PAS: Privacy Algorithms in Systems

Philip Yu
University of Illinois Chicago
psyu@uic.edu

Olivera Kotevska
Oak Ridge National Laboratory
kotevskao@ornl.gov

Tyler Derr
Vanderbilt University
tyler.derr@vanderbilt.edu

ABSTRACT

Today we face an explosion of data generation, ranging from health monitoring to national security infrastructure systems. More and more systems are connected to the Internet that collects data at regular time intervals. These systems share data and use machine learning methods for intelligent decisions, which resulted in numerous real-world applications (e.g., autonomous vehicles, recommendation systems, and heart-rate monitoring) that have benefited from it. However, these approaches are prone to identity thief and other privacy related cyber-security attacks. So, how can data privacy be protected efficiently in these scenarios? More dedicated efforts are needed to propose the integration of privacy techniques into existing systems and develop more advanced privacy techniques to address the complex challenges of multi-system connectivity and data fusion. Therefore, we have introduced Privacy Algorithms in Systems (PAS) at CIKM which provides a venue to gather academic researchers and industry researchers/practitioners to present their research in an effort to advance the frontier of this critical direction of privacy algorithms in systems.

KEYWORDS

privacy algorithms, differential privacy, machine learning, systems

ACM Reference Format:
Philip Yu, Olivera Kotevska, and Tyler Derr. 2022. PAS: Privacy Algorithms in Systems. In *Proceedings of the 31th ACM International Conference on Information and Knowledge Management (CIKM '22), October 17–21, 2022, Hybrid Event, Atlanta, GA, U.S.A. (Conference acronym 'PAS).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 WORKSHOP THEME AND TOPICS

Recently privacy has been widely used in many scientific domains and everyday communications and systems. Even privacy regulations have been developed on a national and international level. The topic of privacy affects all kinds of complex datasets such as text, video, audio, image, streaming, and graph format. They are produced by surveillance systems, home management systems, user tracking devices, data storage, communication systems, social networks, etc. Machine learning/decision making algorithmic systems have been widely adopted to facilitate efficient systems. As our ability to generate and collect data constantly increases unprecedentedly, the complex data we are facing in the modern era are becoming more and more diverse and large-scale. This raises privacy concerns for users, systems, and infrastructure, leading to more efforts to develop effective privacy preservation algorithms and deploy them efficiently for real-world applications.

This workshop aims to discuss the recent research progress of privacy algorithms in both theoretical foundations and practical applications for systems. We invite submissions that focus on recent advances in research and development of privacy algorithms and their applications. Theory and methodology papers are welcome from any of the following areas, including but not limited to:

- Theory of privacy algorithms (e.g., differential privacy, local differential privacy, pan-privacy, data anonymization)
- Privacy preservation of complex data (e.g., image, text, video, audio, streaming data, graph data) and data sharing
- Privacy preservation decision making algorithms, machine and federated learning, transfer and semi-supervised learning, meta-learning
- Privacy preservation in deep learning models (e.g., convolutional neural networks, graph neural networks, recurrent neural networks, transformer-based networks, etc.)
- Benchmark analysis of privacy algorithms
- Relation between privacy guarantee, fairness, and bias
- Metrics for data privacy
- Implementation of privacy policies and regulations in privacy algorithms
- Privacy attacks on complex data and methods

and application papers focused on but not limited to: recommender systems, computer vision, natural language processing, biomedical, healthcare, insurance, cybersecurity, financial security, consumer protection, transportation/mobility networks, along with cloud, edge, and HPC systems.

2 WORKSHOP OBJECTIVES AND EXPECTED OUTCOME

Our main objective is to bring together and strengthen the privacy community at CIKM while also building connections with domain experts to establish innovative new research direction and industry applications. We also seek to provide a unified venue for those interested in these topics and get more privacy researchers to CIKM to increase CIKM's position in this domain. Data privacy is becoming a must-have in many systems, so creating a community where experts can share their knowledge is the goal of this workshop, and CIKM is a perfect venue since the focus is on knowledge management and data. This workshop is an excellent opportunity for students to present their work and network with their peers and experts. We also seek to provide a diverse and welcoming environment across the organizing committee, keynote speakers, panellists, and participating students.

