IBIMA Publishing
Journal of Information Assurance & Cyber security
https://ibimapublishing.com/articles/JIACS/2023/892906/
Vol. 2023 (2023), Article ID 892906, 9 pages, ISSN: 2165-9923
https://doi.org/10.5171/2023.892906

Research Article

Performance and Security Considerations for Differential Privacy Model with Adaptive Metrics

Olga DZIEGIELEWSKA and Boleslaw SZAFRANSKI

Military University of Technology, Warsaw, Poland

Correspondence should be addressed to: Olga DZIEGIELEWSKA; olga.dziegielewska@wat.edu.pl

Received date:13 March 2023; Accepted date:27 July 2023; Published date: 17 October 2023

Academic Editor: Agnieszka Stanimir

Copyright © 2023. Olga DZIEGIELEWSKA and Boleslaw SZAFRANSKI. Distributed under Creative Commons Attribution 4.0 International CC-BY 4.0

Abstract

Over time, the differential privacy model has undergone numerous extensions and implementations by various researchers. Recently, a notable advancement introduced two adaptive metrics aimed at achieving improved calibration of security and accuracy by considering the risk associated with statistical queries. The efficacy of this extension highlights its potential superiority over the original version under specific conditions. Building upon this progress, our paper presents further enhancements to the extended differential privacy model, focusing on selected aspects of its implementation. These modifications hold promise for enhancing the overall performance and security of the solution at a system level.

Keywords: differential privacy model, VIOLAS Framework, retrofitting of security mechanisms, statistical databases security

Introduction

As defined by Dwork (2006), the differential privacy model is a perturbative statistical disclosure control mechanism that involves adding appropriate random noise to the statistical queries to provide a countermeasure against inference attacks directed at statistical answers. Dziegielewska further explains (2017) that the differential privacy model allows a data solicitor to collect data and infer meaningful information from the data without individual record attribution, i.e., the mechanism allows a solicitor to collect sensitive data, but the data cannot be attributed to any party.

Until today, the differential privacy model is still quite an unexplored statistical disclosure control

Cite this Article as: Olga DZIEGIELEWSKA and Boleslaw SZAFRANSKI (2023), "Performance and Security Considerations for Differential Privacy Model with Adaptive Metrics", Journal of Information Assurance & Cyber security, Vol. 2023 (2023), Article ID 892906, https://doi.org/10.5171/2023.892906

mechanism, however, some researchers attempted to extend it and improve it in various ways (e.g., Dwork et al. (2006) propose distributed noise generation, Hall et al. (2011) introduce randomization, or Chatzikokolakis et al. (2013) define new boundaries). One of the extended versions of this model, introduced by Dziegielewska (2020), proposes two adaptive metrics – risk-accuracy and information score – to adjust the noise that is added over the calculations to improve the security of the retrieved statistics without losing the accuracy aspect of the data:

- Risk-Accuracy Metric – an adjustable metric that allows manipulating the statistical noise in such a way that the result would provide a selected level of a tradeoff between the accuracy and the security, which could be easily adapted for the real-life usages of statistical databases, e.g., the different tradeoffs for data processed only internally by an entity and sent to external entities or providing role-based tradeoffs at the database level.
- Information Score Metric – an incremental metric that by design automatically changes the noise of the same or similar query retrieved multiple times. This metric intends to mitigate the risk of statistical disclosure in case of dynamic inference attacks while the results remain statistically integral. This factor can potentially be expanded to provide a multi-source inference control mechanism, i.e., in attack scenarios where more than one database is used.

