Consequently they can not be combined. The last input set that con not be realized is the set of Inclusion

and d-Presence, because of the fact that Inclusion is an extension of d-Presence and arx does not support more than one instance of d-Presence[3].

With these out the way we can now focus on the remaining substitutions. Here we distinguish between two scenarios. In the first scenario, the input models are the same (e.g Input 1 = k_{i1}-Map, input 2 = k_{i2}-Map). Here the output model is the same as input 1 or input 2 (k_{r1}-Map). The privacy model attribute k_{r1} is calculated in this scenario as follows. We take the one value of the input set, which delivers the highest privacy restriction. For example, for $k_{i1} = 2$ and $k_{i2} = 4$, k_{i2} imposes the higher privacy on the dataset. Consequently, k_{r1} needs to be set to 4. This method is used for every input set, where input 1 equals input 2.
Now we will take a look at the second scenario, where the input models are not the same. That means that either the substitution can be done via their differentiation in mitigated attacks or by their performance. Those substitutions will be explained now.
For the input set of k_{i1}-Map and k_{i2}-Anonymity, we generate an output model of k_{r1}-Map as k-Map and k-Anonymity are directly related but k-Map can reduce the loss of information in comparison to k-Anonymity[11]. Consequently, the output attribute is the minimum of both input attribute values.
The next pair of input model we consider are k_{i1}-Map and Inclusion. The output model generated here is k_{r1}-Map as k-Map does mitigate Record Linkage attacks but Inclusion does not enforce any privacy guarantees[3, 11].
For k_{i1}-Map and Entropy-l_{i2}-Diversity, we determine Entropy-l_{r1}-Diversity as the output model, because l-Diversity mitigates not only Record Linkage attacks but also Attribute Linkage attacks. As k-Map and l-Diversity use similar definitions for their privacy, we can treat their attribute values as equivalent. That means that the resulting value of l_{r1} is the maximum of the two input model attributes k_{i1} and l_{i2}. The substitutions for k-Map with the other instances of l-Diversity is done the same.
As the four privacy models of profitability mitigate the same attack as k-Map and furthermore do not let the user specify privacy guarantees, it is recommended to use k_{r1}-Map as the output model for all four input pairs.
For the input of k_{i1}-Map and Average Risk (r_{i2}), we generate an output model of k_{r1}-Map as both models prevent from Record Linkage attacks but the performance of k-Map is better thean the performance of Average Risk. This can be seen in Figures 3.3m and 3.3a.
TODO: PpoUn, Sample Un
The remainder of input pairs containing k-Map can not be substituted due to the fact that these models mitigate a different set of attack types.

Now we will take a look at all input pairs containing k-Anonymity. As k-Anonymity is directly related to k-Map, the substitutions for this privacy model are more or less the same as for k-Map.
Next up are the input pairs, containing d-Disclosure-Privacy. The first input pair we will take a look at is the pair of d_{i1}-Disclosure Privacy and Inclusion. As the privacy model Inclusion does not mitigate any attacks, the output model here is d_{r1}-Disclosure Privacy.
For the input pair of d_{i1}-Disclosure Privacy and Basic-b_{i2}-Likeness, we get an output model of Basic-b_{r1}-Likeness. Although both models mitigate the same attacks and the performance is nearly identical, this substitution is due to the fact that b-Likeness pays attention to sensitive attribute values with lower frequency. This is not the case with d-Disclosure Privacy.
An equivalent substitution can be found for d_{i1}-Disclosure Privacy and Enhanced-b_{i2}-Likeness, as it is a more advanced instance of b-Likeness than its basic form.
For the input pairs of d_{i1}-Disclosure Privacy with the different forms of t_{i2}-Closeness(equal distance, hierarchical distance, ordered distance), we denote d_{r1}-Disclosure Privacy as the output model. The fact that d-disclosure privacy uses a multiplicative approach, which makes it stricter than t-closeness, justifies this assumption.
The rest of the input pair containing d-disclosure Privacy do not make it possible to determine a single output model, as they cover a different set of attacks. This means that both models have to be applied for the remaining input pairs.
Next up are all pairs containing the privacy model of Inclusion. As we remember from the definition in chapter 2.7.1, we know that this model does not enforce any privacy constraints upon data sets when applied. This means that no attacks can be mitigated and therefor it is never the output model. If we take the input set of Inclusion with k_{i2}-Anonymity, we denote k_{r1}-Anonymity as the output, where $k_{r1} = k_{i2}$. The remaining substitutions can be done equivalently.
The next set of input pairs we will take a look at, are the pairs containing d-Presence. There is only one privacy model, where a definite substitution can be done. This is the case with the input pair of -$_{i1}$-Presence and (e_{i2},d_{i2})-differential Privacy, as (e,d)-differential Privacy covers Table Linkage attacks just like d-Presence but also Probabilistic attacks. Consequently, we denote (e_{r1},d_{r1})-differential Privacy as the output model.
The rest of the input pairs that contain d-Presence can not be substituted, as like in previous cases, the sets of mitigated attacks do not overlap. Thus, both models have to be put as output model for the remainder of input pairs.
The next set of input pairs we will consider, are the pairs containing Entropy-

l-Diversity. The first input pair we will take a look at Entropy-l-Diversity combined with Recursive-(c,l)-Diversity. As both models are a form of l-Diversity, they mitigate the same attacks and can therefore be substituted by their performance. Because of the fact that Recursive-(c,l)-Diversity is slightly better in performance with slightly lower execution times, we define Recursive-(c,l)-Diversity as the output model for this substitution. The resulting privacy model attribute is the maximum of both l input values.
The same is the case for Entropy-l-Diversity and Distinct-l-Diversity. Distinct-l-Diversity performs much better than Entropy-l-Diversity and therefore it is denoted as the output model. The output model attribute will be the maximum of both input l values.
The next set of input models is Entropy-l-Diversity with Profitability Journalist. As the models of Profitability only mitigate Record Linkage Attacks and l-Diversity prevents Record Linkage as well as Attribute Linkage Attacks, we define Entropy-l-Diversity as the output model for the input sets of Entropy-l-Diversity with the four different instances of the Profitability privacy model. The same approach can be used with the input sets of Entropy-l-Diversity combined with the models of Population Uniqueness, Sample Uniqueness and Average Risk as all three of these models only prevent Record Linkage Attacks. Consequently, Entropy-l-Diversity is the output model for all these input sets. The rest of the input sets containing Entropy-l-Diversity can not be substituted due to the fact that input model 1 and input model 2 differ in the set of attacks they mitigate. Therefore, both model are put in the output model set.
The next input sets we will take a look at are the remaining input sets containing Recursive-(c,l)-Diversity. The first set is Recursive-(c,l)-Diversity combined with Distinct-l-Diversity. Here we can once more make a decision based on the performance of both models, as the mitigate the same attacks. As Distinct-l-Diversity performs significantly better, it will be denoted as the output model. The privacy model attribute will be the maximum of both input model attributes as both input models are instances of l-Diversity.
Similar to the substitutions containing Entropy-L-Diversity, when we combine Recursive-(c,l)-Diversity with the models of Profitability, we always denote Recursive-(c,l)-Diversity as the output model, because of the fact that Recursive-(c,l)-Diversity mitigates Record Linkage(just like Profitability) as well as Attribute Linkage attacks. The same fact causes us to define Recursive-(c,l)-Diversity the output model, when combined with Population Uniqueness, Sample Uniqueness and Average Risk.
The rest of the input model sets containing Recursive-(c,l)-Diversity can not be substituted, as the input models always differ in the mitigated attacks.
A similar approach can be used when looking at the remaining input model sets

