Keynote: Privacy in Today's World

Rebecca N. Wright
DIMACS and Computer Science
Rutgers University
rebecca.wright@rutgers.edu

ABSTRACT

Privacy has become an issue of increasing importance in today's world of networked computing, mobile devices, embedded computing, and large-scale data analysis. This work describes several privacy solutions and challenges.

1 INTRODUCTION

As computing technology has advanced, personal data is increasingly used and collected at a previously unimaginable scale. These technologies have many positive benefits, but also result in potential privacy concerns and privacy breaches. There is a large body of research in the computer science community and elsewhere seeking to understand privacy and the nature of the privacy problem, put it on firm theoretical grounding, and provide customizable and usable privacy solutions. This keynote talk will highlight a few of these contributions.

2 DP-WHERE: DIFFERENTIALLY PRIVATE MODELING OF HUMAN MOBILITY

The work described in this section is collaborative work with Darakhshan Mir, Sibren Isaacman, Ramón Cáceres, and Margaret Martonosi [5].

Models of human mobility have wide applicability to infrastructure and resource planning, analysis of infectious disease dynamics, ecology, and more. The abundance of spatiotemporal data from cellular telephone networks affords new opportunities to construct such models. Furthermore, such data can be gathered with greater detail at larger scale and lower cost than traditional methods, such as census surveys.

While human mobility models have the potential for great societal benefits, privacy concerns regarding their use of individuals' location data have inhibited their release and wider use. Our work on DP-WHERE enables the creation of human mobility models in a privacy sensitive manner. Specifically, prior work introduced and validated the WHERE (Work and Home Extracted REgions) approach to mobility modeling [3]. WHERE aggregates and distills cellphone Call Detail Records (CDRs) to form a mobility model that can be used to characterize a city's commute patterns and enable the exploration of what-if scenarios regarding changes in residential density, telecommuting popularity, etc.

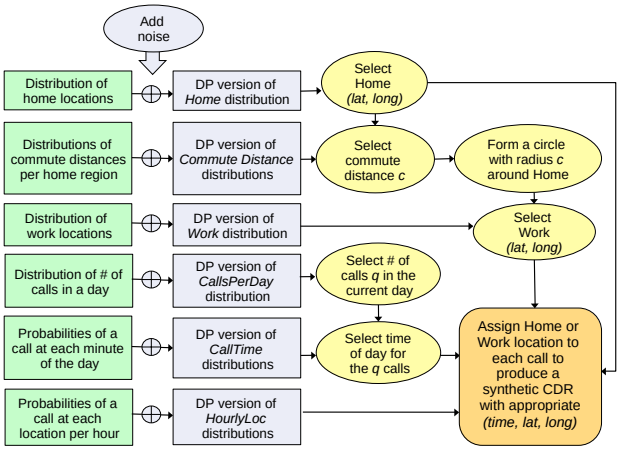

Figure 1: Overview of DP-WHERE, which modifies WHERE by adding noise to achieve differentially private versions of the input probability distributions. The rest of WHERE remains unchanged.

Our work introduced DP-WHERE, a *differentially private* version of WHERE. DP-WHERE satisfies the rigorous requirements of differential privacy while retaining WHERE's usefulness for predicting movement of human populations in metropolitan areas. Differential privacy [2] makes privacy a mathematical requirement on the results of interactions with data. In particular, differential privacy captures the intuition that, in order to provide privacy to individuals, the results of an interaction with a database should be almost the same whether or not any particular individual is present in a database. This is a strong notion of privacy that makes no assumptions about the power or background knowledge of a potential adversary, and that is beginning to see practical adoption. Our work demonstrates the value of a multi-pronged approach to privacy: Our model starts with attributes (such as sampling and aggregation) that make it intrinsically well suited to offering some intuitive degree of privacy. We subsequently modify the steps of the modeling

CyberW'17, October 30, 2017, Dallas, TX, USA

ACM ISBN 978-1-4503-5393-9/17/10...$15.00
https://doi.org/10.1145/3139531.3139536

algorithm to rigorously implement differential privacy, with the results obtaining a reasonable privacy/utility tradeoff because of the earlier sampling and aggregation steps.

DP-WHERE achieves differential privacy by adding controlled noise to the set of empirical probability distributions that WHERE uses, for example distributions of home and work locations. DP-WHERE then proceeds identically to WHERE by systematically sampling these distributions to generate synthetic CDRs containing synthetic locations and associated times, as shown in Figure 1. Because none of these sampling steps require further access to the original CDRs, it would be possible for the data holder to release the noisy distributions while retaining differential privacy. Among possible uses, these distributions would allow others to produce their own synthetic CDR traces for any desired population size, time duration, or other parameters.

