CSC 495.002 – Group Projects

Dr. Özgür Kafalı
North Carolina State University
Department of Computer Science

The main goal of the projects is to give you experience by working on a specific topic related to privacy. Throughout the course, you will be working on a project in groups of 2–3. A project can be chosen by multiple groups. In that case, we will customize the project scope and deliverables to minimize overlap between the groups.

Each project will have the following deliverables expected from the project groups:

1. One page project proposal describing the project goals, research questions, and anticipated contributions of each group member
2. Intermediate report describing current progress towards project goals
3. Final report containing the following sections:
    - Introduction: State your goal and research questions with regards to the project topic. Describe why you chose those research questions, and describe (if applicable) how they deviate from the general project topic.
    - Background and motivation: One page summary of the literature on the subject (challenges, limitations, application areas).
    - Methodology: Explain your approach for achieving your project goal including any manual methodology used, algorithms developed, tools used off the shelf or developed within the course of the project. Describe what the contributions of each group member are.
    - Results: What have you achieved in the project? Explain your findings with the support of figures, tables where applicable.
    - Future Work: Describe open issues and how you would extend the work done in the project.
4. Project specific deliverables (stated below for each project)
5. In class presentations

Important Dates:

- September 11th: Formation of project groups and project proposals due
- October 23rd: Progress reports due
- November 20th: Final reports and deliverables due
- November 20th: In class presentations start

Project 1: Development of a Privacy Ontology

In this project, you will investigate privacy incidents from the Privacy Incidents Database[1], and develop an ontology of privacy breaches. The developed ontology will contain concepts unified from the individual incidents contained in the database as well as their relations with other and associated properties. You will also develop a similarity metric to compare incidents from the database and beyond. Part of the project will consist of aggregating the results from different project groups. Some useful references for this project are [Gharib et al., 2016], [Guo et al., 2014], [Kafalı et al., 2017], [Murukannaiah et al., 2017].

Pros/cons

- Instructor available for guidance (we will also have a lecture on ontologies and semantic similarity)
- Opportunity to exchange ideas with other groups
- Highly publishable work if you do a thorough job
- Requires teamwork and collaboration among groups

Specific project deliverables

1. An ontology developed with the Protégé[2] ontology development tool
2. An implemented similarity metric that takes as input two privacy breaches and queries the ontology to compute the similarity between the incidents

[1] `https://sites.google.com/site/privacyincidentsdatabase/`
[2] `http://protege.stanford.edu/`

Project 2: Classification of Healthcare Privacy Breaches

In this project, you will investigate breaches from the US Department of Health and Human Services breach report[3], and distinguish between security and privacy incidents. We are mainly interested in classification of privacy incidents that are caused by human errors. You will identify common patterns found in breach descriptions (data collection, data usage, data sharing), and report their frequency of occurrence. Part of the project will consist of aggregating the results from different project groups as well as findings from Project 1. A useful reference for this project is [Kafalı et al., 2017].

Pros/cons

- Instructor available for guidance (we will also have a lecture on breaches)
- Opportunity to exchange ideas with other groups
- Highly publishable work if you perform a thorough analysis
- Requires teamwork and collaboration among groups

Specific project deliverables

1. A categorization of privacy related HHS incidents (beyond the categories provided by HHS) with respect to the tags contained in the Privacy Incidents Database
2. Development of a set of common patterns among incidents
3. A list of potential breaches from the HHS datasets as additions to the Privacy Incidents Database

Project 3: Development of a Privacy Card Game

In this project, you will develop a privacy card game for the goal of understanding how people make choices to mitigate privacy risks. First, you will perform a survey of existing privacy games in the literature. By doing so, you will identify the design space of such games, e.g., what are their objectives, what sort of user interfaces and other features they support. Based on your findings, you will design and implement features for the NormDefense game[4], which we have recently started developing. The broad objectives of the game is to (i) prioritize privacy risks and associated mitigation techniques, (ii) act as a testbed for researchers to develop and test privacy related hypotheses, and (iii) serve as a tool for privacy education and training. The new features you will implement will contribute to the above goals, and explore privacy norms and tradeoffs. You will also design user studies using realistic privacy scenarios (with customized card decks for the players). A relevant source of information is the Elevation of Privilege Card Game[5] from Microsoft.

[3] `https://ocrportal.hhs.gov/ocr/breach/`

[4] `https://cps-vo.org/node/34187`

[5] `https://www.microsoft.com/en-us/SDL/adopt/eop.aspx`

Pros/cons

- Instructor highly interested (we will also have a lecture on norms and privacy tradeoffs)
- Limited online information available about privacy games
- Path to publishing longer as evaluation of the game will take more time
- More implementation heavy: Requires web development skills
- Allows for more individual contributions (good for your CV)

Specific project deliverables

1. A short survey of existing privacy games and their supported features
2. Working demo of the new NormDefense components/features
3. A user study with customized game scenarios and associated hypothesis

Project 4: Agent-based Simulation of Privacy Behaviors

In this project, you will design and implement simulations for user sharing behaviors of sensitive content using a Facebook dataset[6] [Viswanath et al., 2009] or similar datasets for content sharing platforms. Agents in the simulation will act based on user content sharing behaviors reported in the literature, and their sharing intentions will be compatible with Westin's privacy category distribution among the general public. You might use an agent development environment such as JADE[7] to implement agents. You will design various sharing scenarios, develop hypothesis, and report sharing and violation statistics. Some useful references for this project are [Johnson et al., 2012], [Kafalı et al., 2014], [Kumaraguru and Cranor, 2005].