containing Distinct-l-Diversity. Combined with the four instances of Profitability, Population Uniqueness, Sample Uniqueness or Average Risk, we denote Distinct-l-Diversity as the output model, since Distinct-l-Diversity mitigates Record Linkage as well as Attribute Linkage attacks.
The remainder of input sets, that contain Dinstinct-l-Diversity, can not be substituted. This is due to the fact that, once more, Dinstinct-l-Diversity and the second input model differ in their mitigated attacks.
Next up are the residual input sets containing (e,d)-Differential Privacy. When we take a look at the classification of (e,d)-Differential Privacy in section 2.9, we see that this model mitigates Table Linkage Attacks and Probabilistic Attacks. As we have no privacy models in the remaining input sets, that cover all attacks or more than (e,d)-Differential Privacy, but models where the set of mitigated attacks differs from (e,d)-Differential Privacy completely, no explicit substitution for these input sets is possible. Both models are denoted as the output.
The next set of input pairs we will deal with are the remaining pairs containing Basic-B-Likeness. When combined with Enhanced-B-Likeness, we can see that both models prevent the same types of attacks, as well as that both model perform in similar execution times. Consequently, we denote Enhanced-B-Likeness as the output model as this model is a more strict version of its predecessor.
The input pairs consisting of Basic-B-Likeness and the instances of t-Closeness can be treated as follows. As both privacy models mitigate the same set of attacks, we will take a look at their performance to define an unambiguous output model. Here we can see, that Basic-B-Likeness has similar execution times as the models of t-Closeness. We denote Basic-B-Likeness as the output model, because it pays attention to less frequent values of sensitive attributes as opposed to t-Closeness.
The rest of the input pairs with Basic-B-Likeness can not be substituted, because of the difference in mitigated attacks.
For the input pairs containing Enhanced-B-Likeness, the same substitutions can be defined as for its predecessor as it is a more advanced version with similar execution times. Consequently, when combined with the privacy models of t-Closeness, we denote Enhanced-B-Likeness as the output model. The rest of the privacy model pairs can not be substituted, just like with Basic-B-Likeness.
The next input pairs we will take into consideration are the residual pairs containing the instances of t-Closeness. We will start with Equal Distance t-Closeness. As Equal Distance t-Closeness and Ordered Distance t-Closeness differ in their semantic definition, they can not be combined, as mentioned

above. If we take a look at Equal Distance t-Closeness combined with Hierarchical Distance t-Closeness, we notice that both privacy model mitigate the same set of attacks as they both are instances of t-Closeness. Consequently, the decision to put Equal Distance t-Closeness as the output model is based on the observed performance for both models, as Equal Distance t-Closeness performs slightly better.
The rest of the input pairs with Equal Distance t-Closeness can once again not be substituted, because of the difference in attacks mitigated by them.
The next input pairs are the remaining pairs containing Ordered Distance t-Closeness. As mentioned above, it can not be combined with Hierarchical Distance t-Closeness. The remaining input model pairs differ in the set of prevented attacks and therefore can not be substituted. The same is the case for the last pairs of privacy models that contain Hierarchical Distance t-Closeness.
TODO: Profitbalilities subst
The result of these substitutions can be seen in the full Privacy Model Substitution Table in Figure 3.2.

Input 1	Input 2	Output	Privacy Model Attribute
k_{i1}-Map	k_{i2}-Map	k_{r1}-Map	$\{(k_{r1}, \max(k_{i1}, k_{i2}))\}$
k_{i1}-Map	k_{i2}-Anonymity	k_{r1}-Map	$\{(k_{r1}, \max(k_{i1}, k_{i2}))\}$
k_{i1}-Map	d_{i2}-Disclosure-Privacy	k_{r1}-Map, d_{r2}-Disclosure-Privacy	$\{(k_{r1}, k_{i1}), (d_{r2}, d_{i2})\}$
k_{i1}-Map	Inclusion	k_{r1}-Map	$\{(k_{r1}, k_{i1})\}$
k_{i1}-Map	(d_{i2min}, d_{i2max})-Presence	k_{r1}-Map, (d_{r2min}, d_{r2max})-Presence	$\{(k_{r1}, k_{i1}), (d_{r2min}, d_{i2min}), (d_{r2max}, d_{i2max})\}$
k_{i1}-Map	Entropy-l_{i2}-Diversity	Entropy-l_{r1}-Diversity	$\{(l_{r1}, \max(k_{i1}, l_{i2}))\}$
k_{i1}-Map	Recursive-$c_{i2}l_{i2}$-Diversity	Recursive-$c_{r1}l_{r1}$-Diversity	$\{(c_{r1}, c_{i2}), (l_{r1}, \max(k_{i1}, l_{i2}))\}$
k_{i1}-Map	Distinct-l_{i2}-Diversity	Distinct-l_{r1}-Diversity	$\{(l_{r1}, \max(k_{i1}, l_{i2}))\}$
k_{i1}-Map	(e_{i2},d_{i2})-differential-Privacy	k_{r1}-Map, (e_{r2},d_{r2})-differential-Privacy	$\{(k_{r1}, k_{i1}), (e_{r2}, e_{i2}), (d_{r2}, d_{i2})\}$
k_{i1}-Map	Basic-B_{i2}-Likeness	k_{r1}-Map, Basic-B_{r2}-Likeness	$\{(k_{r1}, k_{i1}), (B_{r2}, B_{i2})\}$
k_{i1}-Map	Enhanced-B_{i2}-Likeness	k_{r1}-Map, Enhanced-B_{r2}-Likeness	$\{(k_{r1}, k_{i1}), (B_{r2}, B_{i2})\}$
k_{i1}-Map	Equal-Distance-t_{i2}-Closeness	k_{r1}-Map, Equal-Distance-t_{r2}-Closeness	$\{(k_{r1}, k_{i1}), (t_{r2}, t_{i2})\}$
k_{i1}-Map	Ordered-Distance-t_{i2}-Closeness	k_{r1}-Map, Ordered-Distance-t_{r2}-Closeness	$\{(k_{r1}, k_{i1}), (t_{r2}, t_{i2})\}$
k_{i1}-Map	Hierarchical-Distance-t_{i2}-Closeness	k_{r1}-Map, Hierarchical-Distance-t_{r2}-Closeness	$\{(k_{r1}, k_{i1}), (t_{r2}, t_{i2})\}$
k_{i1}-Map	Profitability-Journalist	k_{r1}-Map	TBD

k_{i1}-Map	Profitability-Journalist-No-Attack	k_{r1}-Map	TBD
k_{i1}-Map	Profitability-Prosecutor	k_{r1}-Map	TBD
k_{i1}-Map	Profitability-Prosecutor-No-Attack	k_{r1}-Map	TBD
k_{i1}-Map	Population-Uniqueness (p_{i2})	k_{r1}-Map	TBD
k_{i1}-Map	Sample-Uniqueness (s_{i2})	k_{r1}-Map	TBD
k_{i1}-Map	Average Risk (r_{i2})	k_{r1}-Map	TBD
k_{i1}-Anonymity	k_{i2}-Anonymity	k_{r1}-Anonymity	$\{(k_{r1}, \max(k_{i1}, k_{i2}))\}$
k_{i1}-Anonymity	d_{i2}-Disclosure-Privacy	k_{r1}-Anonymity, d_{r2}-Disclosure-Privacy	$\{(k_{r1}, k_{i1}), (d_{r2}, d_{i2})\}$
k_{i1}-Anonymity	Inclusion	k_{r1}-Anonymity	$\{(k_{r1}, k_{i1})\}$
k_{i1}-Anonymity	(d_{i2min}, d_{i2max})-Presence	k_{r1}-Anonymity, (d_{r2min}, d_{r2max})-Presence	$\{(k_{r1}, k_{i1}), (d_{r2min}, d_{i2min}), (d_{r2max}, d_{i2max})\}$
k_{i1}-Anonymity	Entropy-l_{i2}-Diversity	Entropy-l_{r1}-Diversity	$\{(l_{r1}, \max(k_{i1}, l_{i2}))\}$
k_{i1}-Anonymity	Recursive-$c_{i2}l_{i2}$-Diversity	Recursive-$c_{r1}l_{r1}$-Diversity	$\{(c_{r1}, c_{i2}), (l_{r1}, \max(k_{i1}, l_{i2}))\}$
k_{i1}-Anonymity	Distinct-l_{i2}-Diversity	Distinct-l_{r1}-Diversity	$\{(l_{r1}, \max(k_{i1}, l_{i2}))\}$
k_{i1}-Anonymity	(e_{i2},d_{i2})-differential-Privacy	k_{r1}-Anonymity, (e_{r2},d_{r2})-differential-Privacy	$\{(k_{r1}, k_{i1}), (e_{r2}, e_{i2}), (d_{r2}, d_{i2})\}$
k_{i1}-Anonymity	Basic-B_{i2}-Likeness	k_{r1}-Anonymity, Basic-B_{r2}-Likeness	$\{(k_{r1}, k_{i1}), (B_{r2}, B_{i2})\}$
k_{i1}-Anonymity	Enhanced-B_{i2}-Likeness	k_{r1}-Anonymity, Enhanced-B_{r2}-Likeness	$\{(k_{r1}, k_{i1}), (B_{r2}, B_{i2})\}$