The efficiency of the extended differential privacy method was determined by the statistical security criteria defined by Dziegielewska (2017) in the VIOLAS Framework and threat modeling covered by Dziegielewska (2022) showed that the method as defined can be considered a better candidate than the original version in certain conditions, e.g., when the data can be correlated with multiple external sources to retrieve sensitive statistics.

```
INFORMATION SCORE METRIC DERIVATION
        Prerequisites:
p1      AR_DB
p2      HIST_DB
                Input:
i1      P (a set of parameters of the submitted query)
i2      C (asked statistical characteristic)
i3      S (optional, default=1)
                Algorithm:
a1      if (query was asked before)
a2              return ism=ism_hist
a3      else
a4              while (AR_DB has next)
a5                      if (P and C in AR_DB_i)
a6                              add {P,C,ism=1·S} into HIST_DB
a7                              return ism=1·S
a8                      else
a9                              while (HIST_DB has next)
a10                                     if (P_hist_i,C_hist_i,C,P in AR_DB_i)
a11                                             add {P_hist_i,C_hist_i,P,C,ism=1·S}
                        to HIST_DB
a12                                             return ism=1·S
a13                                     else
a14                                             add {P,C,ism=0} into HIST_DB
a15                                             return ism=0
```

Figure 1 The original algorithm for Information Score Metric Derivation as defined by Dziegielewska (2020)

This paper further improves the extended differential privacy model with selected aspects of the implementation of the mechanism that may contribute to improving the performance and security of the solution.

Statistical Disclosure Control Mechanisms Implementation Considerations

In real-life implementations, the decision on the differential privacy variant's application is based on the system's business requirements and key performance indicators which the system needs to comply with, which vary from one system to another. The ultimate responsibility for the overall security and the performance of the system belongs to the system's owner, meaning that if the components of the system underperform significantly, the system's owner has the authority to approve correcting actions to address the issues. In the case of statistical

disclosure controls, it would also be a system's owner's decision to make a final call on the parameters and features required from the business perspective, however, the technical team needs to design and propose appropriate solutions, e.g., differential privacy mechanism, to meet the desired effect.

On top of the functional requirements of the statistical disclosure control method, non-functional requirements, including security, need to be taken care of. In the case of the security requirements, preventive, detective, and corrective controls at each stage of the SDLC and DBLC are defined and the whole development lifecycle must adhere to them.

Currently, with modern shift-left approach, the security requirements and mechanisms are added much before the implementation stage, what mitigates major security risks by preventing them at early stages of the system's lifecycle

```
RISK-ACCURACY METRIC DERIVATION
                Prerequisites:
p1      r_preset(preset of the risk component of the database)
p2      a_preset(preset of the accuracy component of the database)
                Input:
i1      P (a set of parameters of the submitted query)
i2      C (asked statistical characteristic)
                Algorithm:
a1      calculate ism for {P,C}
a2      risk=r_preset*ism
a3      if risk=0
a4              ram=a_preset-1
a5      else
a6              ram=a_preset*risk+risk
```

Figure 2 The original algorithm for Information Score Metric Derivation as defined by Dziegielewska (2020)

The latest trends can be easily adapted to the newly developed systems, but the transformation for the already existing systems remains a challenge.

Since the IT environments evolve over time, retrofitting is a process that will occur for most of the IT systems to adapt to new functional and non-functional requirements. The extended differential privacy method could be used in both cases – newly developed and already existing solutions, thanks to its post-processing interfaced nature, i.e., the data are derived from the raw data sets, and the algorithm itself can be implemented as an additional end-user interface, without affecting the existing system's implementation. It must be noted, however, that in the case of the systems that collect data that are already affected by some perturbative method at the data collection time, the method will not be as effective in terms of accuracy, as additional noise will be added to existing distorted data.

Implementing the Extended Differential Privacy Model

There are two ways in which perturbative statistical disclosure control methods can be applied: at the time of data collection, where the raw data are modified and the database stores already noisy data, or as a part of the statistical derivation interface of the database, where the raw data are stored in the database without any added noise before the derivations. The advantage of the second case is that neither the data collection method nor the data model of the database must satisfy any statistical security-by-design measures and the extended differential privacy model is an example of such a statistical disclosure control mechanism.

Adopting the post-processing approach makes the method applicable in a broader context, as it can be easily implemented not only for future databases, but more crucially, it can be integrated as an additional end-user interface for existing databases. The demand for retrofitting applications has grown, underscoring the significance of developing a statistical disclosure control method that is compatible with already established and functional systems. Dziegielewska (2020) discusses that this approach would have a greater business impact on the current IT

landscape than designing an infallible method that is only applicable to future databases.

