Qrlew: automatic differential privacy for SQL queries

Anonymous submission

Abstract

AAAI creates proceedings, working notes, and technical reports directly from electronic source furnished by the authors. To ensure that all papers in the publication have a uniform appearance, authors must adhere to the following instructions.

Useful links

PPAI

Last year papers: https://aaai-ppai23.github.io/#sp2 This year program: https://ppai-workshop.github.io/

Comparable open-source projects

- Paszke et al. 2017 - Automatic differentiation in PyTorch https://openreview.net/pdf?id=BJJsrmfCZ
- Frostig et al. 2018 - Compiling machine learning programs via high-level tracing https://mlsys.org/Conferences/2019/doc/2018/146.pdf

Comparable DP SQL papers

- Lessons Learned: Surveying the Practicality of Differential Privacy in the Industry (Garrido et al. 2022)
- Tumult Analytics: a robust, easy-to-use, scalable, and expressive framework for differential privacy (Berghel et al. 2022)
- Differentially Private SQL with Bounded User Contribution (Wilson et al. 2019)
- CHORUS: a Programming Framework for Building Scalable Differential Privacy Mechanisms (Johnson et al. 2020)
- Towards Practical Differential Privacy for SQL Queries (Johnson, Near, and Song 2018)

Introduction

In recent years, the importance of safeguarding privacy when dealing with personal data has continuously increased. Traditional anonymization techniques have proven vulnerable to re-identification, as demonstrated by numerous works (Archie et al. 2018; Dwork et al. 2017; Narayanan and Shmatikov 2008; Sweeney, Abu, and Winn 2013). The total cost of data breaches has also significantly increased (IBM 2023). And governments have introduced stricter data protection laws. Yet, the collection, sharing, and utilization of data hold the potential to generate significant value across various industries, including healthcare, finance, transportation, and energy distribution.

To realize these benefits while managing privacy risks, researchers have turned to *differential privacy (DP)* (Wood et al. 2018; Dwork, Roth et al. 2014), which has become the gold standard in academia since its introduction by Dwork et al. in 2006 (Dwork et al. 2006) due to its provable and automatic privacy guarantees.

Despite the availability of open-source tools, DP adoption remains limited. One of the reasons for this lack of adoption is the relative complexity of the existing tools considered the utility of the results. *Qrlew* has been designed to solve this problem, by providing the following features:

Leverages existing infrastructure *Qrlew* rewrites a SQL query into a *differentially private* equivalent.

Is fully automated *Qrlew* can rewrite a large class of queries into *differentially private* ones.

Leverages synthetic data using jointly *differentially private* mechanisms and *differentially private* Synthetic Data

This In summary, our main contributions are as follows:

Paul on compilation

Victoria on DP mech and DP test

Comparison to other systems

Known limitations

Qrlew relies on the random number generator of the SQL engine used. It is usually not a cryptographic noise.

Qrlew uses the floating-point numbers of the host SQL engine, therefore our system is liable to the vulnerabilities described in

Preparing an Anonymous Submission

This document details the formatting requirements for anonymous submissions. The requirements are the same as for camera ready papers but with a few notable differences:

- Anonymous submissions must not include the author names and affiliations. Write "Anonymous Submission" as the "sole author" and leave the affiliations empty.
- The PDF document's metadata should be cleared with a metadata-cleaning tool before submitting it. This is to prevent leaked information from revealing your identity.
- References must be anonymized whenever the reader can infer that they are to the authors' previous work.
- AAAI's copyright notice should not be included as a footer in the first page.
- Only the PDF version is required at this stage. No source versions will be requested, nor any copyright transfer form.

You can achieve all of the above by enabling the `submission` option when loading the `aaai24` package:

```
\documentclass[letterpaper]{article}
\usepackage[submission]{aaai24}
```

The remainder of this document are the original camera-ready instructions. Any contradiction of the above points ought to be ignored while preparing anonymous submissions.

Camera-Ready Guidelines

Congratulations on having a paper selected for inclusion in an AAAI Press proceedings or technical report! This document details the requirements necessary to get your accepted paper published using PDFLaTeX. If you are using Microsoft Word, instructions are provided in a different document. AAAI Press does not support any other formatting software.

