Privacy policy robustness to reverse engineering

A. Gilad Kusne[1,*], Olivera Kotevska[2]

[1]National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD, 20899, USA
[2]Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN, 37830, USA

Abstract

Differential privacy policies allow one to preserve data privacy while sharing and analyzing data. However, these policies are susceptible to an array of attacks. In particular, often a portion of the data desired to be privacy protected is exposed online. Access to these pre-privacy protected data samples can then be used to reverse engineer the privacy policy. With knowledge of the generating privacy policy, an attacker can use machine learning to approximate the full set of originating data. Bayesian inference is one method for reverse engineering both model and model parameters. We present a methodology for evaluating and ranking privacy policy robustness to Bayesian inference-based reverse engineering, and demonstrated this method across data with a variety of temporal trends.

Keywords

Differential privacy, Bayesian inference, Privacy policy, Privacy defenses

1. Introduction

In recent years, the number of devices connected to the Internet and online services has increased drastically [1] leading to an exponential growth in data generation [2]. This trend is visible across different domains and applications including among many others, streaming medical, personal tracking, and energy use data [1]. Typically, sensing systems are digitized and connected to network-based analysis tools, and the success of these data streaming devices results in increasing adoption and deployment.

Although the proliferation of connected devices increases convenience across many aspects of life, it also creates dangers when sharing sensitive information. This is especially true for sharing unprotected data over the Internet. Around 98% of device traffic is unencrypted and transmitted over the Internet [3]. Cybercriminals have taken notice of this behavior. On average, sensor-based devices are probed for security vulnerabilities around 800 times per hour, with 400 login attempts and 130 successful logins on each device [4].

To address these rapidly growing security risks, significant effort has been devoted to the development of privacy preservation algorithms and their integration into existing platforms. Some of the most used algorithms are randomization, k-anonymity, l-diversity, cryptography, and differential privacy (DP) [5]. These methods have been successfully demonstrated on big data [6], deep learning [7], medical records [8], as well as other domains. Recent studies have shown that DP is the most effective approach due to its rigorous privacy definition and low computational overhead for continuous (i.e., streaming) data sets [9]. A recent survey identified that DP provides successful privacy preservation with the most common DP mechanisms being Laplacian and Gamma distributions and randomized response [10].

A differentially private model ensures that adversaries are incapable of inferring high confidence information about a single record from released models or output results [11]. However, adversaries may manage to infer or identify sensitive information from employing additional unprotected publicly released data, especially equipped with machine learning tools. Some common attack types proposed include: re-identification attack, membership inference attack, model inversion attack, model extraction attack, and model attribute inference attack [12]. These attacks seek to extract information about the data, model, or attributes.

We investigate the use of Bayesian inference-based attacks to identify the privacy policy employed when pre-privacy protected data samples are available. In this preliminary work we evaluate the likelihood of an adversary to differentiate between a DP mechanism employing either Gaussian or Laplacian additive noise. We build on the previous work of [13]. Application to data with different temporal trends are explored. Here the data streams are sampled from zero-mean Gaussian processes using different kernels, resulting in data with different temporal trends. We then use analysis results to rank privacy policy robustness to such reverse engineering.

CIKM'22: Privacy Algorithms in Systems (PAS), October 21, 2022, Atlanta, GA
*Corresponding author.
aaron.kusne@nist.gov (A. G. Kusne); kotevskao@ornl.gov (O. Kotevska)

2. Model

2.1. Bayesian Inference

Bayesian inference [14] is a statistical sampling method for determining the probability of a hypothesis given data, through the use of Bayes' theorem [15]. A common application for Bayesian inference is to identify the most likely parameters values θ of a generating model $M(\theta)$ for observed data Y. Toward this goal, for a given model, a prior over the parameters is needed. The prior belief for the parameter values θ is given by the probability density function $p(\theta)$ (or the probability mass function $P(\theta)$ if the parameters θ take on discrete values.) The probability of observing data Y given particular values for the parameters is given by $p(Y|M(\theta))$, also known as the likelihood. Through the use of Bayes' theorem, the prior and likelihood are combined to determine the probability of different values of θ given the observed Y. This probability is known as the posterior and is represented by $p(M(\theta)|Y)$. Bayesian inference employs statistical sampling of the model parameters' prior and forward computation of the likelihood to evaluate the posterior. For this work, Markov Chain Monte Carlo (MCMC) is the Bayesian inference sampling method used.