k_{i1}-Anonymity	Equal-Distance-t_{i2}-Closeness	k_{r1}-Anonymity, Equal-Distance-t_{r2}-Closeness	$\{(k_{r1}, k_{i1}), (t_{r2}, t_{i2})\}$
k_{i1}-Anonymity	Ordered-Distance-t_{i2}-Closeness	k_{r1}-Anonymity, Ordered-Distance-t_{r2}-Closeness	$\{(k_{r1}, k_{i1}), (t_{r2}, t_{i2})\}$
k_{i1}-Anonymity	Hierarchical-Distance-t_{i2}-Closeness	k_{r1}-Anonymity, Hierarchical-Distance-t_{r2}-Closeness	$\{(k_{r1}, k_{i1}), (t_{r2}, t_{i2})\}$
k_{i1}-Anonymity	Profitability-Journalist	k_{r1}-Anonymity	TBD
k_{i1}-Anonymity	Profitability-Journalist-No-Attack	k_{r1}-Anonymity	TBD
k_{i1}-Anonymity	Profitability-Prosecutor	k_{r1}-Anonymity	TBD
k_{i1}-Anonymity	Profitability-Prosecutor-No-Attack	k_{r1}-Anonymity	TBD
k_{i1}-Anonymity	Population-Uniqueness (p_{i2})	k_{r1}-Anonymity	TBD
k_{i1}-Anonymity	Sample-Uniqueness (s_{i2})	k_{r1}-Anonymity	TBD
k_{i1}-Anonymity	Average Risk (r_{i2})	k_{r1}-Anonymity	TBD
d_{i1}-Disclosure-Privacy	d_{i2}-Disclosure-Privacy	d_{r1}-Disclosure-Privacy	$\{(d_{r1}, \min(d_{i1}, d_{i2}))\}$
d_{i1}-Disclosure-Privacy	Inclusion	d_{r1}-Disclosure-Privacy	$\{(d_{r1}, d_{i1})\}$
d_{i1}-Disclosure-Privacy	(d_{i2min}, d_{i2max})-Presence	d_{r1}-Disclosure-Privacy, (d_{r2min}, d_{r2max})-Presence	$\{(d_{r1}, d_{i1}), (d_{r2min}, d_{i2min}), (d_{r2max}, d_{i2max})\}$

d_{i1}-Disclosure-Privacy	Entropy-l_{i2}-Diversity	d_{r1}-Disclosure-Privacy, Entropy-l_{r2}-Diversity	$\{(d_{r1}, d_{i1}), (l_{r2}, l_{i2})\}$
d_{i1}-Disclosure-Privacy	Recursive-$c_{i2}l_{i2}$-Diversity	d_{r1}-Disclosure-Privacy, Recursive-$c_{r2}l_{r2}$-Diversity	$\{(d_{r1}, d_{i1}), (c_{r2}, c_{i2}), (l_{r2}, l_{i2})\}$
d_{i1}-Disclosure-Privacy	Distinct-l_{i2}-Diversity	d_{r1}-Disclosure-Privacy, Distinct-l_{r2}-Diversity	$\{(d_{r1}, d_{i1}), (l_{r2}, l_{i2})\}$
d_{i1}-Disclosure-Privacy	(e_{i2},d_{i2})-differential-Privacy	d_{r1}-Disclosure-Privacy, (e_{r2},d_{r2})-differential-Privacy	$\{(d_{r1}, d_{i1}), (e_{r2}, e_{i2}), (d_{r2}, d_{i2})\}$
d_{i1}-Disclosure-Privacy	Basic-B_{i2}-Likeness	Basic-B_{r1}-Likeness	TBD
d_{i1}-Disclosure-Privacy	Enhanced-B_{i2}-Likeness	Enhanced-B_{r1}-Likeness	TBD
d_{i1}-Disclosure-Privacy	Equal-Distance-t_{i2}-Closeness	d_{r1}-Disclosure-Privacy	TBD
d_{i1}-Disclosure-Privacy	Ordered-Distance-t_{i2}-Closeness	d_{r1}-Disclosure-Privacy	TBD
d_{i1}-Disclosure-Privacy	Hierarchical-Distance-t_{i2}-Closeness	d_{r1}-Disclosure-Privacy	TBD
d_{i1}-Disclosure-Privacy	Profitability-Journalist	d_{r1}-Disclosure-Privacy, Profitability-Journalist	$\{(d_{r1}, d_{i1})\}$
d_{i1}-Disclosure-Privacy	Profitability-Journalist-No-Attack	d_{r1}-Disclosure-Privacy, Profitability-Journalist-No-Attack	$\{(d_{r1}, d_{i1})\}$

d_{i1}-Disclosure-Privacy	Profitability-Prosecutor	d_{r1}-Disclosure-Privacy, Profitability-Prosecutor	$\{(d_{r1}, d_{i1})\}$
d_{i1}-Disclosure-Privacy	Profitability-Prosecutor-No-Attack	d_{r1}-Disclosure-Privacy, Profitability-Prosecutor-No-Attack	$\{(d_{r1}, d_{i1})\}$
d_{i1}-Disclosure-Privacy	Population-Uniqueness (p_{i2})	d_{r1}-Disclosure-Privacy, Population-Uniqueness (p_{r2})	$\{(d_{r1}, d_{i1}), (p_{r2}, p_{i2})\}$
d_{i1}-Disclosure-Privacy	Sample-Uniqueness (s_{i2})	d_{r1}-Disclosure-Privacy, Sample-Uniqueness (s_{r2})	$\{(d_{r1}, d_{i1}), (s_{r2}, s_{i2})\}$
d_{i1}-Disclosure-Privacy	Average Risk (r_{i2})	d_{r1}-Disclosure-Privacy, Average Risk (r_{r2})	$\{(d_{r1}, d_{i1}), (r_{r2}, r_{i2})\}$
Inclusion	Inclusion	Inclusion	No Attribute
Inclusion	(d_{i2min}, d_{i2max})-Presence	(d_{r1min}, d_{r1max})-Presence	$\{(d_{r1min}, d_{i2min}), (d_{r1max}, d_{i2max})\}$
Inclusion	Entropy-l_{i2}-Diversity	Entropy-l_{r1}-Diversity	$\{(l_{r1}, l_{i2})\}$
Inclusion	Recursive-$c_{i2}l_{i2}$-Diversity	Recursive-$c_{r1}l_{r1}$-Diversity	$\{(l_{r1}, l_{i2})\}$
Inclusion	Distinct-l_{i2}-Diversity	Distinct-l_{r1}-Diversity	$\{(l_{r1}, l_{i2})\}$
Inclusion	(e_{i2},d_{i2})-differential-Privacy	(e_{r1},d_{r1})-differential-Privacy	$\{(e_{r1}, e_{i2}), (d_{r1}, d_{i2})\}$
Inclusion	Basic-B_{i2}-Likeness	Basic-B_{r1}-Likeness	$\{(B_{r1}, B_{i2})\}$
Inclusion	Enhanced-B_{i2}-Likeness	Enhanced-B_{r1}-Likeness	$\{(B_{r1}, B_{i2})\}$
Inclusion	Equal-Distance-t_{i2}-Closeness	Equal-Distance-t_{r1}-Closeness	$\{(t_{r1}, t_{i2})\}$

