Towards Explaining Epsilon:
A Worst-Case Study of Differential Privacy Risks

Luise Mehner
Technische Universität Berlin
mehner@campus.tu-berlin.de

Saskia Nuñez von Voigt
Technische Universität Berlin
saskia.nunezvonvoigt@tu-berlin.de

Florian Tschorsch
Technische Universität Berlin
florian.tschorsch@tu-berlin.de

Abstract—**Differential privacy is a concept to quantify the disclosure of private information that is controlled by the privacy parameter ε. However, an intuitive interpretation of ε is needed to explain the privacy loss to data engineers and data subjects. In this paper, we conduct a worst-case study of differential privacy risks. We generalize an existing model and reduce complexity to provide more understandable statements on the privacy loss. To this end, we analyze the impact of parameters and introduce the notion of a global privacy risk and global privacy leak.**

Index Terms—**differential privacy, ε, privacy risk**

1. Introduction

Differential privacy [1] is a concept that quantifies the risk for data subjects to be identified in a data set. More specifically, a mechanism is ε-differentially private if for two neighboring statistical databases, the ratio of the probabilities of every result is at most e^{ε}. The parameter ε controls the degree of privacy, i.e., the privacy loss.

However, there is often a lack of understanding of the privacy guarantees that an ε provides [2]. The privacy risk depends on the statistical function, the number of data subjects, and the contributed data itself. It is therefore difficult to determine and/or to communicate the privacy loss adequately in advance. Accordingly, one of the main barriers to differential privacy is related to communication issues between data engineers and data subjects [2].

In this paper, we provide an explanation of ε that quantifies the risk of identification and is independent of any other variables. To this end, we adopt the model of Lee and Clifton [3], which rephrases ε as a probability depending on the number of data subjects and the sensitivity. By considering the worst-case privacy risk, we get rid of this dependency. Therefore, we generalize the model and introduce the notion of a *global privacy risk* and a *global privacy leak*, which are global upper bounds.

With our work, we intend to contribute to a more comprehensible explanation of an ε-differentially private mechanism by reducing complexity. To this end, we discuss how our approach can be used to communicate privacy risks to data subjects and at the same time support data engineers to determine reasonable values for ε in advance. Furthermore, we apply the privacy risk to the randomized response technique [4]. In this way, we provide an intuitive narrative to communicate the privacy risk to data subjects.

2. Preliminaries

Differential privacy [1] is a mathematical notion that bounds the risk of being identified in a database.

Definition 1 (Differential Privacy). *A mechanism M gives ε-differential privacy if for all data sets X and X' differing in one entry, all outputs $S \subseteq Range(M)$ satisfy $P[M(X) \in S] \leq \mathrm{e}^{\varepsilon} \cdot P[M(X') \in S]$.*

The definition for ε-differential privacy states that the probabilities of any possible output of an ε-differentially private mechanism do not differ by more than a factor of e^{ε}. The parameter ε thus captures the privacy loss. If ε is small enough, an adversary will not have certainty about whether an output was computed over database X or database X'. As a consequence, the adversary cannot be certain about the presence or absence of any particular data subject. This guarantee holds even for strong adversaries, who know the data of all possible data subjects.

An important notion when dealing with differential privacy is the sensitivity of a function [1]. The sensitivity of a function f quantifies the maximum change a single entry could cause on the function's result.

Definition 2 (Sensitivity). *For $f : D^n \longrightarrow R^n$, the L1-sensitivity of f is $\Delta f = max_{(X,X')}|f(X) - f(X')|_1$ for all X, X' differing in one entry.*

The most commonly used method to achieve differential privacy is to add random noise Y to the function's result, drawn from a Laplace distribution $Lap(\lambda)$ [5].

Definition 3 (Laplace mechanism). *For all f: $D \rightarrow R$, the following mechanism is ε-differentially private: $M(X) = t + Y$ where Y is drawn from $Lap\,(\Delta f/\varepsilon)$.*

The guarantees differential privacy provides are of probabilistic nature. Thus, an inevitable privacy risk remains that boils down to the re-identification of data subjects in the database.