2.2. Gaussian Process

Gaussian process [16] (GP) is a common Bayesian nonparametric regression tool. To learn the function $f : X- > y$ for the data $D = \{(x_i, y_i)\}_i^n$, a prior is assumed $p(f) = N(\mu(x), K(x, x'))$ to quantify epistemic uncertainty. Here $\mu(x)$ is a mean function and $K(x, x')$ is a covariance function. Expected noise in the data (aleatoric uncertainty) is quantified by selected a likelihood, a common one being $p(D|f) = N(f, I\sigma^2)$ which assumes heteroskedastic, normally distributed noise with standard deviation σ. The prior and likelihood are then combined to determine the posterior. When the prior and likelihood are both multivariate normal distributions, the posterior is analytically solvable, giving $p(f|D) = N(\mu_n(x), K_n(x, x'))$, where:

$$\mu_n(x) = \mu(x) + k^T(K + \sigma^2 I)^{-1} y$$
$$K_n(x) = K(x, x') - k^T(K + \sigma^2 I)^{-1} k'$$

Here y is the vector of $\{y_i\}$ and $k_i = K(x, x_i)$ and $k_i' = K(x, x_i)$. For this work, the squared exponential, Matern 5/2, exponential, and Brownian kernels are used to define data stream temporal trends.

2.3. Our approach

We present a methodology for ranking privacy policy robustness to reverse engineering in the presence of exposed or compromised data. We use probabilistic methods to determine the accuracy with which an adversarial actor can identify the privacy policy and its employed parameters from a compromised data stream. Bayesian inference is used to quantify the likelihood (i.e., probability) of each privacy policy being the generating policy and a posterior probability density function for each policy's parameter. Here the target parameter is the privacy loss measure value ε. The privacy policies are then ranked for robustness to this type of attack for the given data.

Ten data stream samples are drawn from each of four Gaussian processes, which differ in kernel. The kernels used are the squared exponential, Matern 5/2, exponential, and Brownian. These data streams are then each privacy protected under two privacy policies - using Gaussian additive noise or Laplacian additive noise. For each policy, we explore the use of varying DP privacy loss measure values of $\varepsilon = [0.1, 0.5, 1.0]$. We assume that x number of pre-privacy protected data samples from each data stream are exposed. We apply Bayesian inference to each of these situations and quantify the sum log likelihood (SLL) of the generating privacy noise policy being either Gaussian or Laplacian. When Bayesian inference identifies that the true generating policy is less likely than the alternative, that policy is ranked greater in robustness to this form of privacy policy reverse engineering. Here, to represent the adversarial attacker's limited knowledge of the privacy policy parameters, the Bayesian inference uses a uniform prior over $[10^{-1}, 10^{-5}]$ and $[10^{-3}, 10]$ for δ and ε, respectively.

For each set of data $D(m_i)$ privacy protected using policy $m_i \in \{Gaussian, Laplacian\}$, we compute the four SLLs: $L_{i,j} = L(m_i|D(m_j))$. The most likely policy is then selected. A measure of whether the used policy is well obfuscated is given by: $\Delta_i = L_{i,j} - L_{i,i}$, with Δ_i positive (negative) if the wrong (right) policy is estimated to be more likely given the data and vice versa. A larger positive value indicates a more difficult challenge for Bayesian inference-based reverse engineering and a larger negative value indicates an easier challenge. We investigate Δ_i as a function of data stream generating kernel, ε value, and size of exposed data stream sample.