Inclusion	Ordered-Distance-t_{i2}-Closeness	Ordered-Distance-t_{r1}-Closeness	$\{(t_{r1}, t_{i2})\}$
Inclusion	Hierarchical-Distance-t_{i2}-Closeness	Hierarchical-Distance-t_{r1}-Closeness	$\{(t_{r1}, t_{i2})\}$
Inclusion	Profitability-Journalist	Profitability-Journalist	No Attribute
Inclusion	Profitability-Journalist-No-Attack	Profitability-Journalist-No-Attack	No Attribute
Inclusion	Profitability-Prosecutor	Profitability-Prosecutor	No Attribute
Inclusion	Profitability-Prosecutor-No-Attack	Profitability-Prosecutor-No-Attack	No Attribute
Inclusion	Population-Uniqueness (p_{i2})	Population-Uniqueness (p_{r1})	$\{(p_{r1}, p_{i2})\}$
Inclusion	Sample-Uniqueness (s_{i2})	Sample-Uniqueness (s_{r1})	$\{(s_{r1}, s_{i2})\}$
Inclusion	Average Risk (r_{i2})	Average Risk (r_{r1})	$\{(r_{r1}, r_{i2})\}$
(d_{i1min}, d_{i1max})-Presence	(d_{i2min}, d_{i2max})-Presence	(d_{r1min}, d_{r1max})-Presence	$\{(d_{r1min}, \max(d_{i1min}, d_{i2min}))$, $(d_{r1max}, \min(d_{i1max}, d_{i2max}))\}$
(d_{i1min}, d_{i1max})-Presence	Entropy-l_{i2}-Diversity	(d_{r1min}, d_{r1max})-Presence, Entropy-l_{r2}-Diversity	$\{(d_{r1min}, d_{i1min})$, (d_{r1max}, d_{i1max}), $(l_{r2}, l_{i2})\}$
(d_{i1min}, d_{i1max})-Presence	Recursive-$c_{i2}l_{i2}$-Diversity	(d_{r1min}, d_{r1max})-Presence, Recursive-$c_{r2}l_{r2}$-Diversity	$\{(d_{r1min}, d_{i1min})$, (d_{r1max}, d_{i1max}), $(l_{r2}, l_{i2})\}$

(d_{i1min}, d_{i1max})-Presence	Distinct-l_{i2}-Diversity	d_{r1}-Presence, Distinct-l_{r2}-Diversity	{(d_{r1min}, d_{i1min}), (d_{r1max}, d_{i1max}), (l_{r2}, l_{i2})}
(d_{i1min}, d_{i1max})-Presence	(e_{i2},d_{i2})-differential-Privacy	(e_{r1},d_{r1})-differential-Privacy	TBD
(d_{i1min}, d_{i1max})-Presence	Basic-B_{i2}-Likeness	(d_{r1min}, d_{r1max})-Presence, Basic-B_{r2}-Likeness	{(d_{r1min}, d_{i1min}), (d_{r1max}, d_{i1max}), (B_{r2}, B_{i2})}
(d_{i1min}, d_{i1max})-Presence	Enhanced-B_{i2}-Likeness	(d_{r1min}, d_{r1max})-Presence, Enhanced-B_{r2}-Likeness	{(d_{r1min}, d_{i1min}), (d_{r1max}, d_{i1max}), (B_{r2}, B_{i2})}
(d_{i1min}, d_{i1max})-Presence	Equal-Distance-t_{i2}-Closeness	(d_{r1min}, d_{r1max})-Presence, Equal-Distance-t_{r2}-Closeness	{(d_{r1min}, d_{i1min}), (d_{r1max}, d_{i1max}), (t_{r2}, t_{i2})}
(d_{i1min}, d_{i1max})-Presence	Ordered-Distance-t_{i2}-Closeness	(d_{r1min}, d_{r1max})-Presence, Ordered-Distance-t_{i2}-Closeness	{(d_{r1min}, d_{i1min}), (d_{r1max}, d_{i1max}), (t_{r2}, t_{i2})}
(d_{i1min}, d_{i1max})-Presence	Hierarchical-Distance-t_{i2}-Closeness	(d_{r1min}, d_{r1max})-Presence, Hierarchical-Distance-t_{r2}-Closeness	{(d_{r1min}, d_{i1min}), (d_{r1max}, d_{i1max}), (t_{r2}, t_{i2})}
(d_{i1min}, d_{i1max})-Presence	Profitability-Journalist	(d_{r1min}, d_{r1max})-Presence, Profitability-Journalist	{(d_{r1min}, d_{i1min}), (d_{r1max}, d_{i1max})}
(d_{i1min}, d_{i1max})-Presence	Profitability-Journalist-No-Attack	(d_{r1min}, d_{r1max})-Presence, Profitability-Journalist-No-Attack	{(d_{r1min}, d_{i1min}), (d_{r1max}, d_{i1max})}
(d_{i1min}, d_{i1max})-Presence	Profitability-Prosecutor	(d_{r1min}, d_{r1max})-Presence, Profitability-Prosecutor	{(d_{r1min}, d_{i1min}), (d_{r1max}, d_{i1max})}
(d_{i1min}, d_{i1max})-Presence	Profitability-Prosecutor-No-Attack	(d_{r1min}, d_{r1max})-Presence, Profitability-Prosecutor-No-Attack	{(d_{r1min}, d_{i1min}), (d_{r1max}, d_{i1max})}

(d_{i1min}, d_{i1max})-Presence	Population-Uniqueness (p_{i2})	(d_{r1min}, d_{r1max})-Presence, Population-Uniqueness (p_{r2})	{(d_{r1min}, d_{i1min}), (d_{r1max}, d_{i1max}), (p_{r2}, p_{i2})}
(d_{i1min}, d_{i1max})-Presence	Sample-Uniqueness (s_{i2})	(d_{r1min}, d_{r1max})-Presence, Sample-Uniqueness (s_{r2})	{(d_{r1min}, d_{i1min}), (d_{r1max}, d_{i1max}), (s_{r2}, s_{i2})}
(d_{i1min}, d_{i1max})-Presence	Average Risk (r_{i2})	(d_{r1min}, d_{r1max})-Presence, Average Risk (r_{r2})	{(d_{r1min}, d_{i1min}), (d_{r1max}, d_{i1max}), (p_{r2}, p_{i2})}
Entropy-l_{i1}-Diversity	Entropy-l_{i2}-Diversity	Entropy-l_{r1}-Diversity	{(l_{r1}, max(l_{i1}, l_{i2}))}
Entropy-l_{i1}-Diversity	Recursive-$c_{i2}l_{i2}$-Diversity	Recursive-$c_{r1}l_{r1}$-Diversity	{(c_{r1}, c_{i2}), (l_{r1}, max(l_{i1}, l_{i2}))}
Entropy-l_{i1}-Diversity	Distinct-l_{i2}-Diversity	Distinct-l_{r1}-Diversity	{(l_{r1}, max(l_{i1}, l_{i2}))}
Entropy-l_{i1}-Diversity	(e_{i2},d_{i2})-differential-Privacy	Entropy-l_{r1}-Diversity, (e_{r2},d_{r2})-differential-Privacy	{(l_{r1}, l_{i1}), (e_{r2}, e_{i2}), (d_{r2}, d_{i2})}
Entropy-l_{i1}-Diversity	Basic-B_{i2}-Likeness	Entropy-l_{r1}-Diversity, Basic-B_{r2}-Likeness	{(l_{r1}, l_{i1}), (B_{r2}, B_{i2})}
Entropy-l_{i1}-Diversity	Enhanced-B_{i2}-Likeness	Entropy-l_{r1}-Diversity, Enhanced-B_{r2}-Likeness	{(l_{r1}, l_{i1}), (B_{r2}, B_{i2})}
Entropy-l_{i1}-Diversity	Equal-Distance-t_{i2}-Closeness	Entropy-l_{r1}-Diversity, Equal-Distance-t_{r2}-Closeness	{(l_{r1}, l_{i1}), (t_{r2}, t_{i2})}
Entropy-l_{i1}-Diversity	Ordered-Distance-t_{i2}-Closeness	Entropy-l_{r1}-Diversity, Ordered-Distance-t_{r2}-Closeness	{(l_{r1}, l_{i1}), (t_{r2}, t_{i2})}
Entropy-l_{i1}-Diversity	Hierarchical-Distance-t_{i2}-Closeness	Entropy-l_{r1}-Diversity, Hierarchical-Distance-t_{r2}-Closeness	{(l_{r1}, l_{i1}), (t_{r2}, t_{i2})}