Conference acronym 'PAS, October 17–21, 2022, Atlanta, GA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

3 WORKSHOP PROGRAM

The workshop is hybrid where at least two of the organizers are going to attend the workshop in person to ensure everything runs smooth. The workshop will be advertised to our local institutions, our collaborator networks, relevant universities, industry, and on social platforms. The full day workshop schedule is planned with two half-day sessions. Our program will consist of the following main components:

- Invited keynotes from experts in the field of privacy algorithms coming from both industry and academia to create a synergistic atmosphere and to stimulate collaborations.
- Contributed research oral talks selected from the set of accepted works into PAS.
- Future directions panel discussion that will be composed of our keynote speakers given their expertise in this domain.
- Contributed poster sessions both before lunch and after the final remarks to allow all those with works accepted into the workshop to present their work and socialize stimulating new ideas and potential collaborations.

Panel. Our anticipation is to have a panel as the final component of the workshop before the poster session, best paper award ceremony, and final remarks. The panellists will be a subset of our keynote speakers and organizers and will focus on the future directions in privacy algorithms (in systems). We believe this is the most fitting formal final component of the workshop and hope that it will help stimulate conversions at the end during the post session consisting of contributed research to the workshop.

3.1 Invited Keynote Speakers

We have a mixture of world experts in privacy from academia and industry. They are the following in alphabetical order.

- Graham Cormode
  - Professor of Computer Science at University of Warwick and Research Scientist at Meta.
- James Honaker
  - Research Scientist at Harvard University and Meta.
- James Joshi
  - Professor of Computer Science at University of Pittsburgh.
- Dan Lin
  - Professor of Computer Science at Vanderbilt University.
- Harsha Nori
  - Senior Data Scientist at Microsoft.
- Norman Sadeh
  - Professor of Computer Science at Carnegie Mellon University

4 WORKSHOP ORGANIZATION

4.1 Workshop Co-Chairs

Philip S. Yu is a Distinguished Professor and Wexler Chair in Information Technology at University of Chicago. Dr. Yu received his Ph.D. in E.E. from Stanford University and M.B.A. from New York University. His main research interests include big data, data mining, privacy preserving data, data streams, and Internet applications and technologies. Dr. Yu has over 150,000 citations and an h-index of 176. Furthermore, he is a Fellow of the ACM and IEEE. He is the recipient of ACM SIHKDD 2016 Innovation Award for his influential research and scientific contributions on mining, fusion and anonymization of big data, the IEEE Computer Society's 2013 Technical Achievement Award for "pioneering and fundamentally innovate contributions to scalable indexing, querying, searching, mining and anonymization of big data", and the Research Contribution Award from IEEE ICDM in 2003 for his pioneering contributions to the field of data mining. More details can he found at: https://cs.uic.edu/profiles/philip-yu/

Olivera Kotevska is a Research Scientist at Oak Ridge National Laboratory (ORNL) in Computer Science and Mathematics Division (CSMD). She received her Ph.D. in Computer Science from University of Grenoble Alpes, France and was part of a joint program with National Institute of Standards and Technology, USA. Her research is in privacy algorithms and machine learning for energy and national security applications. She regularly serves as a program committee member and reviewer in these domains and served in organizational roles including co-chair of IEEE Big Data Industry and Government Program, co-chair of IEEE Power and Energy Society, Computational Analytical Methods Subcommittee, and co-editor of Sensors journal special issue IoT Data Analytics. She is an organizer and chair of IEEE WiE East TN affinity group. She received IEEE Senior membership award in '21 and ORNL CSMD Outreach and Service award in '20.

Tyler Derr is an Assistant Professor at Vanderbilt University in the Department of Computer Science and Data Science Institute. He received his PhD in Computer Science from Michigan State University in 2020. His research is generally in data mining and machine learning, with emphasis on social network analysis, deep learning on graphs, and data science for social good with recent directions in fairness and privacy in graph-structured data. He has published and regularly serves as a SPC/PC member at the top conferences in these domains and served in organizational roles including Publicity Co-Chair of KDD'22, Doctoral Consortium Co-Chair of WSDM'22, Proceedings Co-Chair of KDD'21, and co-organized multiple workshops including Machine Learning on Graphs (MLoG) at WSDM'22 and ICDM'22. He received numerous esteemed awards, such as the Best Student Poster Award at SDM'19, Vanderbilt's Fall'20 Teaching Innovation Award from the School of Engineering, and Best Reviewer Award at ICWSM'19/'21. More details can be found at: http://www.TylerDerr.com

4.2 Additional Workshop Organizers

- Proceedings Chair: Dr. Chris Stanley
  - Research Scientist at Oak Ridge National Laboratory
- Web Chair: Yuying Zhao
  - PhD Student at Vanderbilt University

5 ACKNOWLEDGMENT

This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE 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).