An Efficient Two-Layer Mechanism for Privacy-Preserving Truth Discovery

Type Conference Paper
Author Yaliang Li
Author Chenglin Miao
Author Lu Su
Author Jing Gao
Author Qi Li
Author Bolin Ding
Author Zhan Qin
Author Kui Ren
URL http://doi.acm.org/10.1145/3219819.3219998
Series KDD '18
Place New York, NY, USA
Publisher ACM
Pages 1705–1714
ISBN 978-1-4503-5552-0
Date 2018
Extra event-place: London, United Kingdom
DOI 10.1145/3219819.3219998
Accessed 2.4.2019, 13:12:34
Library Catalog ACM Digital Library
Abstract Soliciting answers from online users is an efficient and effective solution to many challenging tasks. Due to the variety in the quality of users, it is important to infer their ability to provide correct answers during aggregation. Therefore, truth discovery methods can be used to automatically capture the user quality and aggregate user-contributed answers via a weighted combination. Despite the fact that truth discovery is an effective tool for answer aggregation, existing work falls short of the protection towards the privacy of participating users. To fill this gap, we propose perturbation-based mechanisms that provide users with privacy guarantees and maintain the accuracy of aggregated answers. We first present a one-layer mechanism, in which all the users adopt the same probability to perturb their answers. Aggregation is then conducted on perturbed answers but the aggregation accuracy could drop accordingly. To improve the utility, a two-layer mechanism is proposed where users are allowed to sample their own probabilities from a hyper distribution. We theoretically compare the one-layer and two-layer mechanisms, and prove that they provide the same privacy guarantee while the two-layer mechanism delivers better utility. This advantage is brought by the fact that the two-layer mechanism can utilize the estimated user quality information from truth discovery to reduce the accuracy loss caused by

perturbation, which is confirmed by experimental results on real-world datasets. Experimental results also demonstrate the effectiveness of the proposed two-layer mechanism in privacy protection with tolerable accuracy loss in aggregation.

Proceedings Title Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining

Date Added 2.4.2019, 13:12:34

Modified 2.4.2019, 13:12:34

Tags:

differential privacy, truth discovery, two-layer mechanism

Attachments

- ACM Full Text PDF

Bringing Salary Transparency to the World: Computing Robust Compensation Insights via LinkedIn Salary

Type Journal Article

Author Krishnaram Kenthapadi

Author Stuart Ambler

Author Liang Zhang

Author Deepak Agarwal

URL http://arxiv.org/abs/1703.09845

Publication arXiv:1703.09845 [cs]

Date 2017-03-28

Extra arXiv: 1703.09845

DOI 10.1145/3132847.3132863

Accessed 2.4.2019, 13:03:41

Library Catalog arXiv.org

Abstract The recently launched LinkedIn Salary product has been designed with the goal of providing compensation insights to the world's professionals and thereby helping them optimize their earning potential. We describe the overall design and architecture of the statistical modeling system underlying this product. We focus on the unique data mining challenges while designing and implementing the system, and describe the modeling components such as Bayesian hierarchical smoothing that help to compute and present robust compensation insights to users. We report on extensive evaluation with nearly one year of de-identified compensation data collected from over one million LinkedIn users, thereby demonstrating the efficacy of the statistical models. We also highlight the lessons learned through the deployment of our system at LinkedIn.

Short Title Bringing Salary Transparency to the World
Date Added 2.4.2019, 13:03:42
Modified 2.4.2019, 13:03:42

Tags:

Computer Science - Artificial Intelligence, Computer Science - Information Retrieval,

Computer Science - Social and Information Networks

Notes:

Comment: Conference information: ACM International Conference on Information and Knowledge Management (CIKM 2017)

Attachments

- arXiv:1703.09845 PDF
- arXiv.org Snapshot

De-anonymizing Social Networks

Type Conference Paper
Author A. Narayanan
Author V. Shmatikov
Pages 173-187
Date May 2009
DOI 10.1109/SP.2009.22
Library Catalog IEEE Xplore
Conference Name 2009 30th IEEE Symposium on Security and Privacy
Abstract Operators of online social networks are increasingly sharing potentially sensitive information about users and their relationships with advertisers, application developers, and data-mining researchers. Privacy is typically protected by anonymization, i.e., removing names, addresses, etc.We present a framework for analyzing privacy and anonymity in social networks and develop a new re-identification algorithm targeting anonymized social-network graphs. To demonstrate its effectiveness on real-world networks, we show that a third of the users who can be verified to have accounts on both Twitter, a popular microblogging service, and Flickr, an online photo-sharing site, can be re-identified in the anonymous Twitter graph with only a 12% error rate.Our de-anonymization algorithm is based purely on the network topology, does not require creation of a large number of dummy "sybil" nodes, is robust to noise and all existing defenses, and works even when the overlap between the target network and the adversary's auxiliary information is small.