As with any perturbative statistical disclosure control method used as an external non-customized interface, there is an increase in processing time for the executed database queries. However, this cost is necessary to ensure the privacy of the statistical output.

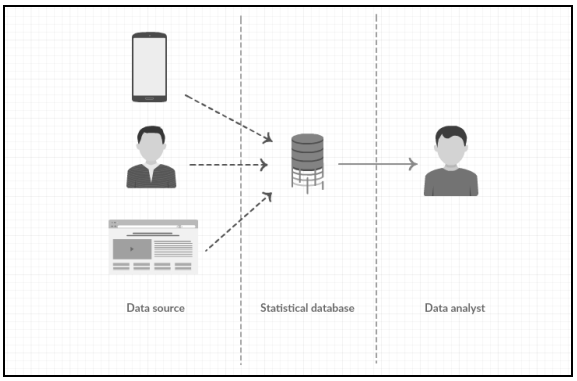

Figure 2 The scheme depicts the place of applying the extended differential privacy (marked with a solid arrow).

The initial considerations regarding the implementation prerequisites were included in the original paper by Dziegielewska (2020) and assume that:

- The data in the statistical database are stored in non-perturbed manner.
- The database has different statistical access roles that have different sensitivity levels assigned to them.
- The association database and historical database are preprocessed.

The process itself starts with the verification of the required presets (the sensitivity level, risk level, accuracy level) for a database access role of a user who submitted a query. The preset values are valid throughout the validity of a user session.

Each query is verified first in the historical database and the associations are verified to determine the risk profile of the query and the *ism* (information score metric output) is either calculated from scratch or gathered from the historical data. Later, the *ism* is used by the risk-accuracy metric and the output of the calculations is passed to the noise generation function as one of the inputs. After the error rate has been determined and the noise generation function has been applied to the raw data, the perturbed output is returned to the user (Dziegielewska, 2020).

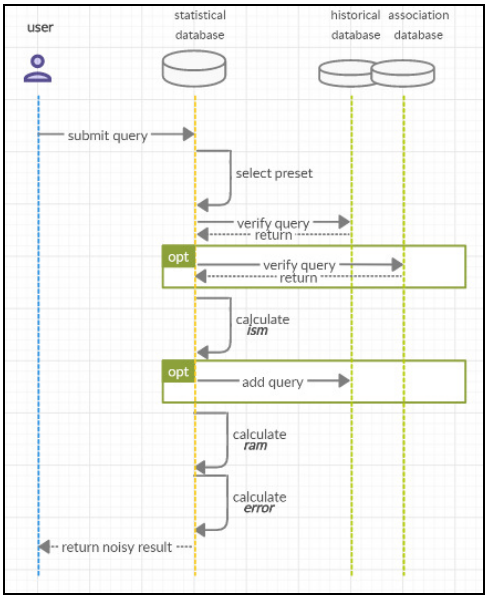

Figure 3: Transaction scheme as defined by Dziegielewska (2020)

The user should not receive any result from the database until the statistical disclosure control process is not terminated. In case an error occurred before the output retrieval, all the transactions made into the supporting databases that could affect future results must be rolled back, i.e., any datum that has been added to the historical database must be deleted (Dziegielewska, 2020).

Implementation-Specific Performance Improvements

In the original paper, Dziegielewska (2020) designed *Enhancing Distribution* algorithm in such a way that it follows the fundamental differential privacy premise, i.e., the noise is added to all the results, regardless of its risk profile or the risk profile of the database.

However, in case the risk profile of the database (*r_preset*) or the risk profile of the retrieved query (*ism*) is within a negligible margin, then some steps may be skipped to improve the performance. This could however affect the classification of the method as a differential privacy mechanism; therefore, it was not proposed in the original design.

The algorithm below shows an improved noise distribution enhancement algorithm that improves the performance of non-critical systems or queries.

```
ENHANCING DISTRIBUITION
                Prerequisites:
p1      D(d_q)
                Input:
i1      q
                Algorithm:
a1      calculate ram for q
a2      if ram≤0
a3              use D(d_q-normalize(ram)[1])
a4      else
a5              use D(d_q+normalize(ram)[1])
```

Figure 3: The original algorithm for Information Score Metric Derivation (paper) 1)In the original paper, the normalization step is defined as a multiplication function, however, it was altered in this paper for better clarity.