2.3.1. Assumption

We assume that x number of raw data points are available. We investigate the robustness of DP Gaussian and Laplacian additive noise to the exposure of varying numbers of data points as well as different values of the DP privacy loss measure ε.

2.3.2. Investigating Privacy Risks

We performed Bayesian inference experiments to determine the privacy policy and its parameters used for pri-

vacy protection. Here the target data stream is privacy protected using the equation: $y_i = y_i + n_i$ with data index i and noise n_i given by either the Gaussian or Laplacian distribution with mean of zero and scale (or standard deviation) given by the following equations:

$$\sigma = \sqrt{2 * log(\frac{1.25}{\delta})} * \frac{sensitivity}{\varepsilon} \quad (1)$$

$$sensitivity = \sqrt{\frac{MaxAE^2}{2}} \quad (2)$$

where MaxAE is maximum allowed error, a function of the raw data. Here the MaxAE is set to one tenth the value of the current data stream value [17]. The adversary is able to obtain data samples prior to privacy protection along with the same data after privacy protection. The pre-privacy protected data can be obtained in a few ways including public exposure by the data stream sensor, by the user (e.g., sharing data on social media), or by the adversary temporarily placing a similar sensing device close to the first, e.g., using a microphone or software hack to listen in to part of a conversation held over a cellphone or IoT device. The resulting pre-privacy protected data can then be used to reverse engineer the privacy policy. Knowledge of the privacy policy can then be used to extract raw data from privacy protected data collected before or after the data exposure occurs as in [13].

3. Results

Figure 1 compares the robustness to Bayesian inference-based model determination for the Laplacian and Gaussian additive noise privacy policies. The difference in SLL $\Delta_i = L_{i,j} - L_{i,i}$ for each investigated case is shown, where $\Delta_{Laplace}$ is plotted in orange and $\Delta_{Gaussian}$ is plotted in green. The means of each are indicated by a solid line and the standard deviation is indicated by the colored regions. Diamond markers indicate a positive mean value and squares indicate a negative mean value - these correspond to the wrong and right policy having greater likelihood, respectively. Interestingly $\Delta_{Laplace}$ tends to be larger than $\Delta_{Gaussian}$, indicating a greater ease in identifying the Gaussian policy over the Laplacian policy. Additionally, both means tend to lower values with increasing number of exposed data samples. In other words, with access to larger amounts of pre-privacy protected data there is an increasing probability in identifying the correct policy, as would be expected. Additionally, a relationship between the choice of kernel or ε and the resulting Δ_i is not clear. Further investigation should be performed where the variance of additive noise is a larger percentage of the data stream variance.

Figure 2 provides a plot of robustness to Bayesian inference-based parameter determination. The deviation in identifying the correct value of ε is plotted for each case. Red, orange, green, and blue indicate $L(m_i|D(m_j))$ for $L(Lap|D(Lap))$, $L(Gaus|D(Lap))$, $L(Lap|D(Gaus))$ and $L(Gaus|D(Gaus))$ respectively. A greater ε increases the difficulty in identifying ε. A greater robustness to parameter determination is shown by $L(Lap|D(Lap))$, while a lower robustness is seen for $L(Gaus|D(Gaus))$. The application of the Laplacian additive noise policy tends to provide greater robustness over the Gaussian policy. As expected, there also appears to be a subtle reduction in parameter estimation error with an increasing number of data points.

4. Conclusion

As more data is shared online, the need for privacy preservation is becoming critical to ensure user confidence in sharing and analyzing personal data with online services. In this paper we investigated the robustness of privacy policies when a subset of pre-privacy protected data is exposed. We demonstrate a methodology for selecting the privacy policy that is more difficult to identify through Bayesian inference-based model and parameter determination. For the range of data stream trends investigated, the Laplacian noise privacy policy was more difficult to identify compared to the Gaussian policy, for both Bayesian inference-based model and parameter determination.