Proceedings Title 2009 30th IEEE Symposium on Security and Privacy
Date Added 2.4.2019, 12:56:22
Modified 2.4.2019, 12:56:22

Tags:

anonymity, anonymized social network graphs, application developers, Companies, data mining, data mining researchers, data privacy, Data privacy, de-anonymizing social networks, Facebook, graph theory, network topology, Peer to peer computing, privacy, Privacy, re-identification algorithm, social networking (online), social networks

Attachments

- IEEE Xplore Abstract Record
- Submitted Version

De-anonymizing Web Browsing Data with Social Networks

Type Conference Paper
Author Jessica Su
Author Ansh Shukla
Author Sharad Goel
Author Arvind Narayanan
URL https://doi.org/10.1145/3038912.3052714
Series WWW '17
Place Republic and Canton of Geneva, Switzerland
Publisher International World Wide Web Conferences Steering Committee
Pages 1261–1269
ISBN 978-1-4503-4913-0
Date 2017
Extra event-place: Perth, Australia
DOI 10.1145/3038912.3052714
Accessed 2.4.2019, 13:26:31
Library Catalog ACM Digital Library
Abstract Can online trackers and network adversaries de-anonymize web browsing data readily available to them? We show---theoretically, via simulation, and through experiments on real user data---that de-identified web browsing histories can be linked to social media profiles using only publicly available data. Our approach is based on a simple observation: each person has a distinctive social network, and thus the set of links appearing in one's feed is unique. Assuming users visit links in their feed with higher probability than a random user, browsing histories contain

tell-tale marks of identity. We formalize this intuition by specifying a model of web browsing behavior and then deriving the maximum likelihood estimate of a user's social profile. We evaluate this strategy on simulated browsing histories, and show that given a history with 30 links originating from Twitter, we can deduce the corresponding Twitter profile more than 50% of the time.To gauge the real-world effectiveness of this approach, we recruited nearly 400 people to donate their web browsing histories, and we were able to correctly identify more than 70% of them. We further show that several online trackers are embedded on sufficiently many websites to carry out this attack with high accuracy. Our theoretical contribution applies to any type of transactional data and is robust to noisy observations, generalizing a wide range of previous de-anonymization attacks. Finally, since our attack attempts to find the correct Twitter profile out of over 300 million candidates, it is---to our knowledge---the largest-scale demonstrated de-anonymization to date.

Proceedings Title Proceedings of the 26th International Conference on World Wide Web
Date Added 2.4.2019, 13:26:31
Modified 2.4.2019, 13:26:31

Tags:

de-anonymization, deanonymization, privacy, social network, social networking, social networks, twitter

Attachments

- ACM Full Text PDF

Deep Learning with Differential Privacy

Type Conference Paper
Author Martin Abadi
Author Andy Chu
Author Ian Goodfellow
Author H. Brendan McMahan
Author Ilya Mironov
Author Kunal Talwar
Author Li Zhang
URL http://doi.acm.org/10.1145/2976749.2978318
Series CCS '16
Place New York, NY, USA
Publisher ACM
Pages 308–318
ISBN 978-1-4503-4139-4

Date	2016
Extra	event-place: Vienna, Austria
DOI	10.1145/2976749.2978318
Accessed	2.4.2019, 13:21:38
Library Catalog	ACM Digital Library
Abstract	Machine learning techniques based on neural networks are achieving remarkable results in a wide variety of domains. Often, the training of models requires large, representative datasets, which may be crowdsourced and contain sensitive information. The models should not expose private information in these datasets. Addressing this goal, we develop new algorithmic techniques for learning and a refined analysis of privacy costs within the framework of differential privacy. Our implementation and experiments demonstrate that we can train deep neural networks with non-convex objectives, under a modest privacy budget, and at a manageable cost in software complexity, training efficiency, and model quality.
Proceedings Title	Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security
Date Added	2.4.2019, 13:21:38
Modified	2.4.2019, 13:21:38

