USCViterbi

CSCI599 Privacy in the World of Big Data

Units: 4
Spring 2016

Time: Mondays, 2-5:20pm
Location: VKC 108
URL: www-bcf.usc.edu/~korolova/teaching/CSCI599Privacy/

Instructor: Aleksandra Korolova
Office: SAL 206
Office Hours: Wed 4:30pm-5:30pm and by appointment
Contact Info: lastname@usc.edu

Teaching Assistant: Brendan Avent
Office: SAL 246
Office Hours: Thu 1:30pm-2:30pm and by appointment
Contact Info: b followed by lastname@usc.edu

Course Description

A graduate level introduction to the privacy challenges that arise as a result of ubiquitous use of technology, dropping data collection, storage, and analysis costs, and data-based technological innovation, as well as algorithmic and technological approaches to addressing them.

The first half of the course will focus on statistical data privacy – the problem of making useful inferences based on data of many individuals while ensuring that each individual's privacy is preserved. We will survey plausible-sounding approaches that fail to achieve this goal, followed by a study of privacy definitions and algorithms for achieving both privacy and utility (including in real-world applications such as publishing search logs and location traces, building recommender systems and telemetry for malware detection).

The second half of the course will survey the technical aspects of topics and technologies at the frontier of current privacy-related discourse, such as web (and other forms of) tracking, advertising, anonymity and surveillance, and algorithmic fairness.

Our aim is to explore cutting-edge research topics in privacy, with a balance between theory and practical applications. The final syllabus and list of topics will be tailored to the backgrounds and interests of enrolled students.

The course is geared toward graduate students who want to gain familiarity with privacy from a scientific perspective. Advanced undergraduates with sufficient mathematical maturity are welcome.

Learning Objectives

This course aims to introduce students to the topics and techniques at the forefront of modern privacy research, thereby contributing to students' ability for advancing the state-of-the-art in the field or addressing privacy-related challenges in their own research. By the end of the course, the students should be able to think critically about the privacy risks posed by collection, analysis or publication of data in various contexts, as well as be able to propose approaches to mitigating those risks.

Prerequisite(s): solid grasp of algorithms, proof-based mathematics, and basic probability

Course Notes

N/A

Technological Proficiency and Hardware/Software Required

N/A

Required Readings and Supplementary Materials

Links to PDF versions of all required readings will be posted on the class website.

Description and Assessment of Assignments

Reading assignments, problem sets, and responses to prompts.

Students will be expected to read the papers whose material is covered in class, and occasionally, one additional paper per week, and submit a response to the prompts provided for those papers. The responses will be graded on a scale of 0 – 3 (0: missing, 1: poor, 2: good, 3: excellent). The goal of assignments is to deepen the knowledge of the topic presented in class, and to practice thinking critically and constructively about privacy research.

Presentation of a research paper to the class.

Over the course of the semester, each student will be expected to give a 45 minute presentation of one of the assigned papers to the class. The presentation should cover the salient ideas of the paper and propose

directions for class discussion. The students are expected to prepare the draft of the presentation and attend office hours to discuss it with the instructor or TA at least 2 days before their chosen class. The goal of presentations is to practice identifying salient research contributions and communicating them.

Course Project.
A major part of the course assignments is a research project, to be done in groups of 2-3 students, that should address a problem related to privacy. This may include:

- Discovery and analysis of a new privacy vulnerability in a recently published dataset, app, or widely used system.
- A design of a protocol, algorithm, or system that improves prior work.
- An implementation or extension of a recently published work or system demonstrating previously unknown aspects of it.

Projects can be tailored to be more theoretical and more applied depending on student background and interests, and may be in any area of privacy, including those not directly covered in the course. The instructor will be available to help formulate project ideas and give feedback and suggestions on project direction throughout the semester.

You will give a project presentation during the last week of classes and submit a conference-style report during the final exam period.

Grading Breakdown

Assignment	% of Grade
Weekly reading assignment and responses to prompts	20%
Class participation	10%
Presentation of a research paper to the class	15%
Course project (proposal)	15%
Course project (progress report)	10%
Course project (presentation)	10%
Course project (final report)	20%

Assignment Submission Policy

Response to weekly reading assignments is to be emailed to the instructor by 2pm on Sunday. Good quality photos or scans of handwritten responses are acceptable.

Course project proposal (including problem definition, motivation, planned approach and evaluation) is to be submitted by email as a 2-3 page PDF by March 7; progress report – by April 4th; final report – by May 2nd. Project presentations are on April 25th.

