E6998-004 Privacy and Online Social Networks

Balachander Krishnamurthy, AT&T Labs–Research, bala@cs.columbia.edu

Course URL: https://www.cs.columbia.edu/~bala/s14
TA: Chris Riederer, cjr2149@columbia.edu, Office: 618 CEPSR

Purpose of course

This is a graduate-level course focusing on privacy issues and online social networks (OSN). Privacy has become a hot topic due to many reasons: demographics on the Internet, popularity of OSNs, and considerable media interest. We will learn about origins of the privacy problem, the various entities involved, technology used to track users and to prevent tracking, and the specific manners of privacy leakages.

The course is *not* about the role of government, legal issues, or policy matters.

The course will let you become familiar with technical publications in this area and gain a first-hand understanding of privacy-related technologies. The focus of the work will be on OSNs. You will write code and present results of your creative work. You are free to use any programming language and Operating System but please note that I will not be able to help you with *any* aspect of coding.

Pre-requisites

The course is open to graduate students but highly interested undergraduates are indeed welcome to convince me and i'd be happy to have you register. You must be comfortable reading technical papers and writing code.

What you will learn

- Full awareness of the privacy problem and how it manifests itself on OSNs
- How to track trackers
- Privacy protection measures
- Possible privacy protection improvement vectors

What you will do

- Participate actively in class
- Read several papers
- Research, design, write code (in groups), present. Your code will be tested by other students.
- Write a potentially submission-quality paper based on the project.

Reading list

- No required books but papers from:

  Conferences (COSN, USENIX Sec. Symp., IEEE Symp on S&P, WWW, IMC, CCS)
  Workshops (WOSN, PETS, W2SP, WEIS, WPES, SOUPS)
  Journals (IEEE S&P etc.).

- See bib (evolving) at http://www.cs.columbia.edu/~bala/html/s14/bib.pdf

Office hours: *NONE*. Email me or meet before/after class, schedule-permitting.

Grading and submissions

Please note: there will be no extensions to homeworks or the various project deadlines under any circumstances. Do not ask for extensions. If you have a serious medical or some genuinely major personal issue, I will refer you to the Dean's office and abide by their suggestion.

Make backups! Make backups of backups!

Grade will be based *roughly* along these lines:

- 10% class attendance and active participation
- 20% homework (including paper reviews, an annotated bibliography with critical evaluation)
- 45% on group project
- 25% on a 12 page potentially submission-quality paper

Paper reviews

A key thing you will learn in this course is critical reading of technical papers. While it is increasingly hard to get papers accepted at good conferences, it is the case that not all published papers are great. When you read papers and write reviews, I want to ensure that you understood the contributions of the paper. However, more importantly, you should be able to *critically* evaluate it by indicating which parts of the paper, if any, are weak and why, what improvements might have been possible, etc. Additionally, this will help you with the final paper that you need to write for the course.

Academic honesty

Read http://www.cs.columbia.edu/education/honesty NOW.

Except for the group project (where you will work with members of just your group), *all* other work has to be just yours. It is important and I take it seriously. If I detect violation, I will ensure that you are referred to the Academic Committee. It *will* affect your grade.

Discussion is ok but copying is *lame*. If you improperly help someone else then I consider you to have cheated as well. I look for leakages in my research; so chances of you escaping scrutiny are low. If it is *not* your idea then cite the source and say how you build on it or differ from it.

Group project timeline

(Note: Each phase will contribute to your overall grade)

You are free to be a group of size one. Actual group sizes will depend on enrollment count but unlikely to be more than 3 or 4. Part of what you learn is to divide the work equally and help each other *within* the group to continue to make progress.

1. Proposal—due by 2/26: Submit proposal of your choice—it can be short or detailed.

Or: my suggestion—build a personal privacy assistant (PPA)
Your project should demonstrate the following (evaluation is based on this)

- originality
- good selection of domain where privacy will be protected (FB/Twitter etc. is ok. Snapchat, ask.fm, kik etc. may be more intriguing).
- Software: iOS, Android, or whatever - I don't care, as long as *you* are comfortable coding. Do *not* ask me for coding help.
- understanding of privacy loss
- demonstration of protection
- awareness of all costs for such protection including loss of fidelity, latency, etc.
- new technologies for improving protection
- any crowdsourcing to spread or improve would be a plus

If you decide to work on PPA:

- points for originality here depends on depth
- an app that watches over your actions tracks your data dribbling at low cost
- estimate cost of privacy protection
- what is the best protection/cost ratio that can be achieved
- what alternate methods can be developed for protection
- only your privacy matters (at best) but can you get others to buy into it?