Tags:

deep learning, differential privacy

Attachments

- ACM Full Text PDF

Geographic Differential Privacy for Mobile Crowd Coverage Maximization

Type	Journal Article
Author	Leye Wang
Author	Gehua Qin
Author	Dingqi Yang
Author	Xiao Han
Author	Xiaojuan Ma
URL	http://arxiv.org/abs/1710.10477
Publication	arXiv:1710.10477 [cs]
Date	2017-10-28
Extra	arXiv: 1710.10477
Accessed	2.4.2019, 13:31:19

Library Catalog arXiv.org

Abstract For real-world mobile applications such as location-based advertising and spatial crowdsourcing, a key to success is targeting mobile users that can maximally cover certain locations in a future period. To find an optimal group of users, existing methods often require information about users' mobility history, which may cause privacy breaches. In this paper, we propose a method to maximize mobile crowd's future location coverage under a guaranteed location privacy protection scheme. In our approach, users only need to upload one of their frequently visited locations, and more importantly, the uploaded location is obfuscated using a geographic differential privacy policy. We propose both analytic and practical solutions to this problem. Experiments on real user mobility datasets show that our method significantly outperforms the state-of-the-art geographic differential privacy methods by achieving a higher coverage under the same level of privacy protection.

Date Added 2.4.2019, 13:31:19

Modified 2.4.2019, 13:31:19

Tags:

Computer Science - Cryptography and Security

Attachments

- arXiv:1710.10477 PDF
- arXiv.org Snapshot

How Public Is My Private Life?: Privacy in Online Dating

Type Conference Paper

Author Camille Cobb

Author Tadayoshi Kohno

URL https://doi.org/10.1145/3038912.3052592

Series WWW '17

Place Republic and Canton of Geneva, Switzerland

Publisher International World Wide Web Conferences Steering Committee

Pages 1231–1240

ISBN 978-1-4503-4913-0

Date 2017

Extra event-place: Perth, Australia

DOI 10.1145/3038912.3052592

Accessed 2.4.2019, 13:40:23

Library Catalog ACM Digital Library

Abstract Online dating services let users expand their dating pool beyond their social network and specify important characteristics of potential partners. To assess compatibility, users share personal information -- e.g., identifying details or sensitive opinions about sexual preferences or worldviews -- in profiles or in one-on-one communication. Thus, participating in online dating poses inherent privacy risks. How people reason about these privacy risks in modern online dating ecosystems has not been extensively studied. We present the results of a survey we designed to examine privacy-related risks, practices, and expectations of people who use or have used online dating, then delve deeper using semi-structured interviews. We additionally analyzed 400 Tinder profiles to explore how these issues manifest in practice. Our results reveal tensions between privacy and competing user values and goals, and we demonstrate how these results can inform future designs.

Proceedings Title Proceedings of the 26th International Conference on World Wide Web
Short Title How Public Is My Private Life?
Date Added 2.4.2019, 13:40:23
Modified 2.4.2019, 13:40:23

Tags:

human values, information disclosure, online dating, privacy, security

Attachments

- ACM Full Text PDF

Identity-Based Remote Data Integrity Checking With Perfect Data Privacy Preserving for Cloud Storage

Type Journal Article
Author Y. Yu
Author M. H. Au
Author G. Ateniese
Author X. Huang
Author W. Susilo
Author Y. Dai
Author G. Min
Volume 12
Issue 4
Pages 767-778
Publication IEEE Transactions on Information Forensics and Security
ISSN 1556-6013
Date April 2017

DOI 10.1109/TIFS.2016.2615853
Library Catalog IEEE Xplore
Abstract Remote data integrity checking (RDIC) enables a data storage server, say a cloud server, to prove to a verifier that it is actually storing a data owner's data honestly. To date, a number of RDIC protocols have been proposed in the literature, but most of the constructions suffer from the issue of a complex key management, that is, they rely on the expensive public key infrastructure (PKI), which might hinder the deployment of RDIC in practice. In this paper, we propose a new construction of identity-based (ID-based) RDIC protocol by making use of key-homomorphic cryptographic primitive to reduce the system complexity and the cost for establishing and managing the public key authentication framework in PKI-based RDIC schemes. We formalize ID-based RDIC and its security model, including security against a malicious cloud server and zero knowledge privacy against a third party verifier. The proposed ID-based RDIC protocol leaks no information of the stored data to the verifier during the RDIC process. The new construction is proven secure against the malicious server in the generic group model and achieves zero knowledge privacy against a verifier. Extensive security analysis and implementation results demonstrate that the proposed protocol is provably secure and practical in the real-world applications.
Date Added 2.4.2019, 13:20:29
Modified 2.4.2019, 13:20:29