Additional Policies

You are allowed a total of 4 late days per semester that can be used for weekly reading assignment responses or course project proposal, progress report or final report.

Course Schedule: A Weekly Breakdown

Week #	Topics	Readings
	PRIVACY VULNERABILITIES IN DATA PUBLISHING	
1	**Data Releases Gone Wrong**	"A Face is exposed for AOL Searcher No. 4417749", The New York Times, Aug 9, 2006
		Narayanan, A. and Shmatikov, V., 2008, May. Robust de-anonymization of large sparse datasets. In Security and Privacy, 2008. SP 2008. IEEE Symposium on (pp. 111-125). IEEE.
		Kumar, R., Novak, J., Pang, B. and Tomkins, A., 2007, May. On anonymizing query logs via token-based hashing. In Proceedings of the 16th international conference on World Wide Web (pp. 629-638). ACM.
		Backstrom, Lars, Cynthia Dwork, and Jon Kleinberg. "Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography." Proceedings of the 16th international conference on World Wide Web. ACM, 2007.
		Homer, Nils, et al. "Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays."PLoS Genet 4.8 (2008): e1000167.
		Narayanan, A., Shi, E. and Rubinstein, B.I., 2011, July. Link prediction by de-anonymization: How we won the kaggle social network challenge. In Neural Networks (IJCNN), The 2011 International Joint Conference on (pp. 1825-1834). IEEE.
		Vijay Pandurangan, On Taxis and Rainbows
1	**Data Uses That Leak Private Info**	Calandrino, J., Kilzer, A., Narayanan, A., Felten, E.W. and Shmatikov, V., 2011, May. " You Might Also Like:" Privacy Risks of Collaborative Filtering. InSecurity and Privacy (SP), 2011 IEEE Symposium on (pp. 231-246). IEEE.
		Korolova, A., 2010, December. Privacy violations using microtargeted ads: A case study. In Data Mining Workshops (ICDMW), 2010 IEEE International Conference on (pp. 474-482). IEEE.
	PRIVACY DEFINITIONS	
2	**k-anonymity, l-diversity, t-closeness**	Sweeney, L., 2002. k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(05), pp.557-570.
		Machanavajjhala, A., Kifer, D., Gehrke, J. and Venkitasubramaniam, M., 2007. l-diversity: Privacy beyond k-anonymity. ACM Transactions on Knowledge Discovery from Data (TKDD), 1(1), p.3.
		Golle, P., 2006, October. Revisiting the uniqueness of simple demographics in the US population. In Proceedings of the 5th ACM workshop on Privacy in electronic society (pp. 77-80). ACM.
		Li, N., Li, T. and Venkatasubramanian, S., 2007, April. t-closeness: Privacy beyond k-anonymity and l-diversity. In Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on (pp. 106-115). IEEE.
2	**Accuracy Limits on Private Query Answering**	Dinur, I. and Nissim, K., 2003, June. Revealing information while preserving privacy. In Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (pp. 202-210). ACM.
3	**Differential Privacy**	Dwork, C. and Roth, A., 2014. The Algorithmic Foundations of Differential Privacy