Lee and Clifton [3] estimate the success probability of an adversary guessing the correct combination of data subjects in a data set based on the output of a differentially private mechanism. They assume that the mechanism adds Laplace noise as introduced before. Lee and Clifton refer to the privacy risk ρ, i.e., the probability of being identified as present/absent in a database, as

$$\rho \leq \frac{1}{1 + (n-1)e^{-\varepsilon \Delta v/\Delta f}} \qquad (1)$$

where n is the number of data subjects, Δf the sensitivity of the function, and Δv the maximum change one of the data entries could cause on the function's result, i.e., the local sensitivity.

Running Example. Consider a school survey on drug abuse. To raise awareness, parents have access to the ε-differentially private results. Statistics such as the average age or the number of drug-using students per class can be obtained.

Bob's mother, Eve, finds out that there is high drug use in her son's class. She wants to know who is using drugs. Eve queries the database for the average age of drug-addicted students of Bob's class, which returns an ε-differentially private answer. Let us assume that the age of the students is between 0 and 25 years and that each class has at least one student who is recorded in the database. Hence, the sensitivity is given by $\Delta f = (25-0)/1 = 25$, i.e., if Bob is 25 years old, he would increase or decrease the average by 25. However, students of the same class are typically the same age. For example, Eve knows that there are a total of 21 students ($n = 21$) in Bob's class, aged between 14 and 18. Additionally, with respect to the privacy risk ρ defined by Lee and Clifton, we assume that only one student is not present in the database. Note that this corresponds to the worst case, since the number of combinations of possible students present is reduced to $n-1$. Accordingly, the local sensitivity yields $\Delta v = (18-14)/20 = 0.2$. Finally, assume the mechanism uses $\varepsilon = \ln 3$. Hence, Eve's probability of finding out which of Bob's classmates use drugs yields $\rho \approx 5\,\%$.

3. Worst-Case Privacy Risks

The privacy risk of Lee and Clifton depends on the number of data subjects n, the ratio of sensitivities $r = \Delta v/\Delta f$, and the privacy loss parameter ε. Often n and r are not known in advance, which makes it difficult to assess the risks. In the following, we analyze the impact of the parameters to provide generalized statements on the privacy risk. Therefore, we consider the worst case, i.e., we determine the global values for r and n.

3.1. Global Sensitivity Ratio

The global sensitivity ratio, i.e., the worst-case value for r, holds the highest impact of a data subject's value. That is, due to the underlying data one data subject has a higher impact on the result and thus increases the probability to identify someone's presence. In this case, the maximum change caused by one of the present data subjects corresponds to the sensitivity of the function, i.e., $\Delta v = \Delta f$. The worst-case ratio therefore yields $r \leq 1$ for all query functions. Consequently, the maximum privacy risk ρ depends on n and ε, given by

$$\rho \leq \frac{1}{1+(n-1)e^{-\varepsilon}}. \qquad (2)$$

For our running example, this means that Bob's age has the maximum possible impact on the mean, i.e., he is 25 years old and the database contains only one person of his class. In this case, Eve would choose the correct present students and thus finding out who is using drugs with 11 % chance for $\varepsilon = \ln 3$ and $n = 21$.

3.2. Global Number of Data Subjects

In our running example we assume that only one student of Bob's class is missing in the database. From this information alone, Eve can randomly guess which students are in the database. We denote to this probability as P_{guess}. If the probability of being in the database is the same for all students, then $P_{guess} = 1/n$. This probability is independent of the released result.

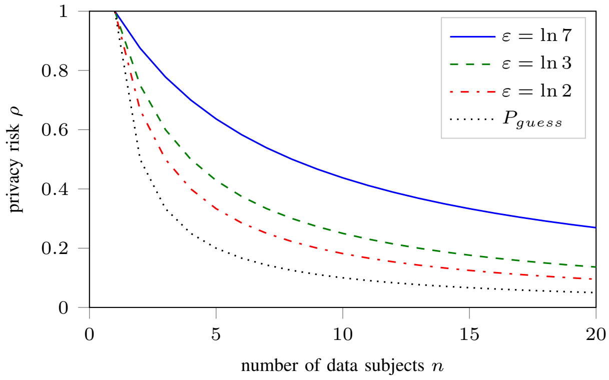

Figure 1. Privacy risk in dependence of the number of n and ε.