The instructions herein are provided as a general guide for experienced LaTeX users. If you do not know how to use LaTeX, please obtain assistance locally. AAAI cannot provide you with support and the accompanying style files are **not** guaranteed to work. If the results you obtain are not in accordance with the specifications you received, you must correct your source file to achieve the correct result.

These instructions are generic. Consequently, they do not include specific dates, page charges, and so forth. Please consult your specific written conference instructions for details regarding your submission. Please review the entire document for specific instructions that might apply to your particular situation. All authors must comply with the following:

- You must use the 2024 AAAI Press LaTeX style file and the aaai24.bst bibliography style files, which are located in the 2024 AAAI Author Kit (aaai24.sty, aaai24.bst).
- You must complete, sign, and return by the deadline the AAAI copyright form (unless directed by AAAI Press to use the AAAI Distribution License instead).
- You must read and format your paper source and PDF according to the formatting instructions for authors.
- You must submit your electronic files and abstract using our electronic submission form **on time.**
- You must pay any required page or formatting charges to AAAI Press so that they are received by the deadline.
- You must check your paper before submitting it, ensuring that it compiles without error, and complies with the guidelines found in the AAAI Author Kit.

Copyright

All papers submitted for publication by AAAI Press must be accompanied by a valid signed copyright form. They must also contain the AAAI copyright notice at the bottom of the first page of the paper. There are no exceptions to these requirements. If you fail to provide us with a signed copyright form or disable the copyright notice, we will be unable to publish your paper. There are **no exceptions** to this policy. You will find a PDF version of the AAAI copyright form in the AAAI AuthorKit. Please see the specific instructions for your conference for submission details.

Formatting Requirements in Brief

We need source and PDF files that can be used in a variety of ways and can be output on a variety of devices. The design and appearance of the paper is strictly governed by the aaai style file (aaai24.sty). **You must not make any changes to the aaai style file, nor use any commands, packages, style files, or macros within your own paper that alter that design, including, but not limited to spacing, floats, margins, fonts, font size, and appearance.** AAAI imposes requirements on your source and PDF files that must be followed. Most of these requirements are based on our efforts to standardize conference manuscript properties and layout. All papers submitted to AAAI for publication will be recompiled for standardization purposes. Consequently, every paper submission must comply with the following requirements:

- Your .tex file must compile in PDFLaTeX — (you may not include .ps or .eps figure files.)
- All fonts must be embedded in the PDF file — including your figures.
- Modifications to the style file, whether directly or via commands in your document may not ever be made, most especially when made in an effort to avoid extra page charges or make your paper fit in a specific number of pages.
- No type 3 fonts may be used (even in illustrations).
- You may not alter the spacing above and below captions, figures, headings, and subheadings.
- You may not alter the font sizes of text elements, footnotes, heading elements, captions, or title information (for references and mathematics, please see the limited exceptions provided herein).
- You may not alter the line spacing of text.
- Your title must follow Title Case capitalization rules (not sentence case).
- LaTeX documents must use the Times or Nimbus font package (you may not use Computer Modern for the text of your paper).
- No LaTeX 209 documents may be used or submitted.

- Your source must not require use of fonts for non-Roman alphabets within the text itself. If your paper includes symbols in other languages (such as, but not limited to, Arabic, Chinese, Hebrew, Japanese, Thai, Russian and other Cyrillic languages), you must restrict their use to bit-mapped figures. Fonts that require non-English language support (CID and Identity-H) must be converted to outlines or 300 dpi bitmap or removed from the document (even if they are in a graphics file embedded in the document).
- Two-column format in AAAI style is required for all papers.
- The paper size for final submission must be US letter without exception.
- The source file must exactly match the PDF.
- The document margins may not be exceeded (no overfull boxes).
- The number of pages and the file size must be as specified for your event.
- No document may be password protected.
- Neither the PDFs nor the source may contain any embedded links or bookmarks (no hyperref or navigator packages).
- Your source and PDF must not have any page numbers, footers, or headers (no pagestyle commands).
- Your PDF must be compatible with Acrobat 5 or higher.
- Your LaTeX source file (excluding references) must consist of a **single** file (use of the "input" command is not allowed.
- Your graphics must be sized appropriately outside of LaTeX (do not use the "clip" or "trim" command) .

