CSC 495.002 – Group Projects

Dr. Özgür Kafalı

North Carolina State University
Department of Computer Science

Fall 2017

Group Work

- Goals:
  - Give you experience (both research and development) on a specific topic related to privacy
  - Collaboration within group members as well as among groups
  - Work with deadlines, prepare deliverables, present work done
- Work in groups of 2–3
- A project can be chosen by multiple groups
- Customize the project scope and deliverables to minimize overlap between groups

Deliverables

- One page project proposal describing the project goals, research questions, and anticipated contributions of each group member
- Intermediate report describing current progress towards project goals
- Final report
- Project specific deliverables
- In class presentations

Final Report

- Introduction: State your goal and research questions with regards to the project topic
  - Describe why you chose those research questions
  - 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
  - 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

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

Development of a Privacy Ontology

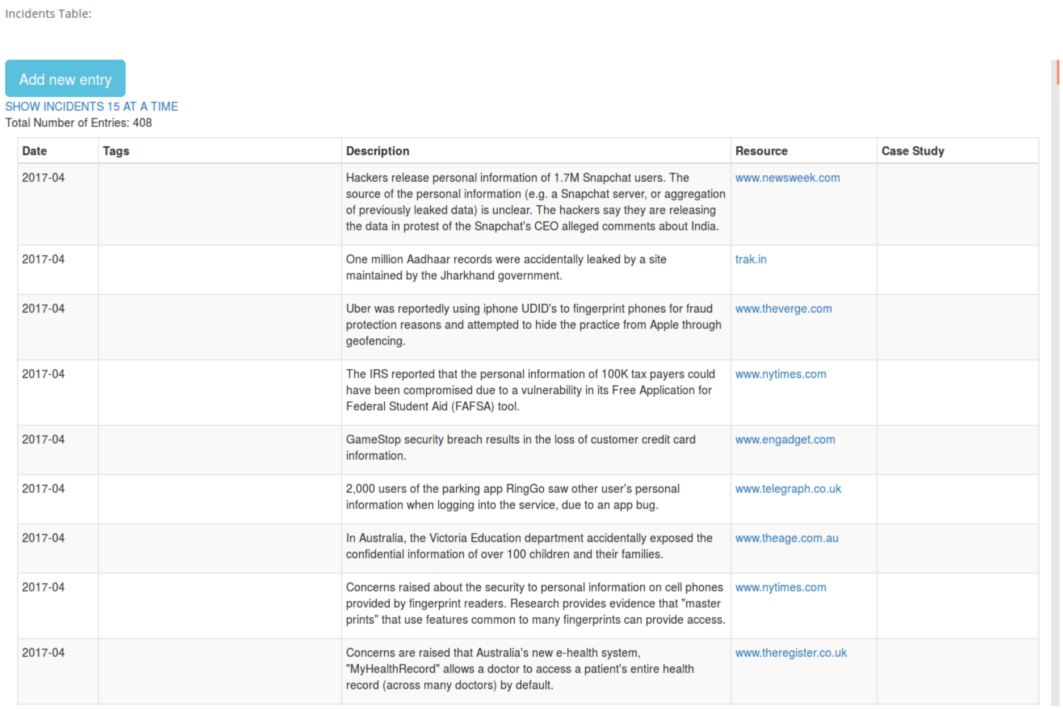
- Investigate privacy incidents from the “Privacy Incidents Database”
- Develop an ontology of privacy breaches
  - Concepts unified from individual incidents
  - Relations among concepts
  - Properties of concepts
- Aggregate results with (potential) other groups
- Potential research questions:
  - What are common concepts associated with incidents? E.g., information disclosure
  - How similar are incidents?
  - How likely is this incident to occur again? Given similar circumstances

Privacy Incidents Database: https://sites.google.com/site/privacyincidentsdatabase/

Privacy Incidents Database

- Incident: An instance of accidental or unauthorized collection, use or exposure of sensitive information about an individual
- Answer questions like
  - What are the common causes of privacy incidents?
  - How do privacy incidents vary by country?
  - Which organizations are commonly involved in privacy incidents?
- Perform analytics: Understand trends and frequency of incident occurrence