		Dwork, C., McSherry, F., Nissim, K. and Smith, A., 2006. Calibrating noise to sensitivity in private data analysis. In Theory of cryptography (pp. 265-284). Springer Berlin Heidelberg.
		Dwork, C., 2011. A firm foundation for private data analysis. Communications of the ACM, 54(1), pp.86-95.
	DIFFERENTIALLY PRIVATE MECHANISMS	
4	**Randomized Response, Laplace Mechanism, Exponential Mechanism**	Dwork, C. and Roth, A., 2014. The Algorithmic Foundations of Differential Privacy
4	**Composition theorems**	Dwork, C. and Roth, A., 2014. The Algorithmic Foundations of Differential Privacy
	USING DIFFERENTIAL PRIVACY FOR DATA RELEASES	
5	**Frequent Itemset Mining, Search Log Release**	Bhaskar, R., Laxman, S., Smith, A. and Thakurta, A., 2010, July. Discovering frequent patterns in sensitive data. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 503-512). ACM.
		Korolova, A., Kenthapadi, K., Mishra, N. and Ntoulas, A., 2009, April. Releasing search queries and clicks privately. In Proceedings of the 18th international conference on World wide web (pp. 171-180). ACM.
		Götz, M., Machanavajjhala, A., Wang, G., Xiao, X. and Gehrke, J., 2012. Publishing search logs—a comparative study of privacy guarantees. Knowledge and Data Engineering, IEEE Transactions on, 24(3), pp.520-532.
5	**Recommender Systems**	McSherry, F. and Mironov, I., 2009, June. Differentially private recommender systems: building privacy into the Netflix Prize Contenders. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 627-636). ACM.
		Kenthapadi, K., Korolova, A., Mironov, I. and Mishra, N., 2012. Privacy via the johnson-lindenstrauss transform. arXiv preprint arXiv:1204.2606.
6	**Location**	Andrés, M.E., Bordenabe, N.E., Chatzikokolakis, K. and Palamidessi, C., 2013, November. Geo-indistinguishability: Differential privacy for location-based systems. In Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security (pp. 901-914). ACM.
		Golle, P. and Partridge, K., 2009. On the anonymity of home/work location pairs. In Pervasive computing (pp. 390-397). Springer Berlin Heidelberg.
		Xiao, Y. and Xiong, L., 2015, October. Protecting locations with differential privacy under temporal correlations. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (pp. 1298-1309). ACM.
		Hien To, Liyue Fan, and Cyrus Shahabi, Differentially Private H-Tree, In proceeding of 2nd Workshop on Privacy in Geographic Information Collection and Analysis (GeoPrivacy 2015). In conjunction with ACM SIGSPATIAL , Seattle, Washington, USA, November 3-6, 2015
7	**The Local Model**	Erlingsson, Ú., Pihur, V. and Korolova, A., 2014. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. InProceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security.
		Fanti, G., Pihur, V. and Erlingsson, U., 2015. Building a RAPPOR with the Unknown: Privacy-Preserving Learning of Associations and Data Dictionaries. arXiv preprint

		arXiv:1503.01214.
		Bassily, R. and Smith, A., 2015. Local, Private, Efficient Protocols for Succinct Histograms. arXiv preprint arXiv:1504.04686.
7	**Optimizing utility for differential privacy**	Ghosh, A., Roughgarden, T. and Sundararajan, M., 2012. Universally utility-maximizing privacy mechanisms. SIAM Journal on Computing, 41(6), pp.1673-1693.
		Hay, M., Rastogi, V., Miklau, G. and Suciu, D., 2010. Boosting the accuracy of differentially private histograms through consistency. Proceedings of the VLDB Endowment, 3(1-2), pp.1021-1032.
		Fredrikson, M., Lantz, E., Jha, S., Lin, S., Page, D. and Ristenpart, T., 2014, August. Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing. In Proceedings of USENIX Security.
	ALTERNATE PRIVACY DEFINITIONS	
8	**Crowd-blending privacy**	Gehrke, J., Hay, M., Lui, E. and Pass, R., 2012. Crowd-blending privacy. InAdvances in Cryptology–CRYPTO 2012 (pp. 479-496). Springer Berlin Heidelberg.
8	**Pufferfish**	Kifer, D. and Machanavajjhala, A., 2012, May. A rigorous and customizable framework for privacy. In Proceedings of the 31st symposium on Principles of Database Systems (pp. 77-88). ACM.
9	**Contextual Integrity**	Nissenbaum, H., 2011. A contextual approach to privacy online. Daedalus,140(4), pp.32-48.
	DATA COLLECTION AND WEB TRACKING	
11	**Web Tracking**	New York Times. How Companies Learn Your Secrets
		WSJ. Julia Angwin. The Web's New Gold Mine: Your Secrets (First in the Wall Street Journal's What They Know series)
		Mayer et al.: Third-Party Web Tracking: Policy and Technology, Oakland 2012
11	**Fingerprinting**	Eckersley, P., 2010, January. How unique is your web browser?. In Privacy Enhancing Technologies (pp. 1-18). Springer Berlin Heidelberg.
		Nikiforakis et al.: Cookieless Monster: Exploring the Ecosystem of Web-based Device Fingerprinting, Oakland 2013
		Bojinov, H., Michalevsky, Y., Nakibly, G. and Boneh, D., 2014. Mobile device identification via sensor fingerprinting. arXiv preprint arXiv:1408.1416.
		Behavioral Profiling, Evercookie
12	**Metadata and Surveillance**	Englehardt, S., Reisman, D., Eubank, C., Zimmerman, P., Mayer, J., Narayanan, A. and Felten, E.W., 2015, May. Cookies That Give You Away: The Surveillance Implications of Web Tracking. In Proceedings of the 24th International Conference on World Wide Web (pp. 289-299). International World Wide Web Conferences Steering Committee.
		Mayer, J., Mutchler, P., MetaPhone: The NSA Three-Hop
12	**Beacons, Cross-device tracking, Connecting online and Offline activities**	Jinyan Zang, Krysta Dummit, James Graves, Paul Lisker, and Latanya Sweeney. Who Knows What About Me? A Survey of Behind the Scenes Personal Data Sharing to Third Parties by Mobile Apps