2. Design document and progress report–due 3/26: Status of the project, references, code component description all compiled into a 4-5 pages report.

3. Presentation of project—on 4/16 and 4/23: 30 min presentation/demo per group. Other students will test your code. Presentation will be evaluated and be a component of your grade. Yes, some will have one less week to finish the code and present it but will really have an extra week to work on the paper (order chosen by lottery, if needed).

4. Submission-quality paper summarizing the project–due 4/30: Group evaluation: I'll allocate both a group grade and ask members to rate other group members confidentially. I'll then decide if outliers need to be evaluated differently.

Lecture outline (*not* set in concrete!)

1/22 Lecture 1: Introduction to privacy and OSNs

- History and definition of Privacy
- Vocabulary, history, terms
- Types of consumers, privacy
- Myths or why we don't care?
- Brief introduction to OSNs

Papers: [19, 6, 18, 21, 10, 4, 14]
Homework assignment 1 – due next week 1/29!

1/29 Lecture 2: Privacy, security, anonymity *Homework assignment 1 – due TODAY!*

- Duality of privacy and security
- Privacy and security
- The role of usability
- Crypto and its failure
- Privacy and anonymity
- Anonymization techniques
- Why hasn't privacy/security duality yielded more?

Papers: [20, 16, 1, 7, 13, 23, 1, 2, 24, 15]

2/5 Lecture 3: Technology – 1

- Terminology and key players
- Tracking
- Technologies for tracking
- Technical vectors of leakage and ways of identifying them
- Role of JavaScript
- Role of protocols

Papers: [8, 3, 12, 11]
Homework assignment 2 – due next week 2/12!

2/12 Lecture 4: Technology – 2 *Homework assignment 2 – due TODAY!*

- PII: What is personally identifiable information
- People search engines
- Online Social Networks: a more detailed look

Papers: [5, 11]

2/19 Lecture 5: Technology – 3

- OSNs (continued)
- Mobile OSNs
- Non-US centric OSNs: .kr, .cn, .jp
- Special purpose OSNs: Pinterest, SnapChat, Ask.fm, WhatsApp
- Privacy settings in OSNs
- PII leakage in OSNs

Papers: [9, 22]
Homework assignment 3 – due 3/5!

2/26 Lecture 6: Technology – 4 *Reminder: Project proposal due TODAY!*

- Linkage
- Semantics and the compositional problem
- Collateral Damage of Privacy

Reminder: Homework assignment 3 – due next week 3/5!

3/5 Lecture 7: Protection – 1 *Reminder: Homework assignment 3 – due TODAY!*

- Early methods
- Cat and Mouse game
- Clever techniques
- Role of usability

3/12 Lecture 8: Protection – 2

- Complex technologies
- Startup solutions
- Thinning/source quenching
- Role of crowdsourcing
- Differential Privacy

Papers: [7]

3/26 Lecture 9: Deployment *Reminder: Design document and progress report due TODAY!*

- Practical issues
- Limitations
- Countermeasures
- Incentives

4/2 Lecture 10: Inter-disciplinary role of privacy

- Possibly: 2 Guest lectures (TBA)

4/9 Lecture 11: Recent topics

- Economics of Privacy
- Privacy of mixed data sets, OSNs
- Privacy across time

Papers: [17]

4/16 Class 12: Short paper presentations by students

4/23 Class 13: Project presentation

4/30 Class 14: Project presentation

- *Final paper due*
- Recall: *No* extensions. Don't ask. Don't yell.

References

[1] A. Adams and M. A. Sasse. Users are not the enemy: why users compromise security mechanisms and how to take remedial measures. *Communications of the ACM*, 42(12):40–46, 1999.

[2] D. Chaum. Security without identification: Transaction systems to make big brother obsolete. *Communications of the ACM*, 28(10):1030–1044, 1985.

[3] Joshua Gomez, Travis Pinnick, and Ashkan Soltani. Knowprivacy: The current state of web privacy, data collection and information sharing, June 2009. `http://knowprivacy.org/report/KnowPrivacy_Final_Report.pdf`.

[4] Graham Cormode and Balachander Krishnamurthy. Key differences between Web 1.0 and Web 2.0. *First Monday*, 13(6), June 2008. `http://www.research.att.com/~bala/papers/web1v2.pdf`.