If you do not follow these requirements, your paper will be returned to you to correct the deficiencies.

What Files to Submit

You must submit the following items to ensure that your paper is published:

- A fully-compliant PDF file.
- Your LaTeX source file submitted as a **single** .tex file (do not use the "input" command to include sections of your paper — every section must be in the single source file). (The only allowable exception is .bib file, which should be included separately).
- The bibliography (.bib) file(s).
- Your source must compile on our system, which includes only standard LaTeX 2020 TeXLive support files.
- Only the graphics files used in compiling paper.
- The LaTeX-generated files (e.g. .aux, .bbl file, PDF, etc.).

Your LaTeX source will be reviewed and recompiled on our system (if it does not compile, your paper will be returned to you. **Do not submit your source in multiple text files.** Your single LaTeX source file must include all your text, your bibliography (formatted using aaai24.bst), and any custom macros.

Your files should work without any supporting files (other than the program itself) on any computer with a standard LaTeX distribution.

Do not send files that are not actually used in the paper. Avoid including any files not needed for compiling your paper, including, for example, this instructions file, unused graphics files, style files, additional material sent for the purpose of the paper review, intermediate build files and so forth.

Obsolete style files. The commands for some common packages (such as some used for algorithms), may have changed. Please be certain that you are not compiling your paper using old or obsolete style files.

Final Archive. Place your source files in a single archive which should be compressed using .zip. The final file size may not exceed 10 MB. Name your source file with the last (family) name of the first author, even if that is not you.

Using LaTeX to Format Your Paper

The latest version of the AAAI style file is available on AAAI's website. Download this file and place it in the TeX search path. Placing it in the same directory as the paper should also work. You must download the latest version of the complete AAAI Author Kit so that you will have the latest instruction set and style file.

Document Preamble

In the LaTeX source for your paper, you **must** place the following lines as shown in the example in this subsection. This command set-up is for three authors. Add or subtract author and address lines as necessary, and uncomment the portions that apply to you. In most instances, this is all you need to do to format your paper in the Times font. The helvet package will cause Helvetica to be used for sans serif. These files are part of the PSNFSS2e package, which is freely available from many Internet sites (and is often part of a standard installation).

Leave the setcounter for section number depth commented out and set at 0 unless you want to add section numbers to your paper. If you do add section numbers, you must uncomment this line and change the number to 1 (for section numbers), or 2 (for section and subsection numbers). The style file will not work properly with numbering of subsubsections, so do not use a number higher than 2.

Your paper must compile in PDFLaTeX. Consequently, all your figures must be .jpg, .png, or .pdf. You may not use the .gif (the resolution is too low), .ps, or .eps file format for your figures.

Figures, drawings, tables, and photographs should be placed throughout the paper on the page (or the subsequent page) where they are first discussed. Do not group them together at the end of the paper. If placed at the top of the paper, illustrations may run across both columns. Figures must not invade the top, bottom, or side margin areas. Figures must be inserted using the \usepackage{graphicx}. Number figures sequentially, for example, figure 1, and so on. Do not use minipage to group figures.

If you normally create your figures using pgfplots, please create the figures first, and then import them as pdfs with proper bounding boxes, as the bounding and trim boxes created by pfgplots are fragile and not valid.

When you include your figures, you must crop them **outside** of LaTeX. The command \includegraphics*[clip=true, viewport 0 0 10 10]... might result in a PDF that looks great, but the image is **not really cropped.** The full image can reappear (and obscure whatever it is overlapping) when page numbers are applied or color space is standardized. Figures **??**, and **??** display some unwanted results that often occur.

If your paper includes illustrations that are not compatible with PDFTeX (such as .eps or .ps documents), you will need to convert them. The epstopdf package will usually work for eps files. You will need to convert your ps files to PDF in either case.

