Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9764363, 1 page
https://doi.org/10.1155/2023/9764363

Retraction

Retracted: The Research of Adaptive Data Desensitization Method Based on Middle Platform

Wireless Communications and Mobile Computing

Received 13 September 2023; Accepted 13 September 2023; Published 14 September 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Wang, M. Xu, and K. Lu, "The Research of Adaptive Data Desensitization Method Based on Middle Platform," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 5348637, 7 pages, 2022.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 5348637, 7 pages
https://doi.org/10.1155/2022/5348637

Research Article

The Research of Adaptive Data Desensitization Method Based on Middle Platform

Jijun Wang,[1] **Mingsheng Xu**,[2] **and Kang Lu**[2]

[1]*State Grid Jiangsu Electric Power Co., LTD., Nanjing, Jiangsu 210000, China*
[2]*Jiangsu Electric Power Information Technology Co., LTD., Nanjing, Jiangsu 210000, China*

Correspondence should be addressed to Jijun Wang; 201903301@stu.ncwu.edu.cn

Received 26 June 2022; Revised 27 July 2022; Accepted 1 August 2022; Published 18 August 2022

Academic Editor: Aruna K K

With the popularization of Middle Platform characterized by data aggregating and governance, the security protection of data is paying more and more attention, and the data desensitization technology is widely used. In order to solve the high threshold for use, high customization, and lack of stability caused by conventional data desensitization methods, a desensitization strategy configuration system based on desensitization intensity and desensitization algorithm weight was established. The methods of reidentification risk assessment and information security attribute assessment were used to classify and quantify configuration items, and then, an adaptive desensitization strategy configuration method was proposed which not only simplifies the configuration process but also provides reliable desensitization data flexibly and stably for application requirements. It is beneficial to the development of an intelligent automatic data desensitization system.

1. Introduction

Middle Platform is built on a big data platform, abstracting and encapsulating data structures into service APIs by the aggregation and governance of crossdomain data (data from different sources). It can make up for the speed difference between data development and application development and coordinate developments to improve overall development efficiency. More and more Middle Platforms are being built in China. Also, in Silicon Valley, although there is no specific title like "Middle Platform," there are many practical applications of similar functions on data platforms [1]. The architecture of big data platform is shown in Figure 1.

A large amount of data is deposited to Middle Platform for high-frequency access, query, processing, and calculation. The user privacy or trade secrets carried in the data constitute sensitive data. International regulations in various industries require that data should be privacy-protected before open use [2, 3]. Therefore, the safe use of sensitive data on Middle Platform is a challenge. In the current practices of privacy protection, data desensitization is a common and efficient technical means, which transforms the original sensitive data into desensitized data with reduced sensitivity. Desensitized data is a certain degree of distortion in exchange for the improvement of data security and still retains a part of the data value.

The application of data desensitization mainly relies on three concepts, desensitization algorithm, desensitization rule, and desensitization strategy. A conventional desensitization system, with some desensitization rules built-in for each sensitive data [4], combines the various desensitization rules to perform desensitization tasks. In this method, desensitization strategies are driven by desensitization rules to meet application requirements. Using this method, users have to learn to desensitize algorithms and desensitization rules and accumulate application requirements in the

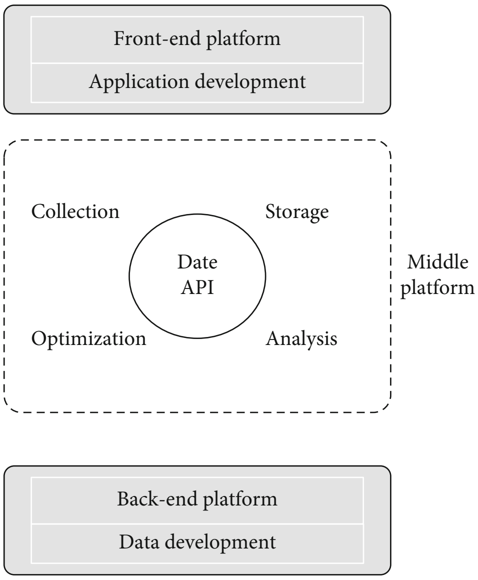

Figure 1: Big data platform architecture.