Tags:

cloud computing, Cloud computing, cloud storage, Cloud storage, cryptographic protocols, data integrity, data privacy, Data privacy, data storage server, ID-based RDIC protocol, identity-based cryptography, identity-based RDIC protocol, identity-based remote data integrity checking, key-homomorphic cryptographic primitive, PKI, privacy preserving, Protocols, Public key, public key cryptography, public key infrastructure, security analysis, Servers, system complexity, zero knowledge privacy

Attachments

- IEEE Xplore Abstract Record
- Submitted Version

K-anonymity: A Model for Protecting Privacy

Type Journal Article
Author Latanya Sweeney
URL http://dx.doi.org/10.1142/S0218488502001648

Volume	10
Issue	5
Pages	557–570
Publication	Int. J. Uncertain. Fuzziness Knowl.-Based Syst.
ISSN	0218-4885
Date	October 2002
DOI	10.1142/S0218488502001648
Accessed	2.4.2019, 12:55:34
Library Catalog	ACM Digital Library
Abstract	Consider a data holder, such as a hospital or a bank, that has a privately held collection of person-specific, field structured data. Suppose the data holder wants to share a version of the data with researchers. How can a data holder release a version of its private data with scientific guarantees that the individuals who are the subjects of the data cannot be re-identified while the data remain practically useful? The solution provided in this paper includes a formal protection model named k-anonymity and a set of accompanying policies for deployment. A release provides k-anonymity protection if the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appears in the release. This paper also examines re-identification attacks that can be realized on releases that adhere to k- anonymity unless accompanying policies are respected. The k-anonymity protection model is important because it forms the basis on which the real-world systems known as Datafly, μ-Argus and k-Similar provide guarantees of privacy protection.
Short Title	K-anonymity
Date Added	2.4.2019, 12:55:34
Modified	2.4.2019, 12:55:34

Tags:

data anonymity, data fusion, data privacy, privacy, re-identification

Attachments

- Submitted Version

Location Privacy-Preserving Task Allocation for Mobile Crowdsensing with Differential Geo-Obfuscation

Type	Conference Paper
Author	Leye Wang
Author	Dingqi Yang
Author	Xiao Han

Author Tianben Wang
Author Daqing Zhang
Author Xiaojuan Ma
URL https://doi.org/10.1145/3038912.3052696
Series WWW '17
Place Republic and Canton of Geneva, Switzerland
Publisher International World Wide Web Conferences Steering Committee
Pages 627–636
ISBN 978-1-4503-4913-0
Date 2017
Extra event-place: Perth, Australia
DOI 10.1145/3038912.3052696
Accessed 2.4.2019, 13:36:13
Library Catalog ACM Digital Library
Abstract In traditional mobile crowdsensing applications, organizers need participants' precise locations for optimal task allocation, e.g., minimizing selected workers' travel distance to task locations. However, the exposure of their locations raises privacy concerns. Especially for those who are not eventually selected for any task, their location privacy is sacrificed in vain. Hence, in this paper, we propose a location privacy-preserving task allocation framework with geo-obfuscation to protect users' locations during task assignments. Specifically, we make participants obfuscate their reported locations under the guarantee of differential privacy, which can provide privacy protection regardless of adversaries' prior knowledge and without the involvement of any third-part entity. In order to achieve optimal task allocation with such differential geo-obfuscation, we formulate a mixed-integer non-linear programming problem to minimize the expected travel distance of the selected workers under the constraint of differential privacy. Evaluation results on both simulation and real-world user mobility traces show the effectiveness of our proposed framework. Particularly, our framework outperforms Laplace obfuscation, a state-of-the-art differential geo-obfuscation mechanism, by achieving 45% less average travel distance on the real-world data.
Proceedings Title Proceedings of the 26th International Conference on World Wide Web
Date Added 2.4.2019, 13:36:13
Modified 2.4.2019, 13:36:13

Tags:

crowdsensing, differential location privacy, task allocation

Attachments

- ACM Full Text PDF

PATE-GAN: Generating Synthetic Data with Differential Privacy Guarantees

Type	Journal Article
Author	James Jordon
Author	Jinsung Yoon
Author	Mihaela van der Schaar
URL	https://openreview.net/forum?id=S1zk9iRqF7
Date	2018/09/27
Accessed	2.4.2019, 13:24:31
Library Catalog	openreview.net
Abstract	Machine learning has the potential to assist many communities in using the large datasets that are becoming more and more available. Unfortunately, much of that potential is not being realized...
Short Title	PATE-GAN
Date Added	2.4.2019, 13:24:31
Modified	2.4.2019, 13:24:31