[5] B. Krishnamurthy and C. Wills. On the leakage of personally identifiable information via online social networks. In *Proceedings of the Workshop on Online Social Networks*, August 2009. `http://www.research.att.com/~bala/papers/wosn09.pdf`.

[6] Balachander Krishnamurthy. I know what you will do next summer. *ACM SIGCOMM CCR*, 40(5), 2010. `http://www.research.att.com/~bala/papers/ccr10-priv.pdf`.

[7] Balachander Krishnamurthy. Privacy and online social networks: Can colorless green ideas sleep furiously? *IEEE Security and Privacy*, May-June 2013.

[8] Balachander Krishnamurthy and Craig E. Wills. Privacy diffusion on the web: A longitudinal perspective. In *WWW*, 2009. `http://www.research.att.com/~bala/papers/www09.pdf`.

[9] Balachander Krishnamurthy and Craig E. Wills. Privacy leakage in mobile online social networks. In *Proceedings of the Workshop on Online Social Networks*, June 2010. `http://www.research.att.com/~bala/papers/pmob.pdf`.

[10] Mary Madden, Susannah Fox, Aaron Smith, and Jessica Vitak
. Digital footprints. http://www.pewinternet.org/Reports/2007/Digital-Footprints/1-Sum%mary-of-Findings.aspx.

[11] Marco Balduzzi, et al. Abusing social networks for automated user profiling. In *RAID*, 2010. http://www.iseclab.org/papers/raid2010.pdf.

[12] Jonathan R. Mayer and John C. Mitchell. Third-party web tracking: Policy and technology. In *Proceedings of IEEE Symposium on Security and Privacy*, May 2012. https://stanford.edu/~jmayer/papers/trackingsurvey12.pdf.

[13] Arvind Narayanan. What happened to the crypto dream? parts 1 and 2. *IEEE Security and Privacy*, March-April, May-June 2013.

[14] Helen Nissenbaum. A contextual approach to privacy online. *Daedalus the Journal of the American Academy of Arts & Sciences*, 140(4):32–48, Fall 2011. http://www.amacad.org/publications/daedalus/11_fall_nissenbaum.p%df.

[15] Lee Rainie, Sara Kiesler, Ruogu Kang, and Mary Madden. Anonymity, privacy, and security online. http://pewinternet.org/Reports/2013/Anonymity-online.aspx.

[16] M. K. Reiter and A. D Rubin. Anonymous web transactions with crowds. *Communications of the ACM*, 42(2):32–48, 1999.

[17] Christopher Riederer, Vijay Erramilli, Augustin Chaintreau, Balachander Krishnamurthy, and Pablo Rodriguez. For sale: Your data, by: You. In *Proceedings of the Workshop on Hot Topics in Networking*, Cambridge, MA USA, November 2011. http://www.research.att.com/~bala/papers/hotnets11.pdf.

[18] Robert L. Rothman. A guide to privacy law, 2010. http://www.privassoc.com/Documents/ICLE%203rd%20Annual%20Informa%tion%20Technology%20Law%20Seminar%209-22-2010.pdf.

[19] Daniel J. Solove. "'i've got nothing to hide' and other misunderstandings of privacy". *San Diego Law Review, Vol. 44, p. 745*, 2007. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=998565.

[20] Latanya Sweeney. k-anonymity: a model for protecting privacy. *International Journal of Uncertain. Fuzziness and Knowledge-Based Systems*, 10(5):557–570, 2002.

[21] James Waldo, Herbert S. Lin, and Lynette I. Millett. Thinking about privacy. In *Engaging privacy and information technology in a digital age*. National Academic Press, 2007. http://www.nap.edu/openbook.php?record_id=11896.

[22] David Wetherall, David Choffnes, B. Greenstein, Seungyeop Han, Peter Hornyack, Jaeyeon Jung, Stuart Schechter, and Xiao Wang. Privacy revelations for web and mobile apps. In *Proceedings of HotOS*, May 2011. http://appanalysis.org/jjung/jaeyeon-pub/hotos2011-revelations.p%df.

[23] Alma Whitten and J.D. Tygar. Why johnny can't encrypt: A usability case study of pgp 5.0. In *8th USENIX Security Symposium*, August 1999.

[24] Sergey Yekhanin. Private information retrieval. *Communications of the ACM*, 53(4):68–73, 2010.