Pros/cons

- Learn to design and analyze simulation based experiments
- Can be publishable with some additional effort
- Implementation heavy: Instructor less available for support on implementation details

Specific project deliverables

1. Working demo of the simulation environment (no visualization required)
2. A set of user sharing behaviors and associated agent implementations

[6] `http://socialnetworks.mpi-sws.mpg.de/data-wosn2009.html`
[7] `http://jade.tilab.com/`

Project 5: Systematic Investigation of Privacy Policies and Laws

In this project, you will investigate international privacy laws among various countries such as the US, EU, and China[8]. You will develop a systematic and repeatable methodology to identify conflicting clauses, e.g., one policy allows sharing of sensitive user information in certain situations, whereas another policy prohibits. You will represent privacy policies and laws in formal logic, and develop a set of conflict patterns using the logic representation. You will also design (implementation is not necessary) interfaces that enable interaction with a user to confirm conflicts. Some useful references for this project are [Ghanavati et al., 2014], [Breaux and Anton, 2008].

Pros/cons

- You will be responsible for finding content online to investigate
- We will have a lecture on conflicting privacy policies, but pretty late
- Quality of results unpredictable (chances of publishing will rely on results)
- Requires familiarity with formal logic
- Minimal implementation effort

Specific project deliverables

1. A semiautomated methodology (clearly describing human and automated tasks) to identify conflicts in privacy policies
2. A set of conflict patterns
3. A set of identified conflicts and explanations about how they are identified using your methodology
4. Mockup interactive user interface design for identifying conflicts in privacy policies

Project X: Your Own Idea

If you have a project idea other than the above projects and they are related to the topics of the course, you are welcome to discuss it with the instructor. If so, make sure to prepare a short project proposal with expected deliverables.

Pros/cons

- Work on a topic that you are interested in
- Potentially less support from the instructor depending on the topic

[8] `http://searchsecurity.techtarget.com/news/450420139/International-data-privacy-laws-create-inconsistent-rules`

References

[Breaux and Anton, 2008] Breaux, T. and Anton, A. (2008). Analyzing regulatory rules for privacy and security requirements. *Software Engineering, IEEE Transactions on*, 34(1):5–20.

[Ghanavati et al., 2014] Ghanavati, S., Rifaut, A., Dubois, E., and Amyot, D. (2014). Goal-oriented compliance with multiple regulations. In *Requirements Engineering Conference (RE), 2014 IEEE 22nd International*, pages 73–82.

[Gharib et al., 2016] Gharib, M., Giorgini, P., and Mylopoulos, J. (2016). Ontologies for privacy requirements engineering: A systematic literature review. *CoRR*.

[Guo et al., 2014] Guo, J., Monaikul, N., Plepel, C., and Cleland-Huang, J. (2014). Towards an intelligent domain-specific traceability solution. In *Proceedings of the 29th ACM/IEEE International Conference on Automated Software Engineering (ASE)*, pages 755–766. ACM.

[Johnson et al., 2012] Johnson, M., Egelman, S., and Bellovin, S. M. (2012). Facebook and privacy: It's complicated. In *Proceedings of the 8th Symposium on Usable Privacy and Security (SOUPS)*, pages 9:1–9:15.

[Kafalı et al., 2014] Kafalı, Ö., Günay, A., and Yolum, P. (2014). Detecting and predicting privacy violations in online social networks. *Distributed and Parallel Databases*, 32(1):161–190.

[Kafalı et al., 2017] Kafalı, Ö., Jones, J., Petruso, M., Williams, L., and Singh, M. P. (2017). How good is a security policy against real breaches? a HIPAA case study. In *Proceedings of the 39th International Conference on Software Engineering (ICSE)*, pages 530–540, Buenos Aires. IEEE Computer Society.

[Kumaraguru and Cranor, 2005] Kumaraguru, P. and Cranor, L. F. (2005). Privacy indexes: A survey of westin's studies. *ISRI Technical Report*.

[Murukannaiah et al., 2017] Murukannaiah, P. K., Dabral, C., Sheshadri, K., Sharma, E., and Staddon, J. (2017). Learning a privacy incidents database. In *Proceedings of the Hot Topics in Science of Security: Symposium and Bootcamp (HoTSoS)*, pages 35–44.

[Viswanath et al., 2009] Viswanath, B., Mislove, A., Cha, M., and Gummadi, K. P. (2009). On the evolution of user interaction in facebook. In *Proceedings of the 2nd ACM Workshop on Online Social Networks*, pages 37–42.