INTRODUCING PRIVACY TO UNDERGRADUATE COMPUTING STUDENTS

Marina Moore, Maximilian Zinkus, Nathan Lemay, Zachary Peterson, and Bruce DeBruhl
Computer Science and Software Engineering Department
California Polytechnic State University, San Luis Obispo, CA 93407
805-756-1392
bdebruhl@calpoly.edu

ABSTRACT

Since the release of the Kodak Camera in 1888, the relationship of technology and privacy has been a matter of public consciousness [5]. Recently, the rise of ubiquitous computing and 'big data' has created new questions about privacy for policy- and technology-makers alike [16, 28]. However, few computer science programs have incorporated privacy education into their curriculum. Looking at ABET accredited programs as a sample, we find that less than 2% of CS programs have a dedicated undergraduate privacy course. In this paper, we introduce a privacy course tailored to undergraduate computing students. We establish units, readings, assignments, and activities for this course. Lastly, we analyze student feedback from a pilot of this course.

INTRODUCTION

Privacy, or "complying with a person's desires when it comes to handling his or her personal data" [8], and its relation to technology has been of increasing public interest. Adoption of ubiquitous computing [16] and large-scale data aggregation [28] has created new privacy challenges for policy- and technology-makers alike. These challenges are often complicated by varied cultural norms, unclear legal precedent, and technological misunderstandings. Computer scientists must be prepared to take on these privacy problems by carefully balancing social, ethical, and technical constraints. However, very few undergraduate CS students have the opportunity to explore privacy in depth.

CS curricula have increasingly seen demand for the inclusion of computer security coursework [9]. This has resulted in high adoption of computer security courses, with 80% of ABET accredited CS programs offering at least one security course. However, only 2% of programs currently offer courses dedicated to technical privacy. We also see a gap in the literature that provides guidance in developing a meaningful course in this domain. Therefore, we introduce a privacy course tailored for undergraduates including suggested readings, active learning activities, homework assignments, and projects.

In our course, we introduce undergraduate students to historical, social, and legal precedent of the relationship between technology and privacy. We then introduce privacy-preserving techniques and technologies. Throughout the course, discussion of privacy-related technologies reflects back upon social principles and policies—spurring students to continually reflect on the intersection of technology and society. In a pilot of this course, we assessed students learning with an entry/exit knowledge survey [22] of relevant skills. We find a notable increase in student confidence in technical privacy skills and analysis.

THE STATE OF TECHNICAL PRIVACY EDUCATION

We survey the availability of privacy engineering courses in computer science programs. Since privacy and security are often inextricably combined [20] we explore

privacy and security curricula together. We assume the curricula of the 283 ABET accredited computer science programs to be representative of computer science curricula in general because the ABET accreditation process ensures baseline standards generally accepted in both academia and industry [2]. We search each ABET accredited program's course catalog for courses in security or privacy. We were able to find 270 catalogs, 80% of which contain one or more courses explicitly in computer security and/or privacy.

Overall, we found a total of 577 security courses offered which we categorize according to the course title, as summarized in Figure 1. When a course covered two or more topics we divided it into the relevant categories. Overall, 25% of courses offered are in the area of computer security, computer system security, or security engineering. Other common topics include network security (17%), cryptography (12%), and information assurance (10%). Most importantly, only 0.5% of courses (3 courses) are dedicated to privacy.

It may be expected that introductory security courses also cover privacy. However, when we examined daily course schedules for 8 undergrad security courses from prestigious institutions we find that these courses spend, at most, one day on privacy.

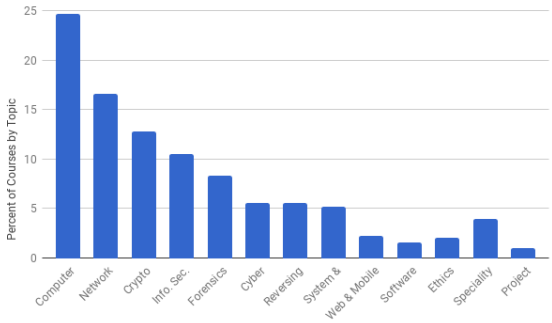

Figure 1. A breakdown of security course offerings in computer science departments by sub-topic at ABET accredited institutions. We group privacy courses with the "specialty" courses.