Attachments

- Full Text PDF
- Snapshot

Practical Secure Aggregation for Privacy-Preserving Machine Learning

Type	Conference Paper
Author	Keith Bonawitz
Author	Vladimir Ivanov
Author	Ben Kreuter
Author	Antonio Marcedone
Author	H. Brendan McMahan
Author	Sarvar Patel
Author	Daniel Ramage
Author	Aaron Segal
Author	Karn Seth
URL	http://doi.acm.org/10.1145/3133956.3133982
Series	CCS '17
Place	New York, NY, USA
Publisher	ACM
Pages	1175–1191

ISBN	978-1-4503-4946-8
Date	2017
Extra	event-place: Dallas, Texas, USA
DOI	10.1145/3133956.3133982
Accessed	2.4.2019, 13:19:53
Library Catalog	ACM Digital Library
Abstract	We design a novel, communication-efficient, failure-robust protocol for secure aggregation of high-dimensional data. Our protocol allows a server to compute the sum of large, user-held data vectors from mobile devices in a secure manner (i.e. without learning each user's individual contribution), and can be used, for example, in a federated learning setting, to aggregate user-provided model updates for a deep neural network. We prove the security of our protocol in the honest-but-curious and active adversary settings, and show that security is maintained even if an arbitrarily chosen subset of users drop out at any time. We evaluate the efficiency of our protocol and show, by complexity analysis and a concrete implementation, that its runtime and communication overhead remain low even on large data sets and client pools. For 16-bit input values, our protocol offers $1.73 x communication expansion for 210 users and 220-dimensional vectors, and 1.98 x expansion for 214 users and 224-dimensional vectors over sending data in the clear.
Proceedings Title	Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security
Date Added	2.4.2019, 13:19:53
Modified	2.4.2019, 13:19:53

Tags:

federated learning, machine learning, privacy-preserving protocols, secure aggregation

Attachments

- ACM Full Text PDF

Privacy Preserving Average Consensus

Type	Journal Article
Author	Y. Mo
Author	R. M. Murray
Volume	62
Issue	2
Pages	753-765
Publication	IEEE Transactions on Automatic Control
ISSN	0018-9286

Date February 2017
DOI 10.1109/TAC.2016.2564339
Library Catalog IEEE Xplore
Abstract Average consensus is a widely used algorithm for distributed computing and control, where all the agents in the network constantly communicate and update their states in order to achieve an agreement. This approach could result in an undesirable disclosure of information on the initial state of an agent to the other agents. In this paper, we propose a privacy preserving average consensus algorithm to guarantee the privacy of the initial state and asymptotic consensus on the exact average of the initial values, by adding and subtracting random noises to the consensus process. We characterize the mean square convergence rate of our consensus algorithm and derive the covariance matrix of the maximum likelihood estimate on the initial state. Moreover, we prove that our proposed algorithm is optimal in the sense that it does not disclose any information more than necessary to achieve the average consensus. A numerical example is provided to illustrate the effectiveness of the proposed design.
Date Added 2.4.2019, 13:21:12
Modified 2.4.2019, 13:21:12

Tags:

Algorithm design and analysis, algorithm theory, asymptotic consensus, covariance matrices, covariance matrix, Databases, distributed computing, Estimation, Heuristic algorithms, maximum likelihood estimate, maximum likelihood estimation, Maximum likelihood estimation, mean square convergence rate, multi-agent systems, networked control systems, privacy, Privacy, privacy preserving average consensus algorithm, Signal processing algorithms, Symmetric matrices

Attachments

- IEEE Xplore Abstract Record

Privacy-preserving Class Ratio Estimation

Type Conference Paper
Author Arun Shankar Iyer
Author J. Saketha Nath
Author Sunita Sarawagi
URL http://doi.acm.org/10.1145/2939672.2939806
Series KDD '16
Place New York, NY, USA
Publisher ACM

Pages	925–934
ISBN	978-1-4503-4232-2
Date	2016
Extra	event-place: San Francisco, California, USA
DOI	10.1145/2939672.2939806
Accessed	2.4.2019, 12:47:53
Library Catalog	ACM Digital Library
Abstract	In this paper we present learning models for the class ratio estimation problem, which takes as input an unlabeled set of instances and predicts the proportions of instances in the set belonging to the different classes. This problem has applications in social and commercial data analysis. Existing models for class-ratio estimation however require instance-level supervision. Whereas in domains like politics, and demography, set-level supervision is more common. We present a new method for directly estimating class-ratios using set-level supervision. Another serious limitation in applying these techniques to sensitive domains like health is data privacy. We propose a novel label privacy-preserving mechanism that is well-suited for supervised class ratio estimation and has guarantees for achieving efficient differential privacy, provided the per-class counts are large enough. We derive learning bounds for the estimation with and without privacy constraints, which lead to important insights for the data-publisher. Extensive empirical evaluation shows that our model is more accurate than existing methods and that the proposed privacy mechanism and learning model are well-suited for each other.
Proceedings Title	Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining
Date Added	2.4.2019, 12:47:54
Modified	2.4.2019, 12:47:54