Figure Captions. The illustration number and caption must appear *under* the illustration. Labels and other text with the actual illustration must be at least nine-point type. However, the font and size of figure captions must be 10 point roman. Do not make them smaller, bold, or italic. (Individual words may be italicized if the context requires differentiation.)

Tables

Tables should be presented in 10 point roman type. If necessary, they may be altered to 9 point type. You may not use any commands that further reduce point size below nine points. Tables that do not fit in a single column must be placed across double columns. If your table won't fit within the margins even when spanning both columns, you must split it. Do not use minipage to group tables.

Table Captions. The number and caption for your table must appear *under* (not above) the table. Additionally, the font and size of table captions must be 10 point roman and must be placed beneath the figure. Do not make them smaller, bold, or italic. (Individual words may be italicized if the context requires differentiation.)

Low-Resolution Bitmaps. You may not use low-resolution (such as 72 dpi) screen-dumps and GIF files—these files contain so few pixels that they are always blurry, and illegible when printed. If they are color, they will become an indecipherable mess when converted to black and white. This is always the case with gif files, which should never be used. The resolution of screen dumps can be increased by reducing the print size of the original file while retaining the same number of pixels. You can also enlarge files by manipulating them in software such as PhotoShop. Your figures should be 300 dpi when incorporated into your document.

LaTeX Overflow. LaTeX users please beware: LaTeX will sometimes put portions of the figure or table or an equation in the margin. If this happens, you need to make the figure or table span both columns. If absolutely necessary, you may reduce the figure, or reformat the equation, or reconfigure the table. **Check your log file!** You must fix any overflow into the margin (that means no overfull boxes in LaTeX). **Nothing is permitted to intrude into the margin or gutter.**

Using Color. Use of color is restricted to figures only. It must be WACG 2.0 compliant. (That is, the contrast ratio must be greater than 4.5:1 no matter the font size.) It must be CMYK, NOT RGB. It may never be used for any portion of the text of your paper. The archival version of your paper will be printed in black and white and grayscale. The web version must be readable by persons with disabilities. Consequently, because conversion to grayscale can cause undesirable effects (red changes to black, yellow can disappear, and so forth), we strongly suggest you avoid placing color figures in your document. If you do include color figures, you must (1) use the CMYK (not RGB) colorspace and (2) be mindful of readers who may happen to have trouble distinguishing colors. Your paper must be decipherable without using color for distinction.

Drawings. We suggest you use computer drawing software (such as Adobe Illustrator or, (if unavoidable), the drawing tools in Microsoft Word) to create your illustrations. Do not use Microsoft Publisher. These illustrations will look best if all line widths are uniform (half- to two-point in size), and you do not create labels over shaded areas. Shading should be 133 lines per inch if possible. Use Times Roman or Helvetica for all figure call-outs. **Do not use hairline width lines** — be sure that the stroke width of all lines is at least .5 pt. Zero point lines will print on a laser printer, but will completely disappear on the high-resolution devices used by our printers.

Photographs and Images. Photographs and other images should be in grayscale (color photographs will not reproduce well; for example, red tones will reproduce as black, yellow may turn to white, and so forth) and set to a minimum of 300 dpi. Do not prescreen images.

Resizing Graphics. Resize your graphics **before** you include them with LaTeX. You may **not** use trim or clip options as part of your \includegraphics command. Resize the media box of your PDF using a graphics program instead.

Fonts in Your Illustrations. You must embed all fonts in your graphics before including them in your LaTeX document.

Algorithms. Algorithms and/or programs are a special kind of figures. Like all illustrations, they should appear floated to the top (preferably) or bottom of the page. However, their caption should appear in the header, left-justified and enclosed between horizontal lines, as shown in Algorithm 1. The algorithm body should be terminated with another horizontal line. It is up to the authors to decide whether to show line numbers or not, how to format comments, etc.

In LaTeX algorithms may be typeset using the `algorithm` and `algorithmic` packages, but you can also use one of the many other packages for the task.