The changes include breaking the execution and returning non-perturbed results if the risk profile of the database (line a1) or the asked query (line a5) is equal to 0. If the risk profile of the query is equal to 0, which renders no potential inference attacks discovered for the query, then the algorithm returns an unmodified result. If the risk profile of the query is not equal to 0, then the randomized generator receives values greater than 0, which reflects that the risk factor is taken into account by the generator, and the noise must be modified accordingly, therefore the error rate must be increased.

ENHANCING DISTRIBUITION – IMPLEMENTATION-IMPROVED VERSION

```
        Prerequisites:
p1      D(d_q)
p2      r_preset
        Input:
i1      q
        Algorithm:
a1      if r_preset = 0
a2          return non-perturbed result
a3      else
a4          calculate ram for q
a5          if ram=0
a6              return non-perturbed result
a7          else
a8              D(d_q+normalize(ram))
```

Additionally, a change in the risk-accuracy metric derivation algorithm needs to be done to adapt to the new conditions – in case *ism=0*, then the *ram=0*, i.e., if the risk profile of the query is equal to 0, what renders no potential inference attacks discovered for the query, then the returned value is 0. If the risk profile of the query is not equal to 0, then the randomized generator receives values greater than 0, which reflects that the risk factor is taken into account by the generator, and the noise must be modified accordingly.

RISK-ACCURACY METRIC DERIVATION - – IMPLEMENTATION-IMPROVED VERSION

```
        Prerequisites:
p1      r_preset(preset of the risk component of the database)
p2      a_preset(preset of the accuracy component of the database)
        Input:
i1      P (a set of parameters of the submitted query)
i2      C (asked statistical characteristic)
        Algorithm:
a1      calculate ism for {P,C}
a2      risk=r_preset*ism
a3      if risk=0
a4          ram=0
a5      else
a6          ram=a_preset*risk+risk
```

It must be noted, however, that without covering all the results with the noise distribution function, the security can be affected, e.g., in case a risk of a particular query is wrongly categorized as negligible risk, due to the insufficient quality of the *AR_DB* or *HIST_DB*, then the retrieved results remain unprotected from the inference attacks. Therefore, the performance-enhancing solution needs to be adopted carefully over the system. Obviously, the proposed performance enhancements may be further adapted, e.g., only *r_preset* verification can be added in the

Enhancing distribution algorithm, leaving the noise addition method in all the rest cases.

Apart from the algorithmic improvements, database-level improvements can be considered. Typically, statistical databases themselves are relational databases, however, to improve the lookup time the *AR_DB* and *HIST_DB* databases could be implemented as non-relational databases. Nonetheless, the decision on the actual database type should not only consider the performance, but also the technological stack of the organization that the solution is implemented.

Additional Security Considerations

As was already mentioned, the extended differential privacy model was evaluated using VIOLAS Framework designed for statistical disclosure control methods (Dziegielewska, 2020) and additional threat modeling to ensure a better quality of the outputs (Dziegielewska, 2022). The evaluations confirmed that the model of the method itself is secure under the statistical disclosure control mechanism conditions, i.e., satisfies statistical confidentiality, statistical integrity, statistical accuracy, and statistical transparency (Dziegielewska, 2020).

The statistical disclosure control methods typically focus on the data layer of the system, as the statistical inference attacks are classified as business logic abuse rather than environment or implementation-related attacks. The attacks leverage the vulnerable data model design as in the inference attack scenarios it is assumed that the access to the dataset is granted by default to a certain group of system users, and the users abuse the legitimate data-level access.

However, since the security of the working environment also plays a major part in the overall security of the system, other factors must also be considered while assessing the risk of the system in scope. For IT systems, the security requirements can be split into two categories: procedural and technical.