Privacy Incidents

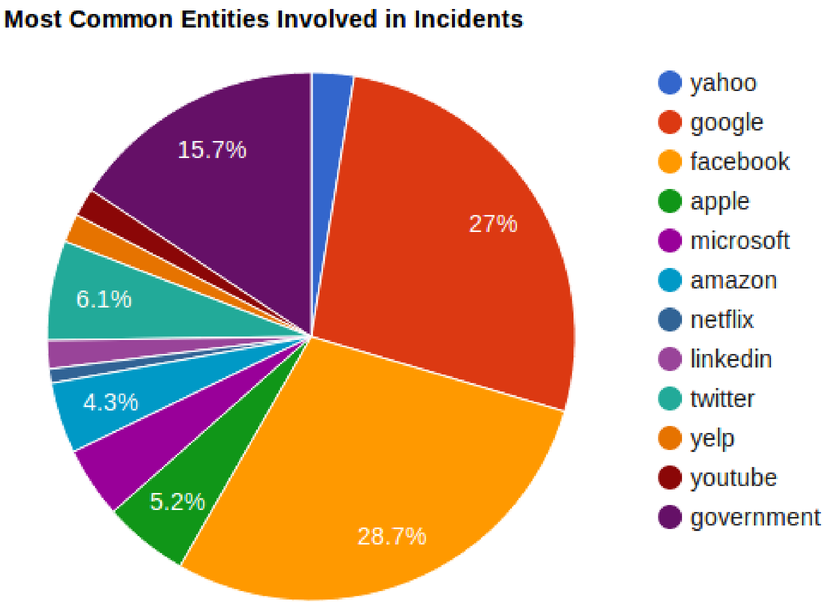


Incidents Table:

Total Number of Entries: 408

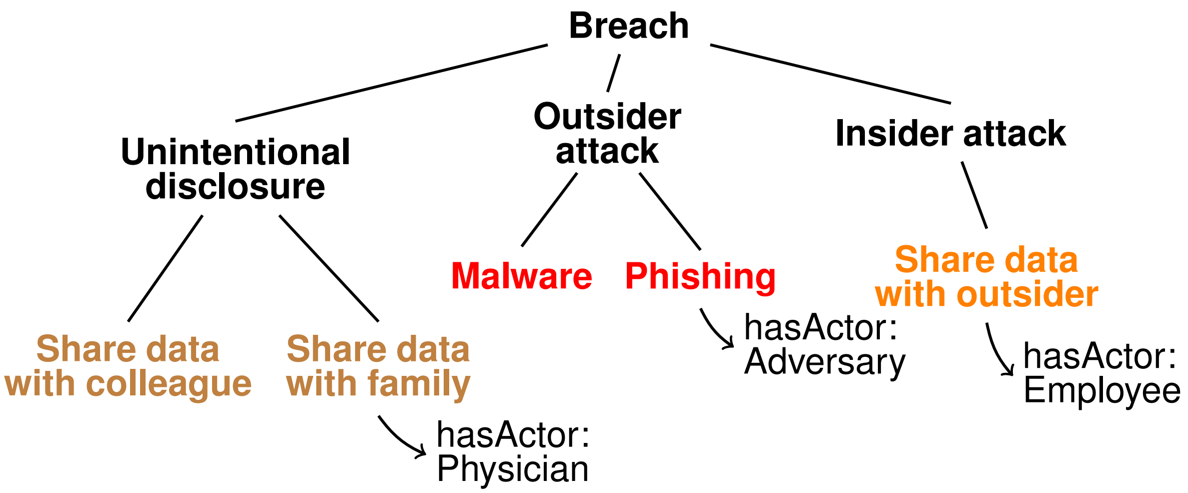
Date	Tags	Description	Resource	Case Study
2017-04		Hackers release personal information of 1.7M Snapchat users. The source of the personal information (e.g. a Snapchat server, or aggregation of previously leaked data) is unclear. The hackers say they are releasing the data in protest of the Snapchat's CEO alleged comments about India.	www.newsweek.com	
2017-04		One million Aadhaar records were accidentally leaked by a site maintained by the Jharkhand government.	trak.in	
2017-04		Uber was reportedly using iphone UDID's to fingerprint phones for fraud protection reasons and attempted to hide the practice from Apple through geofencing.	www.theverge.com	
2017-04		The IRS reported that the personal information of 100K tax payers could have been compromised due to a vulnerability in its Free Application for Federal Student Aid (FAFSA) tool.	www.nytimes.com	
2017-04		GameStop security breach results in the loss of customer credit card information.	www.engadget.com	
2017-04		2,000 users of the parking app RingGo saw other user's personal information when logging into the service, due to an app bug.	www.telegraph.co.uk	
2017-04		In Australia, the Victoria Education department accidentally exposed the confidential information of over 100 children and their families.	www.theage.com.au	
2017-04		Concerns raised about the security to personal information on cell phones provided by fingerprint readers. Research provides evidence that "master prints" that use features common to many fingerprints can provide access.	www.nytimes.com	
2017-04		Concerns are raised that Australia's new e-health system, "MyHealthRecord" allows a doctor to access a patient's entire health record (across many doctors) by default.	www.theregister.co.uk	