Tags:

differential privacy, kernel method, learning theory, maximum mean discrepancy, statistical estimation

Attachments

- ACM Full Text PDF

Privacy-Preserving Distributed Multi-Task Learning with Asynchronous Updates

Type	Conference Paper
Author	Liyang Xie
Author	Inci M. Baytas

Author	Kaixiang Lin
Author	Jiayu Zhou
URL	http://doi.acm.org/10.1145/3097983.3098152
Series	KDD '17
Place	New York, NY, USA
Publisher	ACM
Pages	1195–1204
ISBN	978-1-4503-4887-4
Date	2017
Extra	event-place: Halifax, NS, Canada
DOI	10.1145/3097983.3098152
Accessed	2.4.2019, 13:13:26
Library Catalog	ACM Digital Library
Abstract	Many data mining applications involve a set of related learning tasks. Multi-task learning (MTL) is a learning paradigm that improves generalization performance by transferring knowledge among those tasks. MTL has attracted so much attention in the community, and various algorithms have been successfully developed. Recently, distributed MTL has also been studied for related tasks whose data is distributed across different geographical regions. One prominent challenge of the distributed MTL frameworks is to maintain the privacy of the data. The distributed data may contain sensitive and private information such as patients' records and registers of a company. In such cases, distributed MTL frameworks are required to preserve the privacy of the data. In this paper, we propose a novel privacy-preserving distributed MTL framework to address this challenge. A privacy-preserving proximal gradient algorithm, which asynchronously updates models of the learning tasks, is introduced to solve a general class of MTL formulations. The proposed asynchronous approach is robust against network delays and provides a guaranteed differential privacy through carefully designed perturbation. Theoretical guarantees of the proposed algorithm are derived and supported by the extensive experimental results.
Proceedings Title	Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining
Date Added	2.4.2019, 13:13:26
Modified	2.4.2019, 13:13:26

Tags:

asynchronous proximal optimization, differential privacy, multi-task learning

Attachments

- ACM Full Text PDF

Privacy-preserving outsourced classification in cloud computing

Type	Journal Article
Author	Ping Li
Author	Jin Li
Author	Zhengan Huang
Author	Chong-Zhi Gao
Author	Wen-Bin Chen
Author	Kai Chen
URL	https://doi.org/10.1007/s10586-017-0849-9
Volume	21
Issue	1
Pages	277-286
Publication	Cluster Computing
ISSN	1573-7543
Date	2018-03-01
Journal Abbr	Cluster Comput
DOI	10.1007/s10586-017-0849-9
Accessed	2.4.2019, 13:18:59
Library Catalog	Springer Link
Language	en
Abstract	Classifier has been widely applied in machine learning, such as pattern recognition, medical diagnosis, credit scoring, banking and weather prediction. Because of the limited local storage at user side, data and classifier has to be outsourced to cloud for storing and computing. However, due to privacy concerns, it is important to preserve the confidentiality of data and classifier in cloud computing because the cloud servers are usually untrusted. In this work, we propose a framework for privacy-preserving outsourced classification in cloud computing (POCC). Using POCC, an evaluator can securely train a classification model over the data encrypted with different public keys, which are outsourced from the multiple data providers. We prove that our scheme is secure in the semi-honest model
Date Added	2.4.2019, 13:18:59
Modified	2.4.2019, 13:18:59

Tags:

Classification, Cryptography, Homomorphic encryption, Machine learning, Privacy-preserving