multilevel configuration process, highly dependent on experience. At the same time, the fixed built-in rules are difficult to deal with new application requirements, and the desensitization rules must be modified according to customization. Furthermore, due to excessive manual intervention, limited by the influence of personnel, it is hard to maintain a unified and continuous judgment standard. The desensitization results are uncertain and nonrepeatable, which will lead to multiple operations because of failure to meet the application needs.

In order to solve the above problems, this paper analyzed the desensitization rules in the context of electric power industry scenario and data and defined the desensitization strength and desensitization algorithm weight, according to the reidentification risk assessment theory [5–7] and information security attributes. A quantitative assessment of the desensitization results in terms of both confidentiality and availability was obtained. A method of dynamically generating desensitization rules driven by desensitization strategy is proposed, which is guided by application requirements, so that the desensitization results are evidence-based and repeatable. The use cost went down, and the expansion of algorithms and applications was convenient.

2. Materials and Methods

2.1. Analysis of Desensitization Rules. Desensitization algorithms are the deformation methods used in the desensitization process. A desensitization rule is formed by applying the algorithms to a specific sensitive data. Desensitization rules are named after the sensitive data names. One sensitive data can be mapped to multiple desensitization rules. Table 1 shows some common desensitization algorithms [8]. Table 2 shows several common mobile No. desensitization rules.

Algorithms are theoretically universal for any data, but each has its own applicable data categories and application scenarios.

As shown from Table 2, different desensitization rules processing the same data produced different desensitization results. It was difficult to assess whether the desensitization results meet the application requirements by the descriptions of the desensitization rules alone.

However, it could be found that the description of a desensitization rule consists of two parts: a desensitization algorithm and the location where the desensitization was performed.

The desensitization rules of Table 2 were decomposed: if the algorithm was uniformly set to “mask,” then Table 3 was obtained; if the location was uniformly set to “the bottom 8 bits,” then Table 4 was obtained.

It could be seen from Tables 3 and 4 that, for a sensitive data, there were two factors influencing desensitization results, which were defined as follows:

Definition 1. The effect of the desensitization location on the desensitization result was called the desensitization intensity.

Definition 2. The effect of the desensitization algorithm on the desensitization result was called the algorithm weight.

2.2. Adaptive Desensitization Strategy. By quantifying the desensitization intensity and algorithm weight, respectively, the quantitative evaluation of the desensitization result would be obtained, which was the basis for the flexible configuration of desensitization strategy.

2.2.1. Quantification of the Desensitization Intensity

(i) Estimating reidentification risks

Canadian scholars Emam et al. had proposed three common privacy attack scenarios (prosecutor attack, journalist attack, and marketer attack) and designed risk indices to estimate the risk of reidentification of structured desensitization data (hereinafter referred to as risk). Referring to the experiment [9], the following qualitative conclusions were obtained:

(1) The risks trends of the three attack scenarios were similar, related to the distribution of the probability of data repetition

(2) The lower the probability of data repetition, the higher the desensitization intensity and the lower the risk

(3) The data repetition probability was related to the data encoding structure and rules

(ii) Desensitization intensity grading

Steps are as follows:

Table 1: Common desensitization algorithms.

Algorithm	Description	Example
Mask	Use symbol "*" to replace parts of the data, with the data length unchanged	123456 – >1234 ** 321427198910156223 – >32 **********156223
Floor	Take an integer	19 – >10 19 : 30 : 03 – >19 : 00 : 00
Hashing	Map data into a fixed-length string	Jim – >51talk Tom – >123456
Truncation	Cut parts of the data	13088886666 – >130 010 – 22226666 – >010
Shift	Add a constant offset	233 – >2233 466 – >2466
Synthesis	Simulate new data to replace the original data	13088886666 – >13911007788 010 – 22226666 – >021 – 49494499
Rearrange	Sort a column of values upside-down	20, 30, 40 – >30, 40, 20

Table 2: Common mobile No. desensitization rules.