THE COURSE

We introduce our course content which simultaneously introduces CS students to ethical, technical, and legal precedence in privacy. This course has a prerequisite of a course in "professional responsibilities" to provide students a basis in fundamentals of ethical analysis and legal reasoning. As designed, this course is a ten-week course meeting three times a week for an hour each of lab and lecture. The lecture period includes a mix of lectures, active learning activities, and problem sets. In the lab period, we include two technical projects, one "mini" debate, and one "full" debate.

We divide the course into four units

- Privacy fundamentals: an overview of privacy definitions and history.
- Data privacy: the motivations for data privacy and common implementations.
- Web privacy: Internet privacy policy, usability, technology, and difficulties.
- Mobile privacy: smartphone and IoT privacy.

Although we present these units as a cohesive course, we believe they can be modularized for inclusion in other courses. Due to space limitations, we do not discuss mobile privacy.

Assigned Reading

Technical privacy is a relatively young field and there are limited options for textbooks. Therefore, we choose primary literature sources for students to read from technical journals, legal journals, and conference proceedings.

Privacy Fundamentals - To provide a broad definition and history of privacy, we assign the Stanford Encyclopedia of Philosophy entry on "privacy" [12]. To provide a concrete, non-abstract definition view of privacy and its impact, we assign selections from [7]. Lastly, we assign selections from [24], to demonstrate persuasive arguments for and against privacy. A notable missing work is contextual integrity [21], which we introduce with concrete examples in the mobile privacy section.

Data Privacy - We assign [26], which demonstrates a practical re-identification attack, identifying individuals in an "anonymized" dataset, to motivate data privacy. We assign articles explaining two technical data privacy techniques including k-anonymity [25] and differential privacy [14].

Web Privacy - To introduce students to information privacy and web tracking techniques we assign [15] and [6], which provides a theoretical and practical underpinning for browser-based web tracking. We also propose assign readings for ISP-based web tracking and defenses including [19] and [10]. Since usability is a major difficulty for web browser privacy engineering, we assign multiple readings on the topic [11, 3]. Lastly, there are many important policies related to web tracking. We assign readings on the essential topics of "do not track" [27] and the "right to be forgotten" [4].

Example Active Learning Activities and Homework Assignments

As exemplified below, we incorporate in-class active learning activities to increase student engagement and take-home homework assignments to deepen student understanding.

Active learning: Reading Privacy Policies — Because privacy policies are complicated and opaque [11], very few people read them. In this activity, students bring a privacy policy to class and participate in a "think-pair-share" exercise [17]. First, students read and "think" about their privacy policy by answering open-ended questions. Second, students "pair" with a classmate and share what they learned. The pairs collaboratively identify strengths and weaknesses from both policies. Lastly, students "share" what they learned with a 30-second presentation on their partner's policy.

Active learning: Understanding Web Tracking — Many people, including computer scientists, do not understand the intricacies of web tracking. To allow students to explore web tracking we have them visit Panopticlick [6], a website that uses demonstrates various web-tracking tests. While some tests are self-explanatory, others including super-cookies, canvas fingerprinting, and WebGL fingerprinting are less straightforward. We assign each student an advanced web tracking technique and ask them to explore how it is implemented. Afterwards, students must determine potential privacy preserving techniques for their assigned technique.

Homework: Applying K-Anonymity and L-Diversity — This assignment allows students to practice anonymizing a dataset before it is published or distributed. We give students an example dataset containing fake personal information. We have students classify pieces of data as unique identifiers or quasi-identifiers and then anonymize the data using k-anonymity [25]. Given a target anonymity level, students attempt different amounts of

generalization and suppression to reach the goal. They then assessed the level of l-diversity [18] in their anonymized dataset.

Homework: Applying Differential Privacy — Students write functions to return averages from a dataset, then design an attack to obtain information about individuals. They then add noise to the data using a simplified version of differential privacy [14]. Students then update their attack to account for this noise. Finally, they redesign their aggregation functions to add noise to the aggregate values, and consider how this would affect the attacks they have developed.