In order to illustrate the impact of the number of data subjects n, when releasing the result, we plot the privacy risk ρ from Equation (2) in Figure 1 for varying ε. For larger n, the privacy risk decreases with decreasing impact of larger n (cf. gradient). Hence, a small n has a higher impact on ρ and ε than a large n.

For $n = 1$, the maximum privacy risk yields 1, independent of ε. This makes sense as an adversary will always choose the correct combination of data subjects if there is only one possible combination to choose from. In other words, $P_{guess} = 1$ and the release of the result has no further impact. As a consequence, the worst case, where differential privacy still has an impact but success probabilities are maximized, is given for $n = 2$.

Considering our example under this background, assume that Eve is certain about the presence of all the students in Bob's class except for 2. As there is one student of the class missing from the database, the probability of Eve identifying the correct combination of students in the database is 0.5 before obtaining the average age. With the differentially private ($\varepsilon = \ln 3$) average age the probability, i.e., Eve's chance of inferring the students using drugs, increases to 75 %. This privacy risk is the worst-case privacy risk and depends on ε only. In the following, we generalize this observation as the global privacy risk.

3.3. Global Privacy Risk

We introduce the *global privacy risk* P as a global upper bound on the maximum privacy risk ρ with $r = 1$ and $n = 2$. The global privacy risk is then given by

$$P = \frac{1}{1+e^{-\varepsilon}}. \qquad (3)$$

In Figure 2, we illustrate the dependency of P and ε. The baseline of $P_{guess} = 1/2$ indicates that with $n = 2$ the adversary has a 50-50 chance of guessing the correct data subjects in a database. With increasing ε, the global privacy risk rises steadily and approaches 100 % without reaching it. Yet, a large ε helps Eve to infer who is using drugs with higher probability.

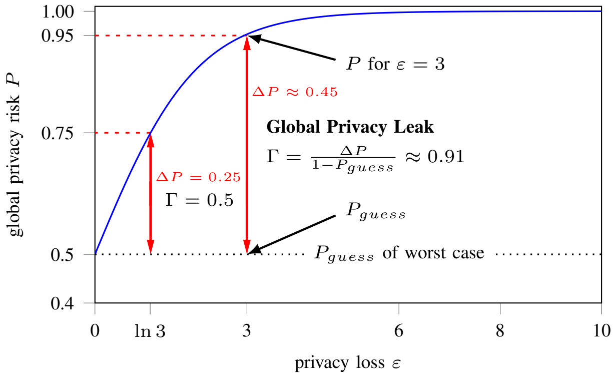

Figure 2. The worst-case privacy risk dependent on ε.

3.4. Global Privacy Leak

The privacy risk indicates the success probability. However, the privacy risk should be considered in relation to P_{guess} to determine the impact of a differentially private outcome on an adversary's success probability. That is, the privacy leak is the increment of the privacy risk caused by the release of an ε-differentially private result, i.e., $\Delta\rho = \rho - P_{guess}$.

Since the privacy leak is an increment to the guessing probability, it is not very intuitive. We therefore suggest to consider the privacy leak as a relative increase by scaling it to a range from 0 to 1. Analogously to the global privacy risk, the maximum relative increase is given for $r = 1$ and $n = 2$. We therefore introduce the *global privacy leak* Γ as

$$\Gamma = \frac{\Delta P}{1 - P_{guess}} \tag{4}$$

where $\Delta P = P - 1/2$. Additionally, we introduce local privacy leak γ, which is defined analogously with ρ instead of P.

In Figure 2, we visualize the global privacy leak. For $n = 2$, an adversary's probability of guessing correctly is $P_{guess} = 0.5$ but increases due to the answer to a query. The increment ΔP is accordingly given by the difference between P and P_{guess}. Large ε reveal more about someone's presence with the release than small ε values. Generally speaking, the privacy leak can be controlled by ε. For instance, for $\varepsilon = 3$ the global privacy leak $\Gamma = 0.91$ compared to $\varepsilon = \ln 3$ where $\Gamma = 0.5$.

In our running example, Eve finds out who is using drugs with a chance of 75 %. The global privacy leak yields $\Gamma = 0.5$ and captures Eve's information gain by taking the result of the differentially private query result into consideration. Accordingly, she became 50 % more certain about the students who are using drugs.