Rule	Description	Example
1	Keep the top 3 & bottom 4, use "mask" for the middle 4 bits.	13088886666 – >130 ****6666
2	Keep the top 3, use "truncation" for the rest.	13088886666 – >130
3	Keep the bottom 4, use "synthesis" for the rest.	13088886666 – >13911006666

Table 3: Mobile No. desensitization examples with same algorithm.

Rule	Reserved bits	Desensitized bits	Algorithm	Example
1	Keep the top 3 & bottom 4	The middle 4 bits	Mask	13088886666 – >130 ****6666
2	Keep the top 3	The bottom 8 bits	Mask	13088886666 – >130 ********
3	Keep the bottom 4	The top 7 bits	Mask	13088886666 – > ** *****6666

Table 4: Mobile No. desensitization examples with same location.

Rule	Reserved bits	Desensitized bits	Algorithm	Example
1	Keep the top 3	The bottom 8 bits	Mask	13088886666 – >130 ********
2	Keep the top 3	The bottom 8 bits	Truncation	13088886666 – >130
3	Keep the top 3	The bottom 8 bits	Synthesis	13088886666 – >13073965031

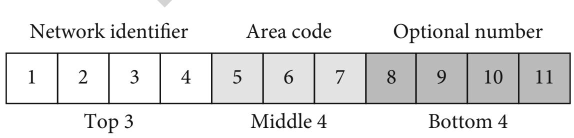

Figure 2: Encoding structure of mobile No.

(1) Analyze the data encoding structure and rules

(2) Estimate the number of data combinations for different desensitization locations. Because the number of combinations is the inverse of the probability of data repetition, the higher the number of combinations, the lower the risk

(3) Sort the desensitization intensities of different desensitization locations according to the degree of risks

(4) The desensitization intensity is divided into three levels according to the risk span

Take mobile No. as an example. Figure 2 shows the encoding structure of mobile No. There are 11 digits, the

Table 5: Intensity ranking.

Intensity	Reserved bits	Desensitized bits	Example	Risk reference
1	Keep the top 2 & bottom 1	The middle 8	13088886666 − >13 ********6	0.06
2	Keep the top 3	The bottom 8	13088886666 − >130 ********	0.65
3	Keep the middle 4	The top 3 & bottom 4	13088886666 − > ** *8888****	0.78
4	Keep the bottom 4	The top 7	13088886666 − > ** *****6666	0.83
5	Keep the top 7	The bottom 4	13088886666 − >1308888 ****	0.96
6	Keep the top 3 & bottom 4	The middle 4	13088886666 − >130 ****6666	0.99

Table 6: Intensity grade.

Intensity grade	Intensity	Reserved bit	Desensitized bits	Example
High	1	Keep the top 2 & bottom 1	The middle 8	13088886666 − >13 ********6
	2	Keep the top 3	The bottom 8	13088886666 − >130 ********
Medium	3	Keep the middle 4	The top 3 & bottom 4	13088886666 − > ** *8888****
	4	Keep the bottom 4	The top 7	13088886666 − > ** *****6666
Low	5	Keep the top 7	The bottom 4	13088886666 − >1308888 ****
	6	Keep the top 3 & bottom 4	The middle 4	13088886666 − >130 ****6666

Table 7: Algorithmic attribute control table.

Attribute	Algorithm description	1	0
Integrity	Whether to keep the encoding structure intact	Y	N
Reality	Whether to reflect data real semantics	Y	N
Reliability	Is data deformation reversible? Random or quantitative?	Irreversible Random Keyless	Reversible Quantitative Keyed

Table 8: The desensitization algorithm weight vectors for mobile No.