Attachments

- Springer Full Text PDF

RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response

Type	Conference Paper
Author	Úlfar Erlingsson
Author	Vasyl Pihur
Author	Aleksandra Korolova
URL	http://doi.acm.org/10.1145/2660267.2660348
Series	CCS '14
Place	New York, NY, USA
Publisher	ACM
Pages	1054–1067
ISBN	978-1-4503-2957-6
Date	2014
Extra	event-place: Scottsdale, Arizona, USA
DOI	10.1145/2660267.2660348
Accessed	2.4.2019, 12:57:05
Library Catalog	ACM Digital Library
Abstract	Randomized Aggregatable Privacy-Preserving Ordinal Response, or RAPPOR, is a technology for crowdsourcing statistics from end-user client software, anonymously, with strong privacy guarantees. In short, RAPPORs allow the forest of client data to be studied, without permitting the possibility of looking at individual trees. By applying randomized response in a novel manner, RAPPOR provides the mechanisms for such collection as well as for efficient, high-utility analysis of the collected data. In particular, RAPPOR permits statistics to be collected on the population of client-side strings with strong privacy guarantees for each client, and without linkability of their reports. This paper describes and motivates RAPPOR, details its differential-privacy and utility guarantees, discusses its practical deployment and properties in the face of different attack models, and, finally, gives results of its application to both synthetic and real-world data.
Proceedings Title	Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security
Short Title	RAPPOR
Date Added	2.4.2019, 12:57:05
Modified	2.4.2019, 12:57:05

Tags:

cloud computing, crowdsourcing, population statistics, privacy protection, statistical inference

Attachments

- ACM Full Text PDF

Robust De-anonymization of Large Sparse Datasets

Type	Conference Paper
Author	Arvind Narayanan
Author	Vitaly Shmatikov
URL	https://doi.org/10.1109/SP.2008.33
Series	SP '08
Place	Washington, DC, USA
Publisher	IEEE Computer Society
Pages	111–125
ISBN	978-0-7695-3168-7
Date	2008
DOI	10.1109/SP.2008.33
Accessed	2.4.2019, 13:00:29
Library Catalog	ACM Digital Library
Abstract	We present a new class of statistical de-anonymization attacks against high-dimensional micro-data, such as individual preferences, recommendations, transaction records and so on. Our techniques are robust to perturbation in the data and tolerate some mistakes in the adversary's background knowledge. We apply our de-anonymization methodology to the Netflix Prize dataset, which contains anonymous movie ratings of 500,000 subscribers of Netflix,the world's largest online movie rental service. We demonstrate that an adversary who knows only a little bit about an individual subscriber can easily identify this subscriber's record in the dataset. Using the Internet Movie Database as the source of background knowledge, we successfully identified the Netflix records of known users, uncovering their apparent political preferences and other potentially sensitive information.
Proceedings Title	Proceedings of the 2008 IEEE Symposium on Security and Privacy
Date Added	2.4.2019, 13:00:29
Modified	2.4.2019, 13:00:29

Tags:

Anonymity, Attack, Privacy

Semi-supervised Knowledge Transfer for Deep Learning from Private Training Data

Type	Journal Article

Author	Nicolas Papernot
Author	Martín Abadi
Author	Úlfar Erlingsson
Author	Ian Goodfellow
Author	Kunal Talwar
URL	https://openreview.net/forum?id=HkwoSDPgg&noteId=HkwoSDPgg
Date	2016/11/02
Accessed	2.4.2019, 13:27:29
Library Catalog	openreview.net
Abstract	Some machine learning applications involve training data that is sensitive, such as the medical histories of patients in a clinical trial. A model may inadvertently and implicitly store some of its...
Date Added	2.4.2019, 13:27:29
Modified	2.4.2019, 13:27:29

Attachments

- Full Text PDF
- Snapshot

The Onions Have Eyes: A Comprehensive Structure and Privacy Analysis of Tor Hidden Services

Type	Conference Paper
Author	Iskander Sanchez-Rola
Author	Davide Balzarotti
Author	Igor Santos
URL	https://doi.org/10.1145/3038912.3052657
Series	WWW '17
Place	Republic and Canton of Geneva, Switzerland
Publisher	International World Wide Web Conferences Steering Committee
Pages	1251–1260
ISBN	978-1-4503-4913-0
Date	2017
Extra	event-place: Perth, Australia
DOI	10.1145/3038912.3052657
Accessed	2.4.2019, 13:41:46
Library Catalog	ACM Digital Library
Abstract	Tor is a well known and widely used darknet, known for its anonymity. However, while its protocol and relay security have already been extensively studied, to date there is no comprehensive analysis of the

structure and privacy of its Web Hidden Service. To fill this gap, we developed a dedicated analysis platform and used it to crawl and analyze over 1.5M URLs hosted in 7257 onion domains. For each page we analyzed its links, resources, and redirections graphs, as well as the language and category distribution. According to our experiments, Tor hidden services are organized in a sparse but highly connected graph, in which around 10% of the onions sites are completely isolated. Our study also measures for the first time the tight connection that exists between Tor hidden services and the Surface Web. In fact, more than 20% of the onion domains we visited imported resources from the Surface Web, and links to the Surface Web are even more prevalent than to other onion domains. Finally, we measured for the first time the prevalence and the nature of web tracking in Tor hidden services, showing that, albeit not as widespread as in the Surface Web, tracking is notably present also in the Dark Web: more than 40% of the scripts are used for this purpose, with the 70% of them being completely new tracking scripts unknown by existing anti-tracking solutions.