Listings. Listings are much like algorithms and programs. They should also appear floated to the top (preferably) or bottom of the page. Listing captions should appear in the header, left-justified and enclosed between horizontal lines as shown in Listing 1. Terminate the body with another hor-

Algorithm 1: Example algorithm

Input: Your algorithm's input
Parameter: Optional list of parameters
Output: Your algorithm's output

```
Let t = 0.
while condition do
    Do some action.
    if conditional then
        Perform task A.
    else
        Perform task B.
    end if
end while
return solution
```

Listing 1: Example listing `quicksort.hs`

```
quicksort :: Ord a => [a] -> [a]
quicksort []     = []
quicksort (p:xs) = (quicksort lesser) ++
      [p] ++ (quicksort greater)
  where
    lesser  = filter (< p) xs
    greater = filter (>= p) xs
```

izontal line and avoid any background color. Line numbers, if included, must appear within the text column.

References

The AAAI style includes a set of definitions for use in formatting references with BibTeX. These definitions make the bibliography style fairly close to the ones specified in the Reference Examples appendix below. To use these definitions, you also need the BibTeX style file "aaai24.bst," available in the AAAI Author Kit on the AAAI web site. Then, at the end of your paper but before \enddocument, you need to put the following lines:

\bibliography{bibfile1,bibfile2,...}

Please note that the aaai24.sty class already sets the bibliographystyle for you, so you do not have to place any \bibliographystyle command in the document yourselves. The aaai24.sty file is incompatible with the hyperref and navigator packages. If you use either, your references will be garbled and your paper will be returned to you.

References may be the same size as surrounding text. However, in this section (only), you may reduce the size to \small if your paper exceeds the allowable number of pages. Making it any smaller than 9 point with 10 point linespacing, however, is not allowed. A more precise and exact method of reducing the size of your references minimally is by means of the following command:

\fontsize{9.8pt}{10.8pt} \selectfont

You must reduce the size equally for both font size and line spacing, and may not reduce the size beyond {9.0pt}{10.0pt}.

The list of files in the \bibliography command should be the names of your BibTeX source files (that is, the .bib files referenced in your paper).

The following commands are available for your use in citing references:

cite: Cites the given reference(s) with a full citation. This appears as "(Author Year)" for one reference, or "(Author Year; Author Year)" for multiple references.

shortcite: Cites the given reference(s) with just the year. This appears as "(Year)" for one reference, or "(Year; Year)" for multiple references.

citeauthor: Cites the given reference(s) with just the author name(s) and no parentheses.

citeyear: Cites the given reference(s) with just the date(s) and no parentheses.

You may also use any of the *natbib* citation commands.

Proofreading Your PDF

Please check all the pages of your PDF file. The most commonly forgotten element is the acknowledgements — especially the correct grant number. Authors also commonly forget to add the metadata to the source, use the wrong reference style file, or don't follow the capitalization rules or comma placement for their author-title information properly. A final common problem is text (expecially equations) that runs into the margin. You will need to fix these common errors before submitting your file.

Improperly Formatted Files

In the past, AAAI has corrected improperly formatted files submitted by the authors. Unfortunately, this has become an increasingly burdensome expense that we can no longer absorb). Consequently, if your file is improperly formatted, it will be returned to you for correction.

Naming Your Electronic File

We require that you name your LaTeX source file with the last name (family name) of the first author so that it can easily be differentiated from other submissions. Complete file-naming instructions will be provided to you in the submission instructions.

Submitting Your Electronic Files to AAAI

Instructions on paper submittal will be provided to you in your acceptance letter.

Inquiries

If you have any questions about the preparation or submission of your paper as instructed in this document, please contact AAAI Press at the address given below. If you have technical questions about implementation of the aaai style file, please contact an expert at your site. We do not provide technical support for LaTeX or any other software package. To avoid problems, please keep your paper simple, and do not incorporate complicated macros and style files.

AAAI Press
1900 Embarcadero Road, Suite 101

Palo Alto, California 94303-3310 USA
Telephone: (650) 328-3123
E-mail: See the submission instructions for your particular conference or event.