Visualizations



Ontologies

- Describes domain knowledge in a structured way
  - A taxonomy of related concepts
  - Properties of concepts



Breach

Unintentional disclosure

Outsider attack

Insider attack

Share data with colleague

Share data with family

hasActor: Physician

Malware

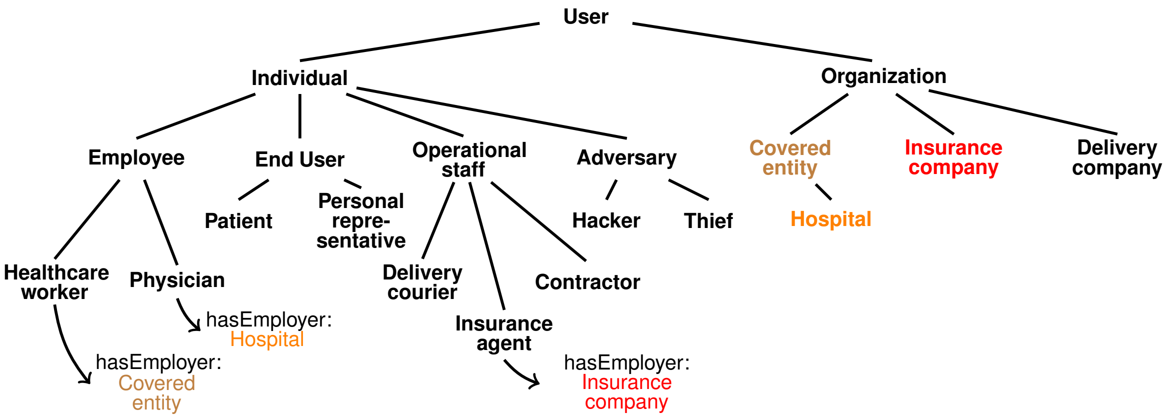
Phishing

hasActor: Adversary

Share data with outsider

hasActor: Employee

Ontology of Healthcare Users



User

Individual

Organization

Employee

End User

Operational staff

Adversary

Covered entity

Insurance company

Delivery company

Patient

Personal repre-sentative

Hacker

Thief

Hospital

Healthcare worker

Physician

Delivery courier

Contractor

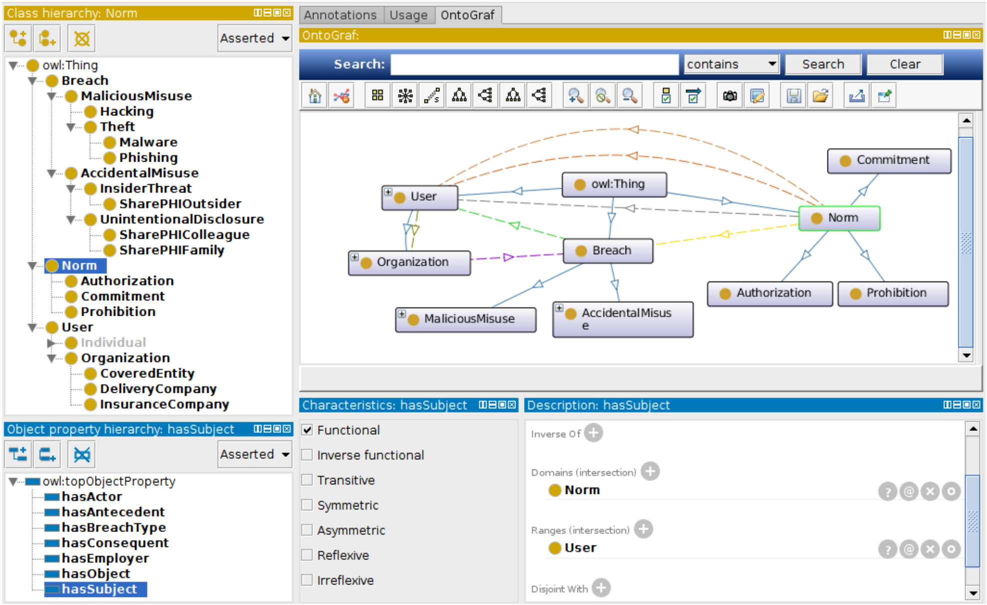
Insurance agent

hasEmployer: Hospital

hasEmployer: Covered entity

hasEmployer: Insurance company

Protégé Ontology Development Tool



Class hierarchy: Norm

```
owl:Thing
  Breach
    MaliciousMisuse
      Hacking
      Theft
        Malware
        Phishing
    AccidentalMisuse
      InsiderThreat
        SharePHIOutsider
      UnintentionalDisclosure
        SharePHIColleague
        SharePHIFamily
  Norm
    Authorization
    Commitment
    Prohibition
  User
    Individual
    Organization
      CoveredEntity
      DeliveryCompany
      InsuranceCompany
```

Object property hierarchy: hasSubject

```
owl:topObjectProperty
  hasActor
  hasAntecedent
  hasBreachType
  hasConsequent
  hasEmployer
  hasObject
  hasSubject
```

Characteristics: hasSubject

☑ Functional
☐ Inverse functional
☐ Transitive
☐ Symmetric
☐ Asymmetric
☐ Reflexive
☐ Irreflexive

Description: hasSubject

Inverse Of

Domains (intersection)
Norm

Ranges (intersection)
User

Disjoint With

Protégé: http://protege.stanford.edu/

Similarity Metric

- Compare individual incidents from the database using elements of the ontology
- How similar are the following incidents?
  - “Yahoo reportedly complied with requests by the NSA and FBI to scan incoming emails for certain keywords/phrases.”
  - “Emails of faculty and staff at Harvard were searched as part of a student cheating investigation, raising a privacy outcry amongst the email account holders.”

Aggregating Results

- Compare ontology concepts and associated relations
- Apply each others’ similarity metrics on the corresponding ontologies (for same pairs of incidents)
- Report similarities, differences, and a methodology to merge individual ontologies