Proceedings Title	Proceedings of the 26th International Conference on World Wide Web
Short Title	The Onions Have Eyes
Date Added	2.4.2019, 13:41:46
Modified	2.4.2019, 13:41:46

Tags:

browser security & privacy, dark web, privacy

Attachments

- ACM Full Text PDF

Towards Practical Differential Privacy for SQL Queries

Type	Journal Article
Author	Noah Johnson
Author	Joseph P. Near
Author	Dawn Song
URL	http://arxiv.org/abs/1706.09479
Publication	arXiv:1706.09479 [cs]
Date	2017-06-28
Extra	arXiv: 1706.09479
DOI	10.1145/3177732.3177733
Accessed	2.4.2019, 13:04:46
Library Catalog	arXiv.org

Abstract Differential privacy promises to enable general data analytics while protecting individual privacy, but existing differential privacy mechanisms do not support the wide variety of features and databases used in real-world SQL-based analytics systems. This paper presents the first practical approach for differential privacy of SQL queries. Using 8.1 million real-world queries, we conduct an empirical study to determine the requirements for practical differential privacy, and discuss limitations of previous approaches in light of these requirements. To meet these requirements we propose elastic sensitivity, a novel method for approximating the local sensitivity of queries with general equijoins. We prove that elastic sensitivity is an upper bound on local sensitivity and can therefore be used to enforce differential privacy using any local sensitivity-based mechanism. We build FLEX, a practical end-to-end system to enforce differential privacy for SQL queries using elastic sensitivity. We demonstrate that FLEX is compatible with any existing database, can enforce differential privacy for real-world SQL queries, and incurs negligible (0.03%) performance overhead.

Date Added 2.4.2019, 13:04:46

Modified 2.4.2019, 13:04:46

Tags:

Computer Science - Cryptography and Security, Computer Science - Databases

Notes:

Comment: Extended & updated from VLDB 2018 version

Attachments

- arXiv:1706.09479 PDF
- arXiv.org Snapshot

Trajectory Recovery From Ash: User Privacy Is NOT Preserved in Aggregated Mobility Data

Type Conference Paper

Author Fengli Xu

Author Zhen Tu

Author Yong Li

Author Pengyu Zhang

Author Xiaoming Fu

Author Depeng Jin

URL https://doi.org/10.1145/3038912.3052620

Series	WWW '17
Place	Republic and Canton of Geneva, Switzerland
Publisher	International World Wide Web Conferences Steering Committee
Pages	1241–1250
ISBN	978-1-4503-4913-0
Date	2017
Extra	event-place: Perth, Australia
DOI	10.1145/3038912.3052620
Accessed	2.4.2019, 13:40:56
Library Catalog	ACM Digital Library
Abstract	Human mobility data has been ubiquitously collected through cellular networks and mobile applications, and publicly released for academic research and commercial purposes for the last decade. Since releasing individual's mobility records usually gives rise to privacy issues, datasets owners tend to only publish aggregated mobility data, such as the number of users covered by a cellular tower at a specific timestamp, which is believed to be sufficient for preserving users' privacy. However, in this paper, we argue and prove that even publishing aggregated mobility data could lead to privacy breach in individuals' trajectories. We develop an attack system that is able to exploit the uniqueness and regularity of human mobility to recover individual's trajectories from the aggregated mobility data without any prior knowledge. By conducting experiments on two real-world datasets collected from both mobile application and cellular network, we reveal that the attack system is able to recover users' trajectories with accuracy about 73%~91% at the scale of tens of thousands to hundreds of thousands users, which indicates severe privacy leakage in such datasets. Through the investigation on aggregated mobility data, our work recognizes a novel privacy problem in publishing statistic data, which appeals for immediate attentions from both academy and industry.
Proceedings Title	Proceedings of the 26th International Conference on World Wide Web
Short Title	Trajectory Recovery From Ash
Date Added	2.4.2019, 13:40:56
Modified	2.4.2019, 13:40:56

Tags:

aggregated mobility data, statistic data privacy, trajectory privacy

Attachments

- ACM Full Text PDF