We hope our results and discussion will be helpful to the community using privacy protection for their data sets.

Acknowledgment

This manuscript has been co-authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

References

[1] "Number of IoT connected devices worldwide",

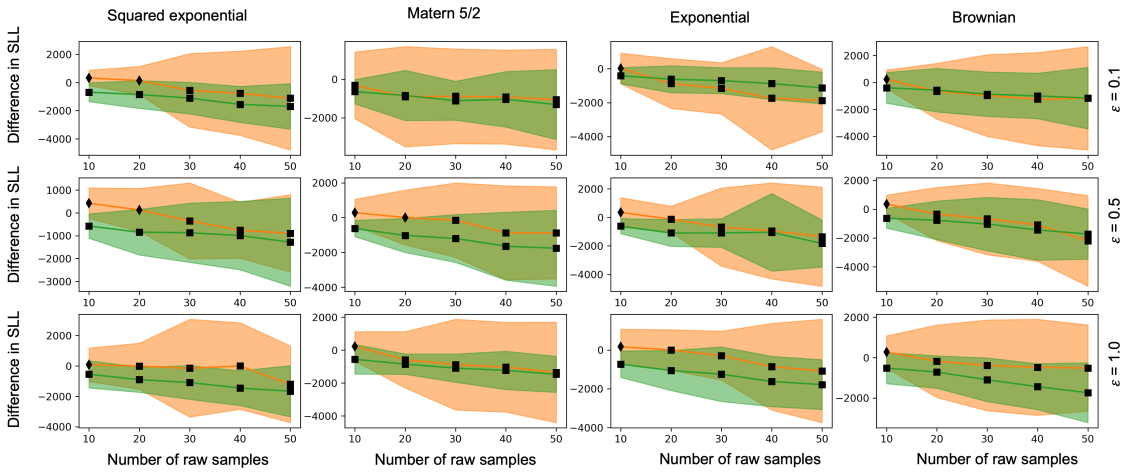

Figure 1: Figure 1. Robustness to Bayesian inference-based model determination. The difference in sum log likelihood (SLL) $\Delta_i = L_{i,j} - L_{i,i}$ for each investigated case. $\Delta_{Laplace}$ is plotted in orange and $\Delta_{Gaussian}$ is plotted in green with the mean indicated by a solid line and standard deviation indicated by the colored region. Diamond markers indicate a positive mean value and squares indicate a negative mean value - these correspond to the wrong and right policy having greater likelihood, respectively.