Entropy-l_{i1}-Diversity	Profitability-Journalist	Entropy-l_{r1}-Diversity	TBD
Entropy-l_{i1}-Diversity	Profitability-Journalist-No-Attack	Entropy-l_{r1}-Diversity	TBD
Entropy-l_{i1}-Diversity	Profitability-Prosecutor	Entropy-l_{r1}-Diversity	TBD
Entropy-l_{i1}-Diversity	Profitability-Prosecutor-No-Attack	Entropy-l_{r1}-Diversity	TBD
Entropy-l_{i1}-Diversity	Population-Uniqueness (p_{i2})	Entropy-l_{r1}-Diversity	TBD
Entropy-l_{i1}-Diversity	Sample-Uniqueness (s_{i2})	Entropy-l_{r1}-Diversity	TBD
Entropy-l_{i1}-Diversity	Average Risk (r_{i2})	Entropy-l_{r1}-Diversity	TBD
Recursive-$c_{i1}l_{i1}$-Diversity	Recursive-$c_{i2}l_{i2}$-Diversity	Recursive-$c_{r1}l_{r1}$-Diversity	$\{(c_{r1}, \min(c_{i1}, c_{i2})), (l_{r1}, \max(l_{i1}, l_{i2}))\}$
Recursive-$c_{i1}l_{i1}$-Diversity	Distinct-l_{i2}-Diversity	Distinct-l_{r1}-Diversity	$\{(l_{r1}, \max(l_{i1}, l_{i2}))\}$
Recursive-$c_{i1}l_{i1}$-Diversity	(e_{i2},d_{i2})-differential-Privacy	Recursive-$c_{r1}l_{r1}$-Diversity, (e_{r2},d_{r2})-differential-Privacy	$\{(c_{r1}, c_{i1}), (l_{r1}, l_{i1}), (e_{r2}, e_{i2}), (d_{r2}, d_{i2})\}$
Recursive-$c_{i1}l_{i1}$-Diversity	Basic-B_{i2}-Likeness	Recursive-$c_{r1}l_{r1}$-Diversity, Basic-B_{r2}-Likeness	$\{(c_{r1}, c_{i1}), (l_{r1}, l_{i1}), (B_{r2}, B_{i2})\}$
Recursive-$c_{i1}l_{i1}$-Diversity	Enhanced-B_{i2}-Likeness	Recursive-$c_{r1}l_{r1}$-Diversity, Enhanced-B_{r2}-Likeness	$\{(c_{r1}, c_{i1}), (l_{r1}, l_{i1}), (B_{r2}, B_{i2})\}$
Recursive-$c_{i1}l_{i1}$-Diversity	Equal-Distance-t_{i2}-Closeness	Recursive-$c_{r1}l_{r1}$-Diversity, Equal-Distance-t_{r2}-Closeness	$\{(c_{r1}, c_{i1}), (l_{r1}, l_{i1}), (t_{r2}, t_{i2})\}$

Recursive-$c_{i1}l_{i1}$-Diversity	Ordered-Distance-t_{i2}-Closeness	Recursive-$c_{r1}l_{r1}$-Diversity, Ordered-Distance-t_{r2}-Closeness	$\{(c_{r1}, c_{i1}), (l_{r1}, l_{i1}), (t_{r2}, t_{i2})\}$
Recursive-$c_{i1}l_{i1}$-Diversity	Hierarchical-Distance-t_{i2}-Closeness	Recursive-$c_{r1}l_{r1}$-Diversity, Hierarchical-Distance-t_{r2}-Closeness	$\{(c_{r1}, c_{i1}), (l_{r1}, l_{i1}), (t_{r2}, t_{i2})\}$
Recursive-$c_{i1}l_{i1}$-Diversity	Profitability-Journalist	Recursive-$c_{r1}l_{r1}$-Diversity	TBD
Recursive-$c_{i1}l_{i1}$-Diversity	Profitability-Journalist-No-Attack	Recursive-$c_{r1}l_{r1}$-Diversity	TBD
Recursive-$c_{i1}l_{i1}$-Diversity	Profitability-Prosecutor	Recursive-$c_{r1}l_{r1}$-Diversity	TBD
Recursive-$c_{i1}l_{i1}$-Diversity	Profitability-Prosecutor-No-Attack	Recursive-$c_{r1}l_{r1}$-Diversity	TBD
Recursive-$c_{i1}l_{i1}$-Diversity	Population-Uniqueness (p_{i2})	Recursive-$c_{r1}l_{r1}$-Diversity	TBD
Recursive-$c_{i1}l_{i1}$-Diversity	Sample-Uniqueness (s_{i2})	Recursive-$c_{r1}l_{r1}$-Diversity	TBD
Recursive-$c_{i1}l_{i1}$-Diversity	Average Risk (r_{i2})	Recursive-$c_{r1}l_{r1}$-Diversity	TBD
Distinct-l_{i1}-Diversity	Distinct-l_{i2}-Diversity	Distinct-l_{r1}-Diversity	$\{(l_{r1}, \max(l_{i1}, l_{i2}))\}$
Distinct-l_{i1}-Diversity	(e_{i2},d_{i2})-differential-Privacy	Distinct-l_{r1}-Diversity, (e_{r2},d_{r2})-differential-Privacy	$\{(l_{r1}, l_{i1}), (e_{r2}, e_{i2}), (d_{r2}, d_{i2})\}$
Distinct-l_{i1}-Diversity	Basic-B_{i2}-Likeness	Distinct-l_{r1}-Diversity, Basic-B_{r2}-Likeness	$\{(l_{r1}, l_{i1}), (B_{r2}, B_{i2})\}$
Distinct-l_{i1}-Diversity	Enhanced-B_{i2}-Likeness	Distinct-l_{r1}-Diversity, Enhanced-B_{r2}-Likeness	$\{(l_{r1}, l_{i1}), (B_{r2}, B_{i2})\}$

