We thank all reviewers for their valuable comments. In what follows, we will respond to their concerns one by one.

To Reviewer #1

The author should evaluate different attack methods and show the experimental results. Thanks for your advice. Here we refer to the attack used in our paper as the threshold attack. In practice, we found that this simple threshold attack suffices to detect some useful information of the training dataset and can validate the relationship between the information leakage and the generalization ability of GANs. Thus, we mainly demonstrated the results of the threshold attack. Actually, we have also shown the results of other attacks (i.e. black-box attack) in Section 4.3. We will highlight these results in the revision.

It is better to demonstrate some evaluations of the information leakage of Bayesian GAN. Actually, we have conducted attack experiments on the Bayesian GAN and found that Bayesian GAN is effective in reducing both the information leakage and the generalization gap of GANs. We will add these results in the revision.

To Reviewer #2

The pure theoretical advancement is good but not strong. Thanks for your advice. Although the pure theoretical part is somewhat straightforward, it demonstrates a novel and practical view of privacy protection to analyze the generalization property. With the help of the properties of differential privacy, we have provided a more general and concise proof of the uniform convergence in contrast to prior works. Moreover, we have mentioned that our theoretical analysis can be extended to the case of joint learning of the generator and the discriminator (via the post-processing property of the differential privacy [1]). This is useful to analyze the information leakage of both the generator and the discriminator. We will enrich this part in the revision.

Section 4 can be improved by providing guidelines of GAN design. This is a good suggestion! The current results have implicitly shown the connection between the information leakage and the generalization ability of GANs. This empirical evidence indicates that we can improve the generalization ability of GANs by reducing the individual's information leakage. This can be done by reducing each individual's influence on the final learned model. Thus we can propose some design principles to reduce the information leakage via limiting each individual's influence, such as clipping the per-sample gradient or adding Lipschitz conditions on both the generator and the discriminator. In the revision, we will add a specific part to summarize these design principles.

To Reviewer #3

Analysis of the joint learning process of generator and discriminator. Thanks for your suggestion! We have mentioned that the analysis of the joint learning process can be accomplished with the post-process property and the composition theory of differential privacy (see line 158-161 in the paper). We will detail this part in the revision.

The effectiveness of the attack method used in this paper. Some recent works have shown that overfitting of a DNN model is a sufficient condition for the success of the membership attack [2]. Thus, we infer that the success of the attack in our paper comes from the overfitting of the discriminator, i.e. we can use the discriminator's output to distinguish the training data from the test data.

The typo. Thanks for your careful review! We will correct it in the revision.

More explanation of the practical impact, or more novel theoretical analysis. This is good advice for improving our paper! For the practical impact, our empirical evidence shows that we can easily reduce the information leakage of the discriminator without the loss of the generated image's quality (via adding Lipschitz constraints). The results of the black-box attack (in Table 2) also show that Lipschitz constraints can reduce the information leakage of the generator (since the black-box attack depends on the information of the generator). This is meaningful in some real-world applications, such as releasing the learned generator to the public for the research purpose. On the other hand, with our theoretical insight, we can provide some design principles of GANs (see details in the second response to the Reviewer # 2). We will add these potential impacts in the revision. For novel theoretical analysis, in the revision, we can formulate the problem of the joint learning process and add the detailed analysis of this learning process (through the composition theory of the generalization bound and differential privacy).

References

[1] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *FTTCS*, 2014.

[2] Long et al. Understanding membership inferences on well-generalized learning models. *CoRR*, 2018.