Algorithm	Description	Example	Weighting
Mask	Use symbol "*" to replace parts of the data, with the data length unchanged	13088886666 − >130 ********	1, 0, 1
Floor	Take an integer	—	0, 0, 0[1]
Hashing	Map data into a fixed-length string	13088886666 − >abcdef	0, 0, 1
Truncation	Cut parts of the data	13088886666 − >130	0, 0, 1
Shift	Add a constant offset	13088886666 − >13088886670	1, 1, 0
Synthesis	Simulate new data to replace the original data	13088886666 − >13011007788	1, 1, 1
Rearrange	Sort a column of values upside-down	—	0, 0, 0[1]

[1]If the algorithm was not applicable for the current data (mobile No.), all the values should be set to 0.

top 3 are network identifier, the middle 4 are area code, and the bottom 3 are optional number for user. For relevant information, there are currently about 40 mobile network identifier codes in China, all starting with 1; area code is 3-digit free combination; maximum number of combination is 1000; user number is 4-digit free combination, and maximum number of combination is 10000.

Six desensitization locations were selected, and the desensitization intensity was sorted from high to low based on risk from low to high. Setting the algorithm to "mask," Table 5 was obtained.

According to the risk references in Table 5, the above 6 desensitization intensities were divided into 3 grades, namely, high intensity (row 1), medium intensity (row 2/3/4), and low intensity (row 5/6), as shown in Table 6.

2.2.2. Quantification of the Desensitization Algorithm Weight. The effects of data deformations under different

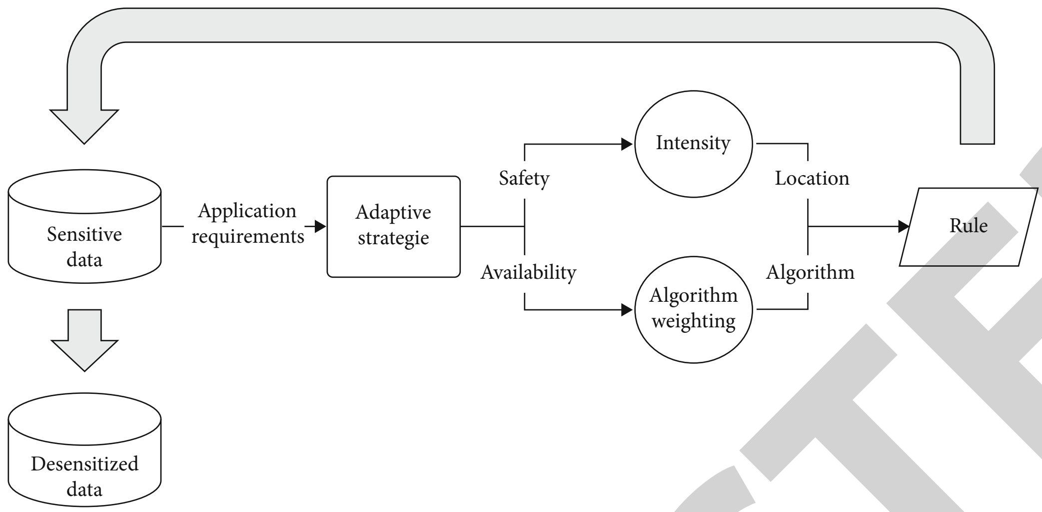

Figure 3: Adaptive desensitization strategy model.

Table 9: Encoding structures and rules of electric power sensitive data.