Distinct-l_{i1}-Diversity	Equal-Distance-t_{i2}-Closeness	Distinct-l_{r1}-Diversity, Equal-Distance-t_{r2}-Closeness	$\{(l_{r1}, l_{i1}), (t_{r2}, t_{i2})\}$
Distinct-l_{i1}-Diversity	Ordered-Distance-t_{i2}-Closeness	Distinct-l_{r1}-Diversity, Ordered-Distance-t_{r2}-Closeness	$\{(l_{r1}, l_{i1}), (t_{r2}, t_{i2})\}$
Distinct-l_{i1}-Diversity	Hierarchical-Distance-t_{i2}-Closeness	Distinct-l_{r1}-Diversity, Hierarchical-Distance-t_{r2}-Closeness	$\{(l_{r1}, l_{i1}), (t_{r2}, t_{i2})\}$
Distinct-l_{i1}-Diversity	Profitability-Journalist	Distinct-l_{r1}-Diversity	TBD
Distinct-l_{i1}-Diversity	Profitability-Journalist-No-Attack	Distinct-l_{r1}-Diversity	TBD
Distinct-l_{i1}-Diversity	Profitability-Prosecutor	Distinct-l_{r1}-Diversity	TBD
Distinct-l_{i1}-Diversity	Profitability-Prosecutor-No-Attack	Distinct-l_{r1}-Diversity	TBD
Distinct-l_{i1}-Diversity	Population-Uniqueness (p_{i2})	Distinct-l_{r1}-Diversity	TBD
Distinct-l_{i1}-Diversity	Sample-Uniqueness (s_{i2})	Distinct-l_{r1}-Diversity	TBD
Distinct-l_{i1}-Diversity	Average Risk (r_{i2})	Distinct-l_{r1}-Diversity	TBD
(e_{i1},d_{i1})-differential-Privacy	(e_{i2},d_{i2})-differential-Privacy	(e_{r1},d_{r1})-differential-Privacy	$\{(e_{r1}, \min(e_{i1}, e_{i2})), (d_{r1}, \min(d_{i1}, d_{i2}))\}$
(e_{i1},d_{i1})-differential-Privacy	Basic-B_{i2}-Likeness	(e_{r1},d_{r1})-differential-Privacy, Basic-B_{r2}-Likeness	$\{(e_{r1}, e_{i1}), (d_{r1}, d_{i1}), (B_{r2}, B_{i2})\}$

(e_{i1},d_{i1})-differential-Privacy	Enhanced-B_{i2}-Likeness	(e_{r1},d_{r1})-differential-Privacy, Enhanced-B_{r2}-Likeness	$\{(e_{r1}, e_{i1}), (d_{r1}, d_{i1}), (B_{r2}, B_{i2})\}$
(e_{i1},d_{i1})-differential-Privacy	Equal-Distance-t_{i2}-Closeness	(e_{r1},d_{r1})-differential-Privacy, Equal-Distance-t_{r2}-Closeness	$\{(e_{r1}, e_{i1}), (d_{r1}, d_{i1}), (t_{r2}, t_{i2})\}$
(e_{i1},d_{i1})-differential-Privacy	Ordered-Distance-t_{i2}-Closeness	(e_{r1},d_{r1})-differential-Privacy, Ordered-Distance-t_{r2}-Closeness	$\{(e_{r1}, e_{i1}), (d_{r1}, d_{i1}), (t_{r2}, t_{i2})\}$
(e_{i1},d_{i1})-differential-Privacy	Hierarchical-Distance-t_{i2}-Closeness	(e_{r1},d_{r1})-differential-Privacy, Hierarchical-Distance-t_{r2}-Closeness	$\{(e_{r1}, e_{i1}), (d_{r1}, d_{i1}), (t_{r2}, t_{i2})\}$
(e_{i1},d_{i1})-differential-Privacy	Profitability-Journalist	(e_{r1},d_{r1})-differential-Privacy, Profitability-Journalist	$\{(e_{r1}, e_{i1}), (d_{r1}, d_{i1})\}$
(e_{i1},d_{i1})-differential-Privacy	Profitability-Journalist-No-Attack	(e_{r1},d_{r1})-differential-Privacy, Profitability-Journalist-No-Attack	$\{(e_{r1}, e_{i1}), (d_{r1}, d_{i1})\}$
(e_{i1},d_{i1})-differential-Privacy	Profitability-Prosecutor	(e_{r1},d_{r1})-differential-Privacy, Profitability-Prosecutor	$\{(e_{r1}, e_{i1}), (d_{r1}, d_{i1})\}$
(e_{i1},d_{i1})-differential-Privacy	Profitability-Prosecutor-No-Attack	(e_{r1},d_{r1})-differential-Privacy, Profitability-Prosecutor-No-Attack	$\{(e_{r1}, e_{i1}), (d_{r1}, d_{i1})\}$
(e_{i1},d_{i1})-differential-Privacy	Population-Uniqueness (p_{i2})	(e_{r1},d_{r1})-differential-Privacy, Population-Uniqueness (p_{r2})	$\{(e_{r1}, e_{i1}), (d_{r1}, d_{i1}), (p_{r2}, p_{i2})\}$
(e_{i1},d_{i1})-differential-Privacy	Sample-Uniqueness (s_{i2})	(e_{r1},d_{r1})-differential-Privacy, Sample-Uniqueness (s_{r2})	$\{(e_{r1}, e_{i1}), (d_{r1}, d_{i1}), (s_{r2}, s_{i2})\}$

(e_{i1},d_{i1})-differential-Privacy	Average Risk (r_{i2})	(e_{r1},d_{r1})-differential-Privacy, Average Risk (r_{r2})	$\{(e_{r1}, e_{i1}), (d_{r1}, d_{i1}), (r_{r2}, r_{i2})\}$
Basic-B_{i1}-Likeness	Basic-B_{i2}-Likeness	Basic-B_{r1}-Likeness	$\{(B_{r1}, \min(B_{i1}, B_{i2}))\}$
Basic-B_{i1}-Likeness	Enhanced-B_{i2}-Likeness	Enhanced-B_{r1}-Likeness	$\{(B_{r1}, \min(B_{i1}, B_{i2}))\}$
Basic-B_{i1}-Likeness	Equal-Distance-t_{i2}-Closeness	Basic-B_{r1}-Likeness	TBD
Basic-B_{i1}-Likeness	Ordered-Distance-t_{i2}-Closeness	Basic-B_{r1}-Likeness	TBD
Basic-B_{i1}-Likeness	Hierarchical-Distance-t_{i2}-Closeness	Basic-B_{r1}-Likeness	TBD
Basic-B_{i1}-Likeness	Profitability-Journalist	Basic-B_{r1}-Likeness, Profitability-Journalist	$\{(B_{r1}, B_{i1})\}$
Basic-B_{i1}-Likeness	Profitability-Journalist-No-Attack	Basic-B_{r1}-Likeness, Profitability-Journalist-No-Attack	$\{(B_{r1}, B_{i1})\}$
Basic-B_{i1}-Likeness	Profitability-Prosecutor	Basic-B_{r1}-Likeness, Profitability-Prosecutor	$\{(B_{r1}, B_{i1})\}$
Basic-B_{i1}-Likeness	Profitability-Prosecutor-No-Attack	Basic-B_{r1}-Likeness, Profitability-Prosecutor-No-Attack	$\{(B_{r1}, B_{i1})\}$
Basic-B_{i1}-Likeness	Population-Uniqueness (p_{i2})	Basic-B_{r1}-Likeness, Population-Uniqueness (p_{r2})	$\{(B_{r1}, B_{i1}), (p_{r2}, p_{i2})\}$

Basic-B_{i1}-Likeness	Sample-Uniqueness (s_{i2})	Basic-B_{r1}-Likeness, Sample-Uniqueness (s_{r2})	$\{(B_{r1}, B_{i1}), (s_{r2}, s_{i2})\}$
Basic-B_{i1}-Likeness	Average Risk (r_{i2})	Basic-B_{r1}-Likeness, Average Risk (r_{r2})	$\{(B_{r1}, B_{i1}), (r_{r2}, r_{i2})\}$
Enhanced-B_{i1}-Likeness	Enhanced-B_{i2}-Likeness	Enhanced-B_{r1}-Likeness	$\{(B_{r1}, \min(B_{i1}, B_{i2}))\}$
Enhanced-B_{i1}-Likeness	Equal-Distance-t_{i2}-Closeness	Enhanced-B_{r1}-Likeness	TBD
Enhanced-B_{i1}-Likeness	Ordered-Distance-t_{i2}-Closeness	Enhanced-B_{r1}-Likeness	TBD
Enhanced-B_{i1}-Likeness	Hierarchical-Distance-t_{i2}-Closeness	Enhanced-B_{r1}-Likeness	TBD
Enhanced-B_{i1}-Likeness	Profitability-Journalist	Enhanced-B_{r1}-Likeness, Profitability-Journalist	$\{(B_{r1}, B_{i1})\}$
Enhanced-B_{i1}-Likeness	Profitability-Journalist-No-Attack	Enhanced-B_{r1}-Likeness, Profitability-Journalist-No-Attack	$\{(B_{r1}, B_{i1})\}$
Enhanced-B_{i1}-Likeness	Profitability-Prosecutor	Enhanced-B_{r1}-Likeness, Profitability-Prosecutor	$\{(B_{r1}, B_{i1})\}$
Enhanced-B_{i1}-Likeness	Profitability-Prosecutor-No-Attack	Enhanced-B_{r1}-Likeness, Profitability-Prosecutor-No-Attack	$\{(B_{r1}, B_{i1})\}$
Enhanced-B_{i1}-Likeness	Population-Uniqueness (p_{i2})	Enhanced-B_{r1}-Likeness, Population-Uniqueness (p_{r2})	$\{(B_{r1}, B_{i1}), (p_{r2}, p_{i2})\}$