Of course, with each additional query, Eve would gain more certainty. Since differential privacy is composable, this situation is well captured, i.e., ε is additive for multiple queries. In consequence, the global privacy risk and global leak increase for additional queries accordingly.

TABLE 1. Table for a simple evaluation of ε

	worst case $n = 2$		$n = 10$		$n = 100$	
ε	P	Γ	ρ	γ	ρ	γ
ln 2	66.7 %	33.3 %	18.2 %	9.1 %	2.0 %	1.0 %
ln 3	75.0 %	50.0 %	25.0 %	16.7 %	2.9 %	2.0 %
ln 7	87.5 %	75.0 %	43.8 %	37.5 %	6.6 %	5.7 %

Finally, note that while we describe the global privacy risk and privacy leak for a specific example and adversary, it has a universal meaning beyond. The crucial assumption is that the adversary has almost global knowledge about the database except for one data point, which holds one of two possible values. To decide on a value, the adversary compares probabilities of an ε-differentially private outcome for each option. The global privacy risk represents the worst-case probability of the adversary choosing the correct value with the global privacy leak representing the corresponding relative information gain. Assuming global knowledge of the adversary is reasonable as the definition of differential privacy is designed to provide privacy protection for adversaries with arbitrary knowledge.

4. Discussion

To convey the implications of ε or to chose an ε in the first place, two perspectives have to be considered: data engineers and data subjects. In the following, we discuss the privacy risks from both perspectives.

4.1. Data Engineer's Perspective

Data engineers face the challenge to choose the "right" ε. In general, they are interested in accurate results and therefore prefer higher values of ε. Yet, it is difficult to estimate the privacy risk, particularly since the expected number of participants n is not known or cannot be guaranteed in advance.

In this case, our study of the global privacy risk and global privacy leak can serve as a framework. The global privacy risk can provide a basic understanding of the maximum privacy risks for varying ε independent of other parameters. Moreover, the global privacy leak captures the information an adversary can gain in the worst case. If a data engineer can make an assumption on n, she can also calculate the local privacy risk ρ and local privacy leak γ. Under this assumption, ρ and γ yield a less conservative assessment of the privacy loss than the global counterparts. In general, we believe that this framework can support finding acceptable values for ε. In Table 1, we summarize the influence for specific values of ε.

In our running example, a data engineer designing the school database could take the privacy risks for the students as well as the privacy leak into account when deciding on the value for ε.

4.2. Data Subject's Perspective

Data subjects often cannot asses the privacy risks of sharing their data. While they prefer lower values of ε, the value is typically fixed in advance. Other values to calculate their privacy risk are unknown. However, the consequences of the privacy parameters remain subject to a personal assessment.

In order to make a qualified decision, our approach makes an effort to make the implications of ε comprehensible for a general audience. The intentions can be understood similar to a weather forecast predicting the chance of rain, i.e., an adversaries possibility to infer someone's presence or absence. Individuals get an idea of what may happen if they go outside, i.e., contribute their data. Thus, individuals are able to make an informed decision about their actions.

The probability of an adversary inferring someone's presence or absence can be explained illustratively using the randomized response technique [4], as it provides an intuitive mechanism to realize differential privacy [2], [6]. The truthful answer is perturbed by flipping a biased coin, which comes up head and tail with a probability p and $1-p$, respectively. If the coin comes up head, the true answer is released, otherwise the opposite answer. Due to the probabilistic nature of the answers, the data subjects gain *plausible deniability*, i.e. an adversary is unable to distinguish between true and forced answers.

We can apply the narrative of randomized response to our findings and particularly use it even to explain Laplace mechanisms. The ratio of the probabilities p and $1-p$ is accordingly at most e^{ε}. Solving this ratio for p yields Equation (3), which corresponds to the global privacy risk P. The randomized response can also be extended to a larger domain of d possible answers. Rearranging the ratio of probabilities gives the privacy risk ρ in Equation (2), where the parameter d corresponds to the number of data subjects n. In this case, the privacy risk of a Laplace mechanism is equal to a randomized response technique. We therefore envisage to explain the privacy risks of differential privacy using randomized responses.