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 Deliverables

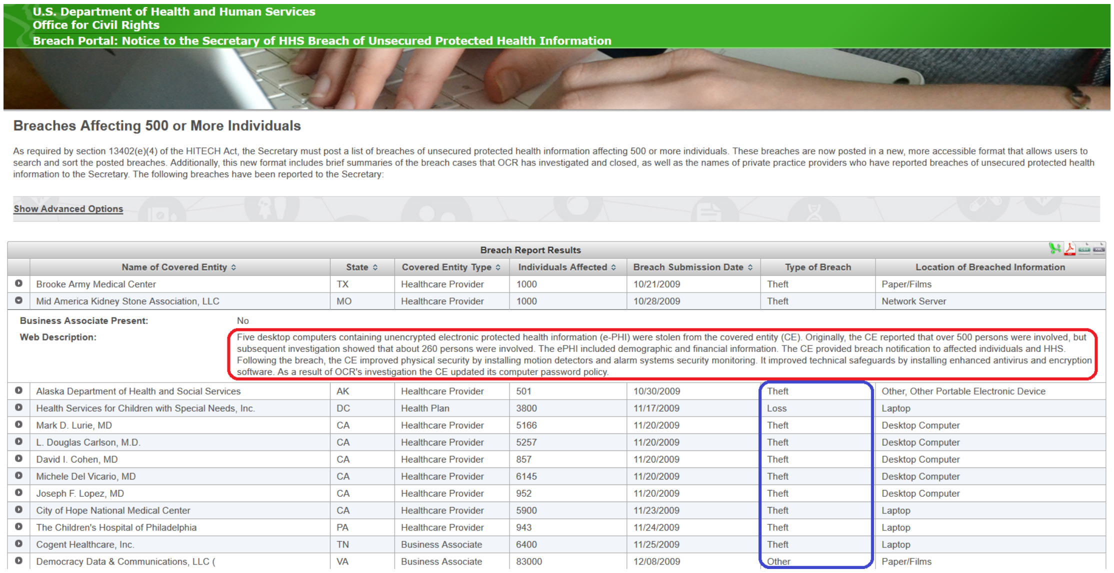
- An ontology developed with Protégé
- An implemented similarity metric that takes as input two privacy incidents and queries the ontology to compute the similarity between the incidents

Classification of Healthcare Privacy Breaches

- Investigate breaches from the “US Department of Health and Human Services” (HHS)
- Potential objectives:
  - Distinguish between security and privacy incidents
  - Classification of privacy incidents caused by human errors
  - Identify common patterns found in breach descriptions (data collection, data usage, data sharing)
  - Report frequency of breach occurrence
- Aggregate results with (potential) other groups as well as Project 1

HHS Breach Report: https://ocrportal.hhs.gov/ocr/breach/

HHS Breach Report



U.S. Department of Health and Human Services
Office for Civil Rights
Breach Portal: Notice to the Secretary of HHS Breach of Unsecured Protected Health Information

Breaches Affecting 500 or More Individuals

As required by section 13402(e)(4) of the HITECH Act, the Secretary must post a list of breaches of unsecured protected health information affecting 500 or more individuals. These breaches are now posted in a new, more accessible format that allows users to search and sort the posted breaches. Additionally, this new format includes brief summaries of the breach cases that OCR has investigated and closed, as well as the names of private practice providers who have reported breaches of unsecured protected health information to the Secretary. The following breaches have been reported to the Secretary:

Breach Report Results

	Name of Covered Entity ⇕	State ⇕	Covered Entity Type ⇕	Individuals Affected ⇕	Breach Submission Date ⇕	Type of Breach	Location of Breached Information
●	Brooke Army Medical Center	TX	Healthcare Provider	1000	10/21/2009	Theft	Paper/Films
●	Mid America Kidney Stone Association, LLC	MO	Healthcare Provider	1000	10/28/2009	Theft	Network Server

Business Associate Present: No

Web Description: Five desktop computers containing unencrypted electronic protected health information (e-PHI) were stolen from the covered entity (CE). Originally, the CE reported that over 500 persons were involved, but subsequent investigation showed that about 260 persons were involved. The ePHI included demographic and financial information. The CE provided breach notification to affected individuals and HHS. Following the breach, the CE improved physical security by installing motion detectors and alarm systems security monitoring. It improved technical safeguards by installing enhanced antivirus and encryption software. As a result of OCR's investigation the CE updated its computer password policy.