Procedural Security Assumptions

The procedural security requirements embed security aspects into the business processes that are supported by an IT system. Typically, such requirements include but are not limited to regulatory compliance, service continuity, system, and data governance, third-party risk management, and access model definition.

In the case of the extended differential privacy model, there are several procedural aspects that must be properly defined and implemented, as they directly influence the effectiveness of the elaborated method; they are:

- Regulatory compliance
- Access model definition
- System and data governance

Other aspects, despite being important from a high-level system perspective, do not immediately influence the elaborated method, therefore will not be commented.

The *regulatory compliance* should be treated with the highest priority as a lack of compliance with local or global legislations or formal requirements may result in legal actions and regulatory fines which may have a direct impact on an entity that intends to implement the proposed mechanism. The personal data privacy is especially sensitive in some geographic areas (e.g., EU, South Korea, Russia) or fields (medical data), therefore before considering the proposed solution, regulatory validation should be performed. When it comes to the compliance of the model against reidentification of the individual records, it is proven by the research that the model can suffice this sort of requirement. However, there may exist regulations that would prevent from using the extended model, e.g., if a legislator bans abusing integrity of the statistics in any way, then both base and extended differential privacy models cannot be implemented.

The *access model* requirements should be carefully defined at different layers. First of all, the database's statistical and non-statistical access must be considered, however, apart from the database, other system components should be also addressed. As the immediate risk for the effectiveness of the proposed method originates from the implementation and configuration of the algorithms, the access model for the system code base pipeline, the algorithm's configuration, and the system's configuration must also be established following the least-privilege principle. Regular users should not have permission to modify algorithms' setup (*S, a_preset, r_preset*, D(d_q)) or to modify the system's configurations

to prevent overwriting the security setup of the system. Additionally, the system's developers should not have a direct possibility to effectively push and render changes in the production environments, to prevent introducing uncontrolled modifications in the algorithm. The code development pipeline should be configured following secure software development standards (e.g., OWASP SAMM [9], BSIMM [10], ISO series: 9000 [11], ISO/IEC 12207 [12], ISO/IEC 15288 [13], ISO/IEC 24748 [14]).

The *system and data governance* aspect that is important for the quality of the results returned by the extended differential privacy model is the assignment of the variable values *S, a_preset, r_preset*. An appropriate strategy is needed to define those values for the IT system to fully leverage the functionalities of the extended differential privacy model. Obviously, the values of the variables are the key to a successful implementation of the model, however, it must be noted that the variables can be set at a different level for different types of users. This way, the results can not only be calibrated at a system layer but also at an individual level per each user or group of users, which makes the proposed model even more adaptive.

Technical Components Security Assumptions

The technical components of security are a separate area that is not covered in the scope of the extended differential privacy model but still states a significant part of the overall security of the implementation. Two aspects are critical in terms of satisfying the security of the statistical disclosure control methods at the technical level: the environment and the implementation security. The best practice for defining specific security requirements in those two areas is to follow one of the commonly used and globally approved standards for IT system development.

One of them is OWASP Application Security Verification Standard [15], which thoroughly covers all the phases of the Software Development Lifecycle with controls to apply in certain aspects of the developed solution. OWASP ASVS v. 4.0.2 covers 70 specific validation points divided into 14 categories. It must be stressed that, for the overall security of the system, it is important to cover all the requirements, however, from the perspective of the extended differential privacy model, the key categories that should be considered mandatory are:

- V1: Architecture, Design and Threat Modeling Requirements
- V4: Access Control Verification Requirements
- V7: Error Handling and Logging Verification Requirements
- V8: Data Protection Verification Requirements
- V9: Communications Verification Requirements
- V11: Business Logic Verification Requirements
- V14: Configuration Verification Requirements

It must be assumed that the environment is secured accordingly and follows best security practices, to prevent any side-channel data leakage, which would lead to releasing sensitive data, thus rendering the model ineffective, in particular:

- the statistical database, historical database, and association database are properly secured at the access level, including external connectivity if it is necessary from the business perspective;
- the sensitive properties: variables (*S, a_preset, r_preset*) and functions (*D(d_q)*) are considered critical and secured with the mechanism preventing or limiting from direct access and modification;
- the communication between the database users and the database is properly secured with the latest communication in-transit recommendations (e.g., the highest available TLS version with secure cipher suites); this is especially important during the data acquisition process;
- the sensitive actions over the systems' components, especially databases, are logged to detect any suspicious alterations or unauthorized access.