Sensitive data item	Encoding structure & rule	Example
Client's No.	10-digit serial number	0257349261
Mobile No.	11 digits, 3 – digit network identifier + 4 – digit area code + 4 – digit serial number	13088886666
Bank card No.	13-19 digits, issuing bank number + card type number + serial number	9558801202106562334
Electricity address	City + district/county + street/town + community/village + road + house number	16th Floor, No. 56, Huaqiao Road, Gulou District, Nanjing
Electricity consumption	Random number	250, 374, 499
Settlement date	8 digits, 4 – digit"year" + 2 – digit"month" + 2 – digit"day"	20210101

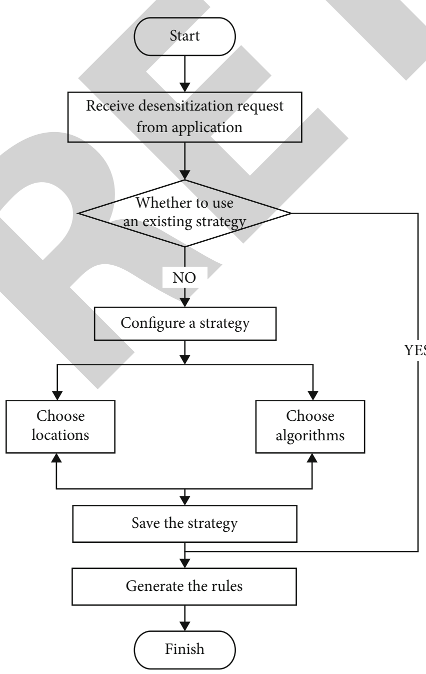

Figure 4: The desensitization strategy configuration procedure.

algorithms were evaluated based on the three dimensions of integrity, reality, and reliability in information security attributes. Assign a value of 0 or 1 to each attribute, respectively. Table 7 shows the comparison between algorithms and attributes.

Set the weight vector according to Table 7 for the common algorithms listed in Table 1 in 2.1. Taking mobile No. as an example, if the location was set to "the bottom 8 bits," Table 8 is obtained.

2.2.3. Configuration of the Desensitization Strategy. Since desensitization intensity was related to the likelihood of occurrence of risk, it would be considered that desensitization intensity characterized the safety of desensitization results, while the algorithm weights characterized the availability of desensitization results.

The desensitization intensity and desensitization algorithm weight were set as the configuration items of the desensitization strategy. User analyzed the expected desensitization results of the application requirements, and set the above items, respectively. The system filtered the conditions from the intensity grade tables and the algorithm weight vector tables, found the suitable desensitization locations and

Table 10: Dynamic rules for the application.

Rule	Intensity	Algorithm	Result
Client's No.	Keep the top 4	Shift	0257349261 – >0257886496
Mobile No.	Keep the middle 4	Hiding	13088886666 – >18988880000
Bank card No.	Keep the top 4 & bottom 4	Synthesis	9558801202106562334 – >9558987123645452334
Electricity address	Keep "city" & "district/county" & "house No."	Synthesis	16th Floor, No. 56, Huaqiao Road, Gulou District, Nanjing- > 16th Floor, No. 1 Hubei Road, Gulou District, Nanjing
Electricity consumption	1	Rearrange	250, 374, 499 – >499, 250, 374
Settlement date	Keep the top 8	Shift	20210101 – >20210108

[1]For random numerical type sensitive data, because there was no coding structure limit, the deformation effects were mainly affected by the algorithm. So the default desensitization intensity was equal to the configured value.

desensitization algorithms, and finally generated the desensitization rules by combining them. This was called adaptive desensitization strategy. The model is shown in Figure 3.

3. Results

Assume that the data were applied for the payment function test on the Middle Platform of the State Grid Jiangsu Electric Power Company.

A data set was extracted from Middle Platform, and the data to be desensitized includes client's No., mobile No., bank card No., electricity address, electricity consumption, and settlement date. Table 9 shows the encoding structures and rules for these sensitive data [10].

Configure the desensitization strategy following the procedure in Figure 4.