	Name of Covered Entity ⇕	State ⇕	Covered Entity Type ⇕	Individuals Affected ⇕	Breach Submission Date ⇕	Type of Breach	Location of Breached Information
●	Alaska Department of Health and Social Services	AK	Healthcare Provider	501	10/30/2009	Theft	Other, Other Portable Electronic Device
●	Health Services for Children with Special Needs, Inc.	DC	Health Plan	3800	11/17/2009	Loss	Laptop
●	Mark D. Lurie, MD	CA	Healthcare Provider	5166	11/20/2009	Theft	Desktop Computer
●	L. Douglas Carlson, M.D.	CA	Healthcare Provider	5257	11/20/2009	Theft	Desktop Computer
●	David I. Cohen, MD	CA	Healthcare Provider	857	11/20/2009	Theft	Desktop Computer
●	Michele Del Vicario, MD	CA	Healthcare Provider	6145	11/20/2009	Theft	Desktop Computer
●	Joseph F. Lopez, MD	CA	Healthcare Provider	952	11/20/2009	Theft	Desktop Computer
●	City of Hope National Medical Center	CA	Healthcare Provider	5900	11/23/2009	Theft	Laptop
●	The Children's Hospital of Philadelphia	PA	Healthcare Provider	943	11/24/2009	Theft	Laptop
●	Cogent Healthcare, Inc.	TN	Business Associate	6400	11/25/2009	Theft	Laptop
●	Democracy Data & Communications, LLC (	VA	Business Associate	83000	12/08/2009	Other	Paper/Films

Classification of Breaches: Security vs Privacy

- Is this a security or a privacy incident?
- “One of the covered entity’s (CE) computers was infected with malware and as a result, data on the infected computer was encrypted and made inaccessible.”

Classification of Breaches: Malicious vs Accidental

- Is this incident caused by malicious intent or due to human error (accidental)?
- “In 2010, an employee in a HIPAA covered entity forgot to erase data contained on disposed photocopiers’ hard drives, which led to disclosure of patient records.”

Aggregating Results

- Compare classifications of security vs privacy, and types of human errors
- Compare common breach patterns
- Report similarities, differences, frequencies of occurrence, potential additions to the Privacy Incidents Database (Project 1)

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, especially on human errors
- Requires teamwork and collaboration among groups

Specific Deliverables

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

Development of a Privacy Card Game

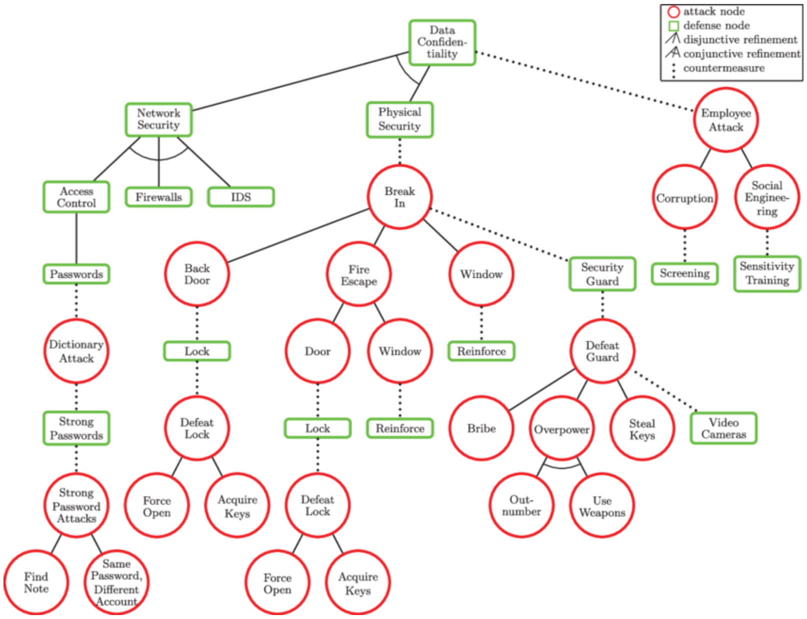
- Goal: Understanding how people make choices to mitigate privacy risks
- Perform a survey of existing privacy games in the literature
  - Identify the design space of such games
  - What are their objectives?
  - What sort of user interfaces and other features do they support?
- Design and implement features for the NormDefense game (recently started developing)