Additional Resources

LaTeX is a difficult program to master. If you've used that software, and this document didn't help or some items were not explained clearly, we recommend you read Michael Shell's excellent document (testflow doc.txt V1.0a 2002/08/13) about obtaining correct PS/PDF output on LaTeX systems. (It was written for another purpose, but it has general application as well). It is available at www.ctan.org in the tex-archive.

Reference Examples

For the most up to date version of the AAAI reference style, please consult the *AI Magazine* Author Guidelines at https://aaai.org/ojs/index.php/aimagazine/about/submissions#authorGuidelines

Acknowledgments

AAAI is especially grateful to Peter Patel Schneider for his work in implementing the original aaai.sty file, liberally using the ideas of other style hackers, including Barbara Beeton. We also acknowledge with thanks the work of George Ferguson for his guide to using the style and BibTeX files — which has been incorporated into this document — and Hans Guesgen, who provided several timely modifications, as well as the many others who have, from time to time, sent in suggestions on improvements to the AAAI style. We are especially grateful to Francisco Cruz, Marc Pujol-Gonzalez, and Mico Loretan for the improvements to the BibTeX and LaTeX files made in 2020.

The preparation of the LaTeX and BibTeX files that implement these instructions was supported by Schlumberger Palo Alto Research, AT&T Bell Laboratories, Morgan Kaufmann Publishers, The Live Oak Press, LLC, and AAAI Press. Bibliography style changes were added by Sunil Issar. \pubnote was added by J. Scott Penberthy. George Ferguson added support for printing the AAAI copyright slug. Additional changes to aaai24.sty and aaai24.bst have been made by Francisco Cruz and Marc Pujol-Gonzalez.

Thank you for reading these instructions carefully. We look forward to receiving your electronic files!

References

Archie, M.; Gershon, S.; Katcoff, A.; and Zeng, A. 2018. Who's watching? de-anonymization of netflix reviews using amazon reviews.

Berghel, S.; Bohannon, P.; Desfontaines, D.; Estes, C.; Haney, S.; Hartman, L.; Hay, M.; Machanavajjhala, A.; Magerlein, T.; Miklau, G.; et al. 2022. Tumult Analytics: a robust, easy-to-use, scalable, and expressive framework for differential privacy. *arXiv preprint arXiv:2212.04133*.

Dwork, C.; McSherry, F.; Nissim, K.; and Smith, A. 2006. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*, 265–284. Springer.

Dwork, C.; Roth, A.; et al. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4): 211–407.

Dwork, C.; Smith, A.; Steinke, T.; and Ullman, J. 2017. Exposed! a survey of attacks on private data. *Annual Review of Statistics and Its Application*, 4: 61–84.

Garrido, G. M.; Liu, X.; Matthes, F.; and Song, D. 2022. Lessons learned: Surveying the practicality of differential privacy in the industry. *arXiv preprint arXiv:2211.03898*.

IBM. 2023. Cost of a Data Breach Report 2023.

Johnson, N.; Near, J. P.; Hellerstein, J. M.; and Song, D. 2020. Chorus: a programming framework for building scalable differential privacy mechanisms. In *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*, 535–551. IEEE.

Johnson, N.; Near, J. P.; and Song, D. 2018. Towards Practical Differential Privacy for SQL Queries. *Proc. VLDB Endow.*, 11(5): 526–539.

Narayanan, A.; and Shmatikov, V. 2008. Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*, 111–125. IEEE.

Sweeney, L.; Abu, A.; and Winn, J. 2013. Identifying participants in the personal genome project by name (a re-identification experiment). *arXiv preprint arXiv:1304.7605*.

Wilson, R. J.; Zhang, C. Y.; Lam, W.; Desfontaines, D.; Simmons-Marengo, D.; and Gipson, B. 2019. Differentially private SQL with bounded user contribution. *arXiv preprint arXiv:1909.01917*.

Wood, A.; Altman, M.; Bembenek, A.; Bun, M.; Gaboardi, M.; Honaker, J.; Nissim, K.; O'Brien, D. R.; Steinke, T.; and Vadhan, S. 2018. Differential privacy: A primer for a non-technical audience. *Vand. J. Ent. & Tech. L.*, 21: 209.