Our experiments show that differential privacy can be achieved with only a modest reduction in accuracy. In particular, across a wide array of experiments involving 10,000 synthetic users moving across more than 14,000 square miles, the distance between synthetic and real population density distributions for DP-WHERE differed by only 0.17–2.2 miles from those of WHERE. An example comparison of population densities for WHERE and DP-WHERE is shown in Figure 2.

Overall, our work shows that modest revisions to a mobility model drawn from real-world and large-scale location data allow for rigorous demonstrations of its privacy without overly compromising its utility. We believe that this demonstrates reason for optimism regarding the judicious use of large data repositories of potentially sensitive information.

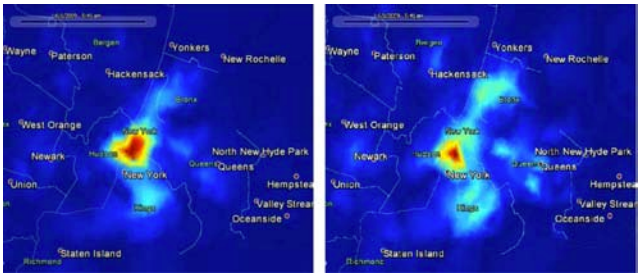

Figure 2: WHERE and DP-WHERE results.

3 SIDE CHANNELS IN FACEBOOK

The work described in this section is collaborative work with Sai Lu and Janne Lindqvist [4].

Over the course of the last decade, Facebook has become an incredibly popular social networking site, reporting around a billion visitors monthly. Like any social networking site, Facebook's design decisions have implications about what is shared, what is not shared, and how much control users have about such sharing. We carried out a systematic analysis of side channels in Facebook—that is, channels that can reveal privacy-sensitive information to users through indirect

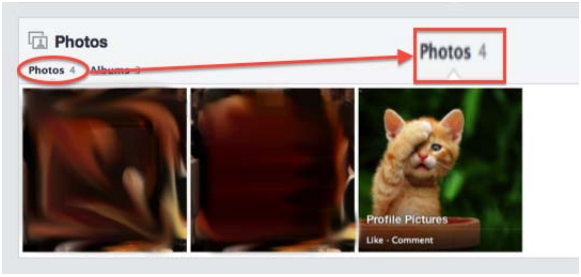

Figure 3: An example of a side channel in Facebook. There are only three photos displayed. In contrast, Facebook lists that there are four photos in this photo album. Users viewing this can infer that one photo has been blocked from them by the user who posted these photos.

mechanisms. While these side channels may not be particularly surprising to the security research community, they still represent potential threats to Facebook users, depending on user expectations and attitudes. We surveyed Facebook users to determine user expectations and attitudes, including whether users are aware of these channels, and whether there are privacy objections to the channels to those aware. We found that many users are unaware of the side channels and express surprise at finding out about the side channels. Among users who were aware of them, some users express concerns while others did not. Based on these results, we also identified design implications for social network sites that wish to provide users with more control over such choices. In our work, we focus on information that is shared via "side channels." Analogous to side channels in other computer systems and applications, such as storage channels, timing channels, and power consumption channels, a side channel is an information channel that is secondary or incidental to the intended communication channel but that can convey additional information. An example of a side channel in Facebook is associated with the number of images in a Facebook photo album, as depicted in Figure 3. Users are allowed to share their photos on Facebook, but also are able to limit who can view them. However, the true count of photos is still shown. In the case shown in the figure, Facebook lists a total of four photos for this photo album but shows only three different photos. Rhe viewer could infer from this that he or she is blocked from viewing one photo.

In our study [4], we systematically investigated the existence of side channels in Facebook and identified ten such side channels of four major types: associated with counts (three side channels), sharing (two side channels), tagging (four side channels), and relationship status (one side channel). Three Facebook functions (Likes, Photos, Notes) show a total number of items to users even though the users do not have access to all of them. Four combinations of privacy settings made by posts' owners and users who get tagged in them will lead a mutual friend to discover that he or she is blocked from seeing a post. Sharing a non-public post in a message

and in a group causes side channels, too. A restricted-access user is able to find out that he or she is blocked by the poster based on the different contents that show on two timelines.

In order to determine whether these side channels actually represent a privacy problem, it is important to understand what people think about the side channels. To this end, we conducted a survey of Facebook users to determine whether users are aware of these channels or not, and whether those who are aware have privacy objections to the channels. Using Amazon Mechanical Turk, we surveyed a total of 80 participants. Among these participants, we found that many were unaware of the side channels and expressed surprise at finding out about them. Among participants who were aware of them, some users expressed concern while others did not.