NormDefense: https://cps-vo.org/node/34187

Objectives & Features

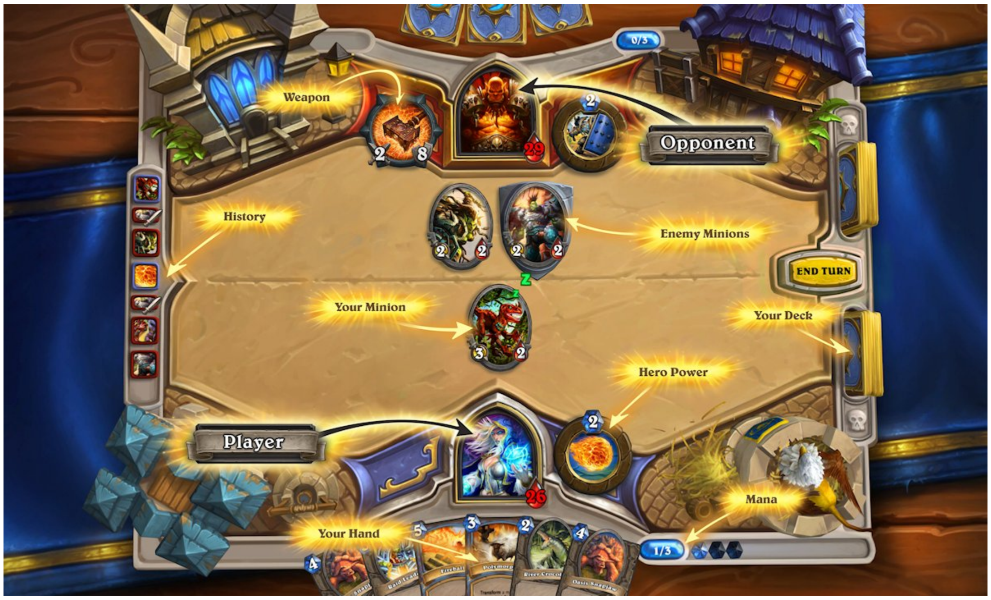
- Broad objectives:
  - Prioritize privacy risks and associated mitigation techniques
  - Act as a testbed for researchers to develop and test privacy related hypotheses
  - Serve as a tool for privacy education and training
- Potential new features to be implemented:
  - Explore tradeoffs among social privacy norms and technical mechanisms
  - Collaboration among players (both defenders and attackers)
  - Develop basic automated strategies (software agents playing the game), and run simulations
  - Design user studies using realistic privacy scenarios (customized card decks), and develop hypotheses

Threat Models: Attack/Defense Trees



Kordy et al. Attack–defense trees. Journal of Logic and Computation, 24(1):55–87, 2014

Inspirations: Interface from Hearthstone



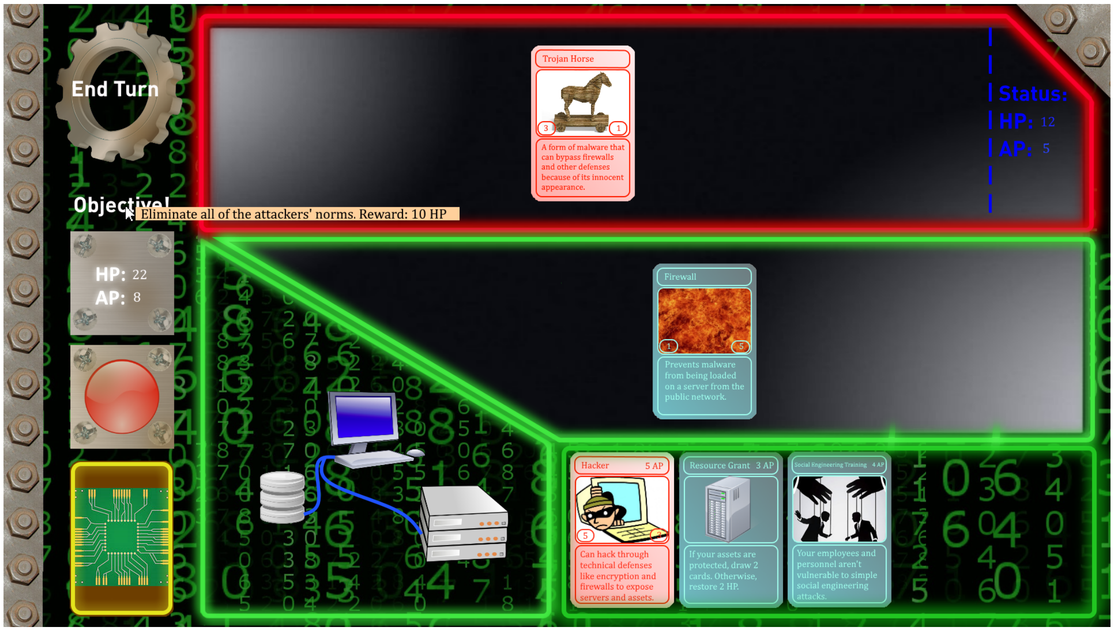
Hearthstone: https://us.battle.net/hearthstone/en/