Enhanced-B_{i1}-Likeness	Sample-Uniqueness (s_{i2})	Enhanced-B_{r1}-Likeness, Sample-Uniqueness (s_{r2})	$\{(B_{r1}, B_{i1}), (s_{r2}, s_{i2})\}$
Enhanced-B_{i1}-Likeness	Average Risk (r_{i2})	Enhanced-B_{r1}-Likeness, Average Risk (r_{r2})	$\{(B_{r1}, B_{i1}), (r_{r2}, r_{i2})\}$
Equal-Distance-t_{i1}-Closeness	Equal-Distance-t_{i2}-Closeness	Equal-Distance-t_{r1}-Closeness	$\{(t_{r1}, \min(t_{i1}, t_{i2}))\}$
Equal-Distance-t_{i1}-Closeness	Hierarchical-Distance-t_{i2}-Closeness	Equal-Distance-t_{r1}-Closeness	$\{(t_{r1}, \min(t_{i1}, t_{i2}))\}$
Equal-Distance-t_{i1}-Closeness	Profitability-Journalist	Equal-Distance-t_{r1}-Closeness, Profitability-Journalist	$\{(t_{r1}, t_{i1})\}$
Equal-Distance-t_{i1}-Closeness	Profitability-Journalist-No-Attack	Equal-Distance-t_{r1}-Closeness, Profitability-Journalist-No-Attack	$\{(t_{r1}, t_{i1})\}$
Equal-Distance-t_{i1}-Closeness	Profitability-Prosecutor	Equal-Distance-t_{r1}-Closeness, Profitability-Prosecutor	$\{(t_{r1}, t_{i1})\}$
Equal-Distance-t_{i1}-Closeness	Profitability-Prosecutor-No-Attack	Equal-Distance-t_{r1}-Closeness, Profitability-Prosecutor-No-Attack	$\{(t_{r1}, t_{i1})\}$
Equal-Distance-t_{i1}-Closeness	Population-Uniqueness (p_{i2})	Equal-Distance-t_{r1}-Closeness, Population-Uniqueness (p_{i2})	$\{(t_{r1}, t_{i1}), (p_{r2}, p_{i2})\}$
Equal-Distance-t_{i1}-Closeness	Sample-Uniqueness (s_{i2})	Equal-Distance-t_{r1}-Closeness, Sample-Uniqueness (s_{i2})	$\{(t_{r1}, t_{i1}), (s_{r2}, s_{i2})\}$

Equal-Distance-t_{i1}-Closeness	Average Risk (r_{i2})	Equal-Distance-t_{r1}-Closeness, Average Risk (r_{i2})	$\{(t_{r1}, t_{i1}), (r_{r2}, r_{i2})\}$
Ordered-Distance-t_{i1}-Closeness	Ordered-Distance-t_{i2}-Closeness	Ordered-Distance-t_{r1}-Closeness	$\{(t_{r1}, \min(t_{i1}, t_{i2}))\}$
Ordered-Distance-t_{i1}-Closeness	Profitability-Journalist	Ordered-Distance-t_{r1}-Closeness, Profitability-Journalist	$\{(t_{r1}, t_{i1})\}$
Ordered-Distance-t_{i1}-Closeness	Profitability-Journalist-No-Attack	Ordered-Distance-t_{r1}-Closeness, Profitability-Journalist-No-Attack	$\{(t_{r1}, t_{i1})\}$
Ordered-Distance-t_{i1}-Closeness	Profitability-Prosecutor	Ordered-Distance-t_{r1}-Closeness, Profitability-Prosecutor	$\{(t_{r1}, t_{i1})\}$
Ordered-Distance-t_{i1}-Closeness	Profitability-Prosecutor-No-Attack	Ordered-Distance-t_{r1}-Closeness, Profitability-Prosecutor-No-Attack	$\{(t_{r1}, t_{i1})\}$
Ordered-Distance-t_{i1}-Closeness	Population-Uniqueness (p_{i2})	Ordered-Distance-t_{r1}-Closeness, Population-Uniqueness (p_{r2})	$\{(t_{r1}, t_{i1}), (p_{r2}, p_{i2})\}$
Ordered-Distance-t_{i1}-Closeness	Sample-Uniqueness (s_{i2})	Ordered-Distance-t_{r1}-Closeness, Sample-Uniqueness (s_{r2})	$\{(t_{r1}, t_{i1}), (s_{r2}, s_{i2})\}$
Ordered-Distance-t_{i1}-Closeness	Average Risk (r_{i2})	Ordered-Distance-t_{r1}-Closeness, Average Risk (r_{r2})	$\{(t_{r1}, t_{i1}), (r_{r2}, r_{i2})\}$
Hierarchical-Distance-t_{i1}-Closeness	Hierarchical-Distance-t_{i2}-Closeness	Hierarchical-Distance-t_{r1}-Closeness	$\{(t_{r1}, \min(t_{i1}, t_{i2}))\}$

Hierarchical-Distance-t_{i1}-Closeness	Profitability-Journalist	Hierarchical-Distance-t_{r1}-Closeness, Profitability-Journalist	$\{(t_{r1}, t_{i1})\}$
Hierarchical-Distance-t_{i1}-Closeness	Profitability-Journalist-No-Attack	Hierarchical-Distance-t_{r1}-Closeness, Profitability-Journalist-No-Attack	$\{(t_{r1}, t_{i1})\}$
Hierarchical-Distance-t_{i1}-Closeness	Profitability-Prosecutor	Hierarchical-Distance-t_{r1}-Closeness, Profitability-Prosecutor	$\{(t_{r1}, t_{i1})\}$
Hierarchical-Distance-t_{i1}-Closeness	Profitability-Prosecutor-No-Attack	Hierarchical-Distance-t_{r1}-Closeness, Profitability-Prosecutor-No-Attack	$\{(t_{r1}, t_{i1})\}$
Hierarchical-Distance-t_{i1}-Closeness	Population-Uniqueness (p_{i2})	Hierarchical-Distance-t_{r1}-Closeness, Population-Uniqueness (p_{r2})	$\{(t_{r1}, t_{i1}), (p_{r2}, p_{i2})\}$
Hierarchical-Distance-t_{i1}-Closeness	Sample-Uniqueness (s_{i2})	Hierarchical-Distance-t_{r1}-Closeness, Sample-Uniqueness (s_{r2})	$\{(t_{r1}, t_{i1}), (s_{r2}, s_{i2})\}$
Hierarchical-Distance-t_{i1}-Closeness	Average Risk (r_{i2})	Hierarchical-Distance-t_{r1}-Closeness, Average Risk (r_{r2})	$\{(t_{r1}, t_{i1}), (r_{r2}, r_{i2})\}$
Profitability-Journalist	Profitability-Journalist	Profitability-Journalist	No Attribute
Profitability-Journalist	Profitability-Journalist-No-Attack	Profitability-Journalist-No-Attack	No Attribute
Profitability-Journalist	Profitability-Prosecutor	Profitability-Journalist	No Attribute
Profitability-Journalist	Profitability-Prosecutor-No-Attack	Profitability-Prosecutor-No-Attack	No Attribute