Example project - web tracking and differential privacy

In this project, we introduce students to web tracking and privacy control technologies. This project is specifically designed to highlight the amount of user information available to modern websites and demonstrate ways to protect that data.

First, students create simple websites with interactive features. They implement server-side tracking techniques to collect data about any visitors to their websites and how those visitors interacted with their sites. This requires students to discover ways to uniquely identify users between sessions. We then simulate user activity with automated programs. These programs visit the sites many times each to allow for tracking through various identifiers. The students store the tracking data and analyze it to find unique users and draw conclusions about patterns in their behavior.

With the collected data, students apply k-anonymity and differential privacy, server-side anonymization techniques. Students compare the anonymized dataset to the original dataset to make observations about the trade-off between privacy and data usefulness. Finally, students submit a write-up that explained the decisions they made and trade-offs they considered throughout the project.

Debates and Term Paper

Debating a topic is a critical skill when discussing privacy, which intersects policy and technology. We require students to debate a position on topics related to ethics, privacy policy, technology, and surveillance using a modified version of the Oxford debate format [1]. We purposely choose a wide range of topics including genetic privacy, information brokerage, web tracking, location tracking, workplace monitoring, and IoT recording.

Since many computing students have not been exposed to formal debate, we first assign a half-length mini-debate with low-stakes grading. This allows students to practice the format and to prepare for the full debate. During the full debate, we require that students come prepared to argue their side thoroughly, prepare rebuttals, and ultimately try to win the audience's approval. This requires students to research arguments for and against their assigned topics, collaborate to develop an effective case, and practice communicating their case.

To assess students' final progress in the course we assign a 3000-word term paper requiring technical and ethical analysis. This assignment has a side benefit of giving students an additional opportunity to write effectively and persuasively, a skill that CS graduates often lack [13]. Students are given considerable freedom to choose a topic of personal interest. The main constraints on the student is that their paper must answer a privacy question with technical and ethical depth.

In the pilot course, we assessed the term paper in three rounds. First, students participated in a peer-review exercise. Second, students were given feedback by the undergraduate teaching assistants using the final paper rubric. Lastly, the paper was graded by the instructor in lieu of a final exam.

ASSESSING THE COURSE

Our assessment of the course is based on student feedback from a paired entry/exit knowledge survey [22] of student confidence in privacy. All of the students in the class were upper division undergraduate students in a computing major and many had a strong interest in security and privacy. Through our survey we found students highly valued privacy ethics, morality, business responsibility, and skills.

For the knowledge survey students were asked to rank statements targeted at measuring their perceived ability to engineer user privacy. The result showed a positive trend in student confidence of their skills to develop privacy-preserving systems, demonstrating the efficacy of the course. The following five questions had statistically significant increases at the $p < 0.05$ level, determined via a Wilcoxon signed-rank test for paired data [23]:

- "When presented with an opinion on a privacy related topic, I would be able to provide reasoned support or opposition."
- "Given a piece of technology (smart-device, website, application, etc), I can reason about the tradeoffs between ethics and business."
- "I can identify and use privacy enhancing technologies for my personal computing."
- "I can identify and use privacy enhancing technologies for integration into development."
- "I can identify techniques to maintain consumer privacy in a database."

Other similar questions also saw significant increases at only slightly larger p-values.

CONCLUSION

While computer security education has received increasing attention, we show that there is a lack of technical privacy education. 80% of CS programs have a course dedicated to computer security while less than 2% have a course dedicated to privacy. Therefore, we introduced an undergraduate course in technical privacy engineering — the first discussed in the literature. This course includes suggested readings, activities, projects, homework, debates, and papers. Finally, we show students had positive outcomes in our pilot course. It is our hope that programs can take our design and implement similar technical privacy courses at other institutions.

REFERENCES

[1] Oxford style debate, 2017, www.uscourts.gov/about-federal-courts/educational-resources/about-educational-outreach/activity-resources/oxford, retrieved January 6 2017.

[2] A valued credential, 2018, www.abet.org/accreditation/why-abet-accreditation-matters/a-valued-credential/, retrieved January 6, 2018.