Some examples that participants shared with use regarding their specific experiences with side channels follow. One participant wrote: *"I have posted on a mutual friends' wall, and I assume the person I unfriended, clicked to add me when she saw that I had posted on our friends' wall. I did not specifically exclude her from seeing my posts though."* Another noted: *"I have tried to make private events, but they show up on my timeline. People not invited can't see the actual event, but they can see that I posted an event."*

4 JANA: PRACTICAL PRIVATE DATA AS A SERVICE

The Jana project is ongoing collaborative work with David Archer and others at Galois, Inc., Anand Sarwate, David Cash, Nigel Smart and others at the University of Bristol, and Dov Gordon. In this project, we seek to provide flexible and efficient solutions for private data as a service. In Jana, contributed data is encrypted at all times, starting before it leaves the subjects possession. Results of queries against data are limited to how much data subjects are willing to reveal, to whom, and when. This is achieved through a combination of tools including secure multiparty computation (based on SPDZ [1]) and differential privacy [2] in order to allow computations to be carried out on large-scale data from multiple data holders in a way that satisfies policies specified by the data owners and other relevant stakeholders.

Challenges to be addressed in such an implementation include:

- Efficient and secure methods to ingest data from multiple parties so that data is securely encrypted while also enabling the necessary computations on the data.
- Combining multiple privacy-preserving techniques in practical, interoperable ways while preventing unintended privacy gaps.
- Understanding what families of policies are appropriate and enforceable for real-world use cases, and determining how best to interact with system designers and users to set specific policy choices within those families.
- Scaling up to practical data volumes with practical throughput and latency for real-world use cases.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under grant numbers CCF-1018445 and CNS-1018557, and the Defense Advanced Research Projects Agency (DARPA) and Space and Naval Warfare Systems Center, Pacific (SSC Pacific) under contract No. N66001-15-C-4070.

REFERENCES

[1] Ivan Damgard, Marcel Keller, Enrique Larraia, Valerio Pastro, Peter Scholl, and Nigel P Smart. 2013. Practical Covertly Secure MPC for Dishonest Majority–or: Breaking the SPDZ Limits. In *ESORICS 2013*.

[2] Cynthia Dwork, Frank Mcsherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In *TCC '06: In Proceedings of the 3rd Theory of Cryptography Conference*. 265–284.

[3] Sibren Isaacman, Richard A. Becker, Ramón Cáceres, Margaret Martonosi, James Rowland, Alexander Varshavsky, and Walter Willinger. Human mobility modeling at metropolitan scales. In *MobiSys '12*.

[4] Sai Lu, Janne Lindqvist, and Rebecca N. Wright. 2014. Uncovering Facebook Side Channels and User Attitudes. In *Online Proceedings of Web 2.0 Security and Privacy, held in conjunction with the IEEE Symposium on Security and Privacy*. IEEE, San Jose, CA. http://www.w2spconf.com/2014/papers/facebook_side_channels.pdf

[5] Darakhshan J. Mir, Sibren Isaacman, Ramón Cáceres, Margaret Martonosi, and Rebecca N. Wright. 2013. DP-WHERE: Differentially private modeling of human mobility. In *Proceedings of the 2013 IEEE International Conference on Big Data*. IEEE, Santa Clara, CA, 580–588. https://doi.org/10.1109/BigData.2013.6691626

SHORT BIOGRAPHY

Dr. Rebecca Wright is a professor in the Computer Science Department and Director of DIMACS at Rutgers. Prior to joining Rutgers, she was a professor in the Computer Science Department at Stevens Institute of Technology in Hoboken, New Jersey and a researcher in the Secure Systems Research Department at AT&T Labs and AT&T Bell Labs. Her research spans the area of information security, including cryptography, privacy, foundations of computer security, and fault-tolerant distributed computing. Recent work includes privacy-preserving data mining, secure multiparty approximations, and improved bounds for Byzantine agreement in the shared memory model. Her ongoing research goals are the design of protocols, systems, and services that perform their specified computational or communication functions even if some of the participants or underlying components behave maliciously, and that balance individual needs such as privacy with collective needs such as network survivability and public safety.

Dr. Wright serves as an editor of the International Journal of Information and Computer Security and the Transactions on Data Privacy. She is a member of the board of the Computer Research Association's Committee on the Status of Women in Computing Research (CRA-W). She was a member of the board of directors of the International Association for Cryptologic Research from 2001 to 2005 and the steering committee of the Information Security Conference from 2006 to 2010, and an editor of the Journal of Computer Security from 2001 to 2011. She was Program Chair of Financial Cryptography 2003 and the 2006 ACM Conference on Computer and Communications Security (CCS) and General Chair of Crypto 2002. She has served on numerous program committees, including Crypto, the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, and the Usenix Security Symposium. She received a Ph.D. in Computer Science from Yale University in 1994 and a B.A. from Columbia University in 1988. She received an honorary M.E. from Stevens Institute of Technology in 2006. She is a member of the IEEE and the ACM.