Profitability-Journalist	Population-Uniqueness (p_{i2})	Profitability-Journalist	No Attribute
Profitability-Journalist	Sample-Uniqueness (s_{i2})	Profitability-Journalist	No Attribute
Profitability-Journalist	Average Risk (r_{i2})	Profitability-Journalist	No Attribute
Profitability-Journalist-No-Attack	Profitability-Journalist-No-Attack	Profitability-Journalist-No-Attack	No Attribute
Profitability-Journalist-No-Attack	Profitability-Prosecutor	Profitability-Journalist-No-Attack	No Attribute
Profitability-Journalist-No-Attack	Profitability-Prosecutor-No-Attack	Profitability-Journalist-No-Attack	No Attribute
Profitability-Journalist-No-Attack	Population-Uniqueness (p_{i2})	Profitability-Journalist-No-Attack	No Attribute
Profitability-Journalist-No-Attack	Sample-Uniqueness (s_{i2})	Profitability-Journalist-No-Attack	No Attribute
Profitability-Journalist-No-Attack	Average Risk (r_{i2})	Profitability-Journalist-No-Attack	No Attribute
Profitability-Prosecutor	Profitability-Prosecutor	Profitability-Prosecutor	No Attribute
Profitability-Prosecutor	Profitability-Prosecutor-No-Attack	Profitability-Prosecutor-No-Attack	No Attribute
Profitability-Prosecutor	Population-Uniqueness (p_{i2})	Profitability-Prosecutor	No Attribute

Profitability-Prosecutor	Sample-Uniqueness (s_{i2})	Profitability-Prosecutor	No Attribute
Profitability-Prosecutor	Average Risk (r_{i2})	Profitability-Prosecutor	No Attribute
Profitability-Prosecutor-No-Attack	Profitability-Prosecutor-No-Attack	Profitability-Prosecutor-No-Attack	No Attribute
Profitability-Prosecutor-No-Attack	Population-Uniqueness (p_{i2})	Profitability-Prosecutor-No-Attack	No Attribute
Profitability-Prosecutor-No-Attack	Sample-Uniqueness (s_{i2})	Profitability-Prosecutor-No-Attack	No Attribute
Profitability-Prosecutor-No-Attack	Average Risk (r_{i2})	Profitability-Prosecutor-No-Attack	No Attribute
Population-Uniqueness (p_{i1})	Population-Uniqueness (p_{i2})	Population-Uniqueness (p_{r1})	$\{(p_{r1}, \min(p_{i1}, p_{i2}))\}$
Population-Uniqueness (p_{i1})	Sample-Uniqueness (s_{i2})	Sample-Uniqueness (s_{r1})	$\{(s_{r1}, \min(p_{i1}, s_{i2}))\}$
Population-Uniqueness (p_{i1})	Average Risk (r_{i2})	Average Risk (r_{r1})	TBD
Sample-Uniqueness (s_{i1})	Sample-Uniqueness (s_{i2})	Sample-Uniqueness (s_{r1})	$\{(s_{r1}, \min(s_{i1}, s_{i2}))\}$
Sample-Uniqueness (s_{i1})	Average Risk (r_{i2})	Sample-Uniqueness (s_{r1})	TBD
Average Risk (r_{i1})	Average Risk (r_{i2})	Average Risk (r_{r1})	$\{(r_{r1}, \min(r_{i1}, r_{i2}))\}$

Table 3.2: Privacy Model Substitution Table

3.3.2 Comparison of Performance

Vergleich table und dumme table

3.3.3 Privacy Model Combinations

Result of the tests –> empfehlungsmatrix

4 Related Work

Arx + PrivacyModels papers

5 Conclusion

6 Bibliography

[1] arx-deidentifier/arx. https://github.com/arx-deidentifier/arx/tree/master/src/main/org/deidentifier/arx/criteria. Last accessed 15 October 2018.

[2] Privacy models. https://arx.deidentifier.org/overview/privacy-criteria/. Last accessed 15 October 2018.

[3] arx deidentifier. arx-deidentifier/arx. https://github.com/arx-deidentifier/arx/blob/master/src/main/org/deidentifier/arx/criteria/Inclusion.java. Last accessed 22 October 2018.

[4] arx deidentifier. arx-deidentifier/arx. https://github.com/arx-deidentifier/arx/blob/master/src/main/org/deidentifier/arx/criteria/SampleUniqueness.java Last accessed 15 October 2018.

[5] Raffael Bild, Klaus A. Kuhn, and Fabian Prasser. SafePub: A truthful data anonymization algorithm with strong privacy guarantees. *Proceedings on Privacy Enhancing Technologies*, 2018(1):67–87, jan 2018.

[6] Justin Brickell and Vitaly Shmatikov. The cost of privacy. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08*. ACM Press, 2008.

[7] Jianneng Cao and Panagiotis Karras. Publishing microdata with a robust privacy guarantee. *Proceedings of the VLDB Endowment*, 5(11):1388–1399, jul 2012.

[8] Fida Kamal Dankar, Khaled El Emam, Angelica Neisa, and Tyson Roffey. Estimating the re-identification risk of clinical data sets. *BMC Medical Informatics and Decision Making*, 12(1), jul 2012.

[9] Damien Desfontaines. k-map, the weird cousin of k-anonymity - ted is writing things. https://desfontain.es/privacy/k-map.html. Last accessed 15 October 2018.

[10] Cynthia Dwork. Differential privacy. In *Automata, Languages and Programming*, pages 1–12. Springer Berlin Heidelberg, 2006.

[11] Khaled El Emam and Fida Kamal Dankar. Protecting privacy using k-anonymity. *Journal of the American Medical Informatics Association*, 15(5):627–637, sep 2008.

[12] Benjamin C. M. Fung, Ke Wang, Rui Chen, and Philip S. Yu. Privacy-preserving data publishing. *ACM Computing Surveys*, 42(4):1–53, jun 2010.

[13] Armin Gerl, Nadia Bennani, Harald Kosch, and Lionel Brunie. LPL, towards a GDPR-compliant privacy language: Formal definition and usage. In *Lecture Notes in Computer Science*, pages 41–80. Springer Berlin Heidelberg, 2018.

[14] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *2007 IEEE 23rd International Conference on Data Engineering*. IEEE, apr 2007.

[15] Ninghui Li, Wahbeh Qardaji, and Dong Su. On sampling, anonymization, and differential privacy or,k-anonymization meets differential privacy. In *Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security - ASIACCS '12*. ACM Press, 2012.

[16] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam. L-diversity: privacy beyond k-anonymity. In *22nd International Conference on Data Engineering (ICDE'06)*. IEEE, 2006.

[17] Mehmet Ercan Nergiz, Maurizio Atzori, and Chris Clifton. Hiding the presence of individuals from shared databases. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data - SIGMOD '07*. ACM Press, 2007.

[18] Fabian Prasser, James Gaupp, Zhiyu Wan, Weiyi Xia, Yevgeniy Vorobeychik, Murat Kantarcioglu, Klaus Kuhn, and Brad Malin. An open source tool for game theoretic health data de-identification. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2017:1430–1439, 2017.

[19] Latanya Sweeney. k-anonymity: a model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, oct 2002.

[20] Zhiyu Wan, Yevgeniy Vorobeychik, Weiyi Xia, Ellen Wright Clayton, Murat Kantarcioglu, Ranjit Ganta, Raymond Heatherly, and Bradley A. Malin. A game theoretic framework for analyzing re-identification risk. 10:e0120592.

Erklärung zur Bachelorarbeit

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Passau, den <date>

<Fabian, Pfeil>