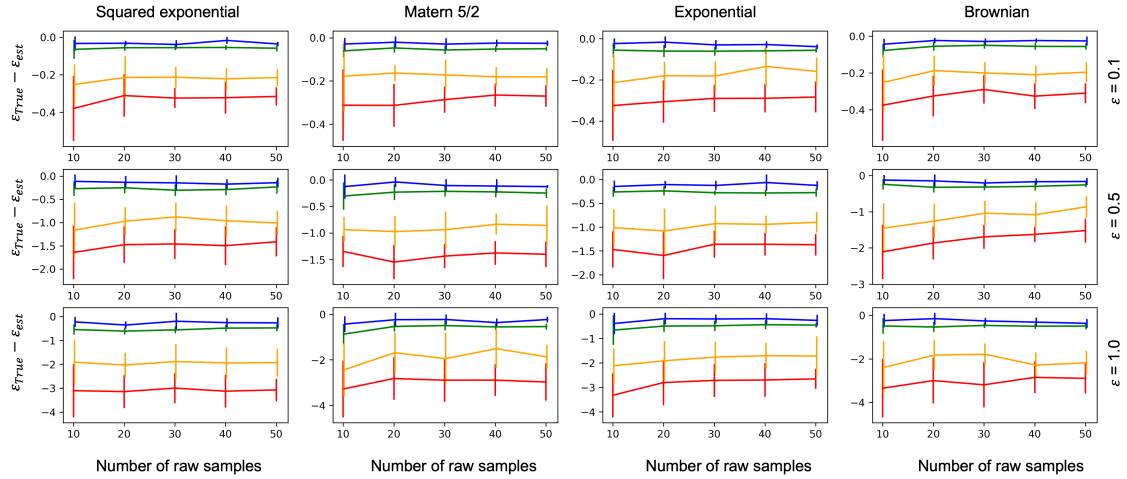

Figure 2: Figure 2. Robustness to Bayesian-inference-based parameter determination. The deviation in identifying the correct value of ε under each case. Red, orange, green, and blue indicating $L(m_i|D(m_j))$ for $L(Lap|D(Lap))$, $L(Gaus|D(Lap))$, $L(Lap|D(Gaus))$ and $L(Gaus|D(Gaus))$ respectively. A greater ε increases the difficulty in identifying ε. The most difficulty is shown with $L(Lap|D(Lap))$, while the greatest ease is shown with $L(Gaus|D(Gaus))$. There also appears to be a subtle reduction estimate error with increasing number of data points. Error bars are slightly shifted for visibility.

https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/, Accessed: August 9, 2022.

[2] "The Growth in Connected IoT Devices", shorturl.at/lmnv1, Accessed: August 9, 2022.

[3] "2020 Unit 42 IoT Threat Report", https://start.paloaltonetworks.com/unit-42-iot-threat-report, Accessed: August 17, 2022.

[4] "4 Ways Cyber Attackers May Be Hacking Your IoT Devices Right Now", shorturl.at/bdjm1, Accessed: August 17, 2022.

[5] M. Khan, S. Foley, B. O'Sullivan, From k-anonymity to differential privacy: A brief introduction to formal privacy models (2021).

[6] S. H. Begum, F. Nausheen, A comparative analysis of differential privacy vs other privacy mechanisms

for big data, in: 2018 2nd International Conference on Inventive Systems and Control (ICISC), 2018, pp. 512–516.

[7] J. Vasa, A. Thakkar, Deep learning: Differential privacy preservation in the era of big data, Journal of Computer Information Systems (2022) 1–24.

[8] A. Kumar, R. Kumar, Privacy preservation of electronic health record: Current status and future direction, Handbook of Computer Networks and Cyber Security (2020) 715–739.

[9] M. Peralta-Peterson, O. Kotevska, Effectiveness of privacy techniques in smart metering systems, in: 2021 International Conference on Computational Science and Computational Intelligence (CSCI), IEEE, 2021, pp. 675–678.

[10] Y. Zhao, J. Chen, A survey on differential privacy for unstructured data content, ACM Computing Surveys (CSUR) (2022).

[11] M. Al-Rubaie, J. M. Chang, Privacy-preserving machine learning: Threats and solutions, IEEE Security & Privacy 17 (2019) 49–58.

[12] M. Rigaki, S. Garcia, A survey of privacy attacks in machine learning, arXiv preprint arXiv:2007.07646 (2020).

[13] O. Kotevska, J. Johnson, A. G. Kusne, Analyzing data privacy for edge systems, in: 2022 IEEE International Conference on Smart Computing (SMARTCOMP), IEEE, 2022, pp. 223–228.

[14] G. E. Box, G. C. Tiao, Bayesian inference in statistical analysis, John Wiley & Sons, 2011.

[15] H. Pishro-Nik, Introduction to probability, statistics, and random processes (2016).

[16] C. E. Rasmussen, Gaussian processes in machine learning, in: Summer school on machine learning, Springer, 2003, pp. 63–71.

[17] M. U. Hassan, M. H. Rehmani, R. Kotagiri, J. Zhang, J. Chen, Differential privacy for renewable energy resources based smart metering, Journal of Parallel and Distributed Computing 131 (2019) 69–80.