Unfortunately, the global privacy risk alone provides a distorted picture of the actual success of identification. For instance, $\varepsilon = \ln 2$ yields $P \approx 66.6\,\%$, which is by construction highly pessimistic. In contrast, the global privacy leak yields $\Gamma \approx 33.3\,\%$, which provides an understanding of the *relative privacy loss* (in the worst case) caused by sharing personal data. Since personal risk aversion may differ greatly, we believe that the global privacy leak should be used as a basis for an individual assessment. Overall, such an approach would contribute to a more transparent communication of privacy risks.

In our example, the school could communicate the privacy risks and privacy leak beforehand. In this way, students can make an informed decision about participation or even help in selecting an appropriate value for ε.

5. Related Work

According to Dwork [7], the value of ε is a "social question". Since then, research has been concerned with the explainability of ε in order to communicate its impact.

Naldi et al. [8] propose a method for analyzing the level of differential privacy achieved with the Laplace mechanism for count queries. They chose ε depending on a probability of inferring the true result within a given range. However, the true value is needed. In contrast, we communicate a generic risk for an associated ε.

The approach by He et al. [9] entails representing ε as an identification risk. The risk is given by the adversary's probability of guessing the range of introduced noise and thus the true value. However, the specification of the risk for other functions than count queries remains open.

The economic implications of ε-differential privacy have been explored by Hsu et al. [10]. The value of ε is set based on a given accuracy and the number of contributors. An analyst pays the contributors according to their privacy cost of participating. The method facilitates obtaining appropriate values for ε. However, the accuracy function has to be recalculated for each scenario and the privacy loss parameter ε is not understandable generically.

The groundwork for our approach on the explainability of the privacy loss parameter ε is the work of Lee and Clifton [3]. The authors assume that an adversary chooses a combination of present contributors according to the released result. The probability of the combination being the correct one represents the disclosure risk. In our paper, we modified the risk and connect it to the randomized response technique in order to improve the explainability.

6. Conclusion and Future Work

In this paper, we introduced the notion of a global privacy risk and global privacy leak that can serve as the basis to explain the impact of ε and make qualified choices. In future work, we intend to explore narratives and communication strategies. In particular, it is known that communicating risks as probabilities without context is usually difficult to understand and should instead be communicated as natural frequencies [11]. Moreover, we intend to verify the explainability of ε using the randomized response and our proposed metrics in user studies. This includes considerations of how the randomized response should be explained for clarity and trust [12].

References

[1] C. Dwork, "Differential privacy," in *ICALP '06*. Springer, pp. 1–12.

[2] C. Dwork, N. Kohli, and D. Mulligan, "Differential privacy in practice: Expose your epsilons!" *Journal of Privacy and Confidentiality*, vol. 9, no. 2.

[3] J. Lee and C. Clifton, "How much is enough? choosing ϵ for differential privacy," in *ISC '11*. Springer, pp. 325–340.

[4] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60.309, pp. 63–69.

[5] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *TCC '06*. Springer, pp. 265–284.

[6] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3-4, pp. 211–407.

[7] C. Dwork, "Differential privacy: A survey of results," in *TAMC '08*. Springer, pp. 1–19.

[8] M. Naldi and G. D'Acquisto, "Differential privacy: An estimation theory-based method for choosing epsilon," *arXiv preprint*, vol. abs/1510.00917.

[9] X. He, Y. Hong, and Y. Chen, "Exploring the privacy bound for differential privacy: From theory to practice," *EAI Endorsed Transactions on Security and Safety*, vol. 5.

[10] J. Hsu, M. Gaboardi, A. Haeberlen, S. Khanna, A. Narayan, B. C. Pierce, and A. Roth, "Differential privacy: An economic method for choosing epsilon," in *CSF '14*. IEEE Computer Society, pp. 398–410.

[11] U. Hoffrage and G. Gigerenzer, "Using natural frequencies to improve diagnostic inferences," *Academic Medicine*, vol. 73, no. 5, pp. 538–540.

[12] B. Bullek, S. Garboski, D. J. Mir, and E. M. Peck, "Towards understanding differential privacy: When do people trust randomized response technique?" in *CHI '17*. ACM, pp. 3833–3837.