Game Elements

- New card suits
  - Attacker: Microsoft's STRIDE (Elevation of Privilege) + social engineering
  - Defender: Social norms, technical mechanisms, assumptions
  - Some card suits: Accountability, logging, forensics
- Maintenance defenses vs achievement defenses
- From card selection to strategy (tradeoffs)
  - Cards that provide overall security → blindly counter all attacks
  - Cards that provide protection against a specific attack → gather intelligence about attacker

Elevation of Privilege: https://www.microsoft.com/en-us/SDL/adopt/eop.aspx

NormDefense Interface



Eliminate all of the attackers' norms. Reward: 10 HP

Trojan Horse: A form of malware that can bypass firewalls and other defenses because of its innocent appearance.

Firewall: Prevents malware from being loaded on a server from the public network.

Hacker 5 AP: Can hack through technical defenses like encryption and firewalls to expose servers and assets.

Resource Grant 3 AP: If your assets are protected, draw 2 cards. Otherwise, restore 2 HP.

Social Engineering Training 4 AP: Your employees and personnel aren't vulnerable to simple social engineering attacks.

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 Deliverables

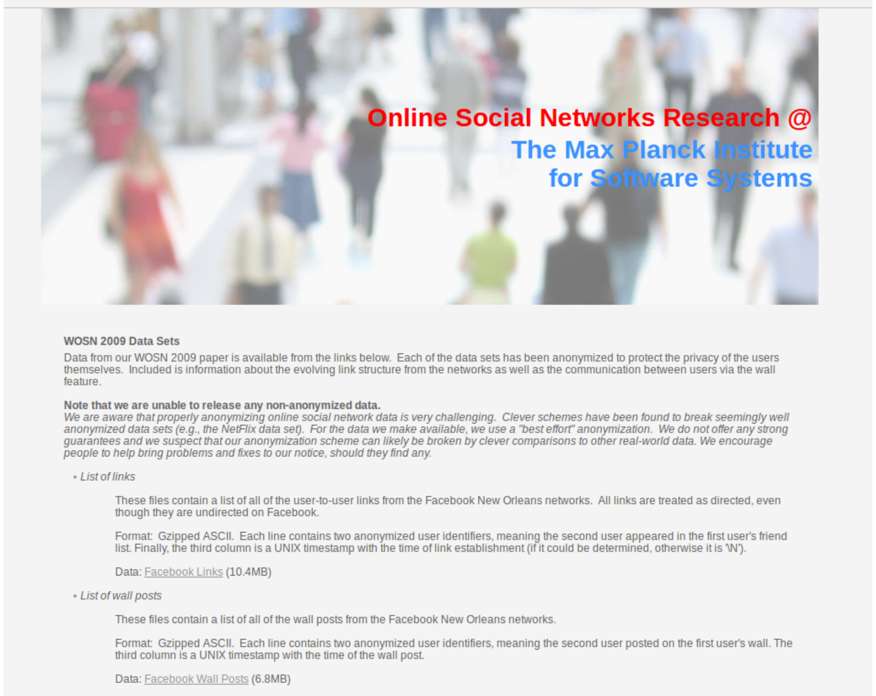
- A short survey of existing privacy games and their supported features
- Working demo of the new NormDefense components/features
- A user study with customized game scenarios and associated hypotheses