[3] Acquisti, A., and Grossklags, J., Privacy and rationality in individual decision making, *IEEE Security & Privacy* 3, (1), 26-33, 2005.
[4] Bennett, S. C. The right to be forgotten: Reconciling EU and US perspectives, *Berkeley J. Int'l L. 30*, 161, 2012.
[5] Bratman, B., Brandeis and warren's the right to privacy and the birth of the right to privacy, *Tenn. L. Rev.* 69, 623, 2001.
[6] Budington, B., Panopticlick: Fingerprinting your web presence, *Usenix: Enigma 2016,* 2016.
[7] Calo, R., The boundaries of privacy harm, *Indiana L.J., 86,* (3), 2011.
[8] Cannon, J., *Privacy: what developers and IT professionals should know*, Addison-Wesley Professional, 2004.
[9] Cassel, L., Clements, A., Davies, G., Guzdial, M., McCauley, R., McGettrick, A., Sloan, B., Snyder, L., Tymann, P., and Weide, B. W., Computer science curriculum 2008: An interim revision of cs, 2001.
[10] Chaum, D. L., Untraceable electronic mail, return addresses, and digital pseudonyms, *Communications of the ACM, 24,* (2), 84–90, 1981.
[11] Cranor, L., P3p: Making privacy policies more useful, *IEEE Security & Privacy*, *99*, (6) , 50-55, 2003.
[12] DeCew, J., Privacy, *Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University, 2015.
[13] Dugan Jr, R. F., and Polanski, V. G., Writing for computer science: a taxonomy of writing tasks and general advice, *Journal of Computing Sciences in Colleges*, 21, (6), 2006.
[14] Dwork, C., A firm foundation for private data analysis, *Communications of the ACM*, *54*, (1), 86-95, 2011.
[15] Eckersley, P., A primer on information theory and privacy, *EFF*, 2010.
[16] Harris, K. D., Privacy on the go, recommendations for the mobile ecosystem, Attorney General of California Department of Justice: Privacy Enforcement and Protection Unit, 2013.
[17] Kothiyal, A., Majumdar, R., Murthy, S., and Iyer, S., Effect of think-pair-share in a large cs1 class: 83% sustained engagement, *Proc. of the 9th annual international ACM conference on Int'l computing education research, 137-144,* 2013.
[18] Machanavajjhala, A., Gehrke, J., Kifer, D., and Venkitasubramaniam, M., l-diversity: Privacy beyond k-anonymity, *Proc. of the 22nd IEEE International Conference on Data Engineering*, 24, 2005.
[19] McLachlan, J., Tran, A., Hopper, N., and Kim, Y., Scalable onion routing with torsk, *Proc of the 16th ACM conference on Computer and communications security*, 590-599, 2009.
[20] Newhouse, W. D., Keith, S., Scribner, B., and Witte, G., Nice cybersecurity workforce framework: National initiative for cybersecurity education,Tech. rep., 2017.
[21] Nissenbaum, H., A contextual approach to privacy online, Daedalus, *140*, (4), 32-48, 2011.
[22] Nuhfer, E., and Knipp, D., The knowledge survey: A tool for all reasons, *To improve the academy, 21*, (1), 59-78, 2003.
[23] Pratt, J. W., Remarks on zeros and ties in the Wilcoxon signed rank procedures, *Journal of the American Statistical Association, 54*, (287), 655-667, 1959.
[24] Solove, D. J., I've got nothing to hide and other misunderstandings of privacy, *San Diego L. Rev*., *44*, 745, 2007.

[25] Sweeney, L., k-anonymity: A model for protecting privacy, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*, (05), 557-570, 2002.
[26] Sweeney, L. Matching known patients to health records in washington state data, Harvard University, Data Privacy Lab, 2013.
[27] Tene, O., and Polenetsky, J., To track or do not track: Advancing transparency and individual control in online behavioral advertising. *Minnesota Journal of Law, Science and Technology, 13*, 281, 2012.
[28] Tene, O., and Polonetsky, J., Big data for all: Privacy and user control in the age of analytics, *Nw. J. Tech. & Intell. Prop*., 11, 2012.