For this test requirements, a unified intensity strategy was adopted, while the data were required to maintain true semantics. Therefore, the desensitization intensities of all data were configured "medium intensity"; the algorithm weight required integrity and reality values of 1; reliability was not required; that is, the algorithm weight vectors were $[1, 1, 1]$ or $[1, 1, 0]$.

The results of filtering the above 6 kinds of sensitive data's intensity grade tables [11] and algorithm weight vector tables were combined into dynamic rules Table 10.

To sum up, a desensitization strategy was configured for the payment test application, which could be named and saved by "payment function testing strategy" and added to the strategy library.

4. Discussion

It can be found from the previous section that more than one matching result may be obtained when filtering the desensitization location and desensitization algorithm according to the desensitization strategy configuration items. At this time, it can be selected according to user preferences and subsequently prioritized by machine learning. Or, according to the method described in the article, add other attributes such as timeliness and data volume threshold to the strategy configuration items, and gradually improve the desensitization strategy configuration system with the expansion of the application requirements.

Starting from the desensitization results, the article analyzed the factors affecting the desensitization results in the desensitization process and established a desensitization strategy configuration model based on the desensitization intensity and the desensitization algorithm weigh. The desensitization strategy was closely related to the application requirements, so the desensitization strategy was configurable so that it could not be constrained by the fixed desensitization rules and could efficiently serve the application requirements by expanding the algorithm library at any time. Middle Platform is aimed at driving business development with data development, and its data services would continue to generate diversified desensitization needs. Data desensitization based on adaptive strategies was a well-adapted method for use in Middle Platform.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Wang Jijun was responsible for conceptualization, methodology, formal analysis, supervision, project administration, and funding acquisition. Xu Mingsheng and Lu Kang were responsible for validation. Xu Mingsheng was responsible for investigation, resources, and writing original draft preparation. Lu Kang was responsible for data curation. Lu Kang was responsible for writing, review, and editing. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This research was funded by Research and Application of Data Security Protection Technology Based on Middle Platform, grant number 5210ED2100L. The APC was funded by 5210ED2100L.

References

[1] "InfoQ," November 2021, https://www.infoq.cn/video/u8wDVqmU63E9b6WFOKml.

[2] Gartner, "An international authoritative IT research and consulting company," *Market Guide for Data Masking*, 2019.

[3] Cyberspace Administration of China, *Measures for data security management (draft for comment)*, Cyberspace Administration of China, Beijing, China, 2019.

[4] "HuaweiCloud," November 2021, https://support.huaweicloud.com/usermanual-dsc/dsc_01_0022.html.

[5] K. Benitez and B. Malin, "Evaluating re-identification risks with respect to the HIPAA privacy rule," *Journal of the American Medical Informatics Association*, vol. 17, no. 2, pp. 169–177, 2010.

[6] F. K. Dankar, K. El Emam, A. Neisa, and T. Roffey, "Estimating the re-identification risk of clinical data sets," *BMC Medical Informatics and Decision Making*, vol. 12, no. 1, p. 66, 2012.

[7] V. Janmey and P. L. Elkin, "Re-identification risk in HIPAA de-identified datasets: the MVA attack," *AMIA Annual Symposium Proceedings*, vol. 2018, article 1329, 2018.

[8] China Electricity Council Standardization Management Center, *Implementation specification for power data masking (draft for comment)*, China Electricity Council Standardization Management Center, Beijing, China, 2020.

[9] "Tencent Cloud," November 2021, https://cloud.tencent.com/developer/article/1636078.

[10] P. Ajay, B. Nagaraj, R. Arun Kumar, R. Huang, and P. Ananthi, "Unsupervised hyperspectral microscopic image segmentation using deep embedded clustering algorithm," *Scanning*, vol. 2022, Article ID 1200860, 9 pages, 2022.

[11] G. Veselov, A. Tselykh, A. Sharma, and R. Huang, "Applications of artificial intelligence in evolution of smart cities and societies," *Informatica (Slovenia)*, vol. 45, no. 5, p. 603, 2021.