	CONSEQUENCES OF ONLINE SOCIAL AND INFORMATION SYSTEMS	
9	**Attribute Inference Using Social Networks**	You are who you know: Inferring User profiles in Online Social Networks
		Gaydar: Facebook friendships expose sexual orientation, C. Jernigan and B. Mistree, 2009
		To join or not to join: The illusion of privacy in social networks with mixed public and private user profiles, E. Zheleva and L. Getoor, 2009
		Homophily and Latent Attribute Inference: Inferring Latent Attributes of Twitter Users from Neighbors, F. Zamal, W. Lie and D. Ruths, 2012
10	**Discrimination, Fairness, Accountability**	Sweeney, L., 2013. Discrimination in online ad delivery. Queue, 11(3), p.10.
		Datta, A., Tschantz, M.C. and Datta, A., 2014. Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. arXiv preprint arXiv:1408.6491.
		A. Datta, Privacy through Accountability: A Computer Science Perspective, in Proceedings of 10th International Conference on Distributed Computing and Internet Technology
		Dwork, C., Hardt, M., Pitassi, T., Reingold, O. and Zemel, R., 2012, January. Fairness through awareness. In Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (pp. 214-226). ACM.
	PETS	
13	**Privacy-Preserving Advertising**	Guha, S., Cheng, B. and Francis, P., 2011, March. Privad: Practical Privacy in Online Advertising. In NSDI.
		Toubiana, V., Narayanan, A., Boneh, D., Nissenbaum, H. and Barocas, S., 2010, March. Adnostic: Privacy preserving targeted advertising. InProceedings Network and Distributed System Symposium.
13	**Tracking Protection**	DNT, Ghostery, AdNauseam, TrackMeNot
14	**Tor**	Dingledine, R., Mathewson, N. and Syverson, P., 2004. Tor: The second-generation onion router. Naval Research Lab Washington DC.
14	**Sensitive Data**	Peddinti, S.T., Korolova, A., Bursztein, E. and Sampemane, G., 2014, May. Cloak and swagger: Understanding data sensitivity through the lens of user anonymity. In Security and Privacy (SP), 2014 IEEE Symposium on (pp. 493-508). IEEE.
		Mathias Lecuyer, Guillaume Ducoffe, Francis Lan, Andrei Papancea, Theofilos Petsios, Riley Spahn, Augustin Chaintreau, and Roxana Geambasu. "XRay: Increasing the Web's Transparency with Differential Correlation." In Proceedings of the USENIX Security Symposium, San Diego, CA, August 2014
15	**PROJECT PRESENTATIONS**	

Statement on Academic Conduct and Support Systems

Academic Conduct

Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in *SCampus* in Section 11, *Behavior Violating University Standards*https://scampus.usc.edu/1100-behavior-violating-university-standards-and-appropriate-sanctions/. Other forms of academic dishonesty are equally unacceptable. See additional information in *SCampus* and university policies on scientific misconduct, http://policy.usc.edu/scientific-misconduct/.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the *Office of Equity and Diversity* http://equity.usc.edu/ or to the *Department of Public Safety* http://capsnet.usc.edu/department/department-public-safety/online-forms/contact-us. This is important for the safety whole USC community. Another member of the university community – such as a friend, classmate, advisor, or faculty member – can help initiate the report, or can initiate the report on behalf of another person. *The Center for Women and Men* http://www.usc.edu/student-affairs/cwm/ provides 24/7 confidential support, and the sexual assault resource center webpage sarc@usc.edu describes reporting options and other resources.

Support Systems

A number of USC's schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the *American Language Institute* http://dornsife.usc.edu/ali, which sponsors courses and workshops specifically for international graduate students. *The Office of Disability Services and Programs* http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.htmlprovides certification for students with disabilities and helps arrange the relevant accommodations. If an officially declared emergency makes travel to campus infeasible, *USC Emergency Information http://emergency.usc.edu/*will provide safety and other updates, including ways in which instruction will be continued by means of blackboard, teleconferencing, and other technology.