Apart from the environmental perspective, it is crucial that all the technical security requirements are properly reflected through the implementation of the solution; in particular, it must be assumed that:

- the implementation of the proposed model is free from errors that would affect the security

of the solution, e.g., business logic errors that would affect the processing of the data;
- the system is free from backdoors that would intend to bypass the security mechanisms;
- the system hardening and security configuration (e.g., defined access model implementation) is effectively implemented and free from errors;
- the fail-safe rule is effectively implemented, especially error handling does not reveal any excessive information, sensitive properties are never revealed through the error condition;
- other application-level errors and vulnerabilities are detected at the early stages of the SDLC lifecycle and properly addressed.

Final Remarks

Ensuring the quality of any proposed statistical disclosure control models must always go beyond the design. Even if the mathematical model is assessed and tested, there may be still a way of improving the solution at implementation level. Although achieved measurable and impactful performance improvements beyond algorithmic definition may be hard to achieve, the overall security of any statistical disclosure control model depends on the robust security of the system that applies this model. The paper summarized the fundamental aspects of the implementation-level improvement for the performance and security which can be applied to the extended differential privacy model. Additionally, what is worth noting is that through this paper we can determine that the quality of any statistical disclosure control mechanism should consider ease for implementation and retrofitting which is not covered in the statistical disclosure control evaluation framework (Dziegielewska, 2020), thus the framework itself should be extended by the additional implementation characteristics.

References

- Dwork, C. (2006). Differential Privacy. *In Proceedings of the 33rd international conference on Automata, Languages and Programming - Volume Part II (ICALP'06)*, 1-12.
- Dzięgielewska, O. (2017). Anonymization, tokenization, encryption. How to recover unrecoverable data. *Computer Science and Mathematical Modelling No.6*, pp. 9-13.
- Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., & Naor, M. (2006). Our data, ourselves: Privacy via distributed noise generation. *In Advances in Cryptology-EUROCRYPT 2006*, 486–503.
- Hall, R., Rinaldo, A., & Wasserman, L. (2011). Random differential privacy. *arXiv:1112.2680*.
- Chatzikokolakis, K., Andrés, M., Bordenabe, N., & Palamidessi, C. (2013). Broadening the scope of Differential Privacy using metrics. *In Privacy Enhancing Technologies*, 82–102
- Dzięgielewska, O. (2020), Defeating Inference Threat with Scoring Metrics. 36th IBIMA Conference on 4-5 November 2020 Granada, Spain.: Conference proceedings (ISBN: 978-0-9998551-5-7, Published in the USA).
- Dzięgielewska, O. (2020), Evaluating Quality of Statistical Disclosure Control Methods – VIOLAS Framework. *Lecture Notes in Computer Science, Springer, Cham. , 12276* (Privacy in Statistical Databases. PSD 2020.).
- Dziegielewska, O. (2022), Evaluating adaptive differential privacy model, *Computer Science and Mathematical Modelling No. 13-14*, pp. 5-16.
- OWASP Software Assurance Maturity Model: https://owasp.org/www-project-samm/
- BSIMM. *BSIMM Framework*. Retrieved from https://www.bsimm.com/framework.html
- ISO 9000 FAMILY QUALITY MANAGEMENT: https://www.iso.org/iso-9001-quality-management.html
- ISO/IEC 12207:2008 Systems and software engineering — Software life cycle processes: https://www.iso.org/standard/63712.html
- ISO/IEC 15288:2008 Systems and software engineering — System life cycle processes: https://www.iso.org/standard/63711.html
- Systems and software engineering — Life cycle management — Part 1: Guidelines for life cycle management: https://www.iso.org/standard/72896.html
- OWASP Application Security Verification Standard: https://owasp.org/www-project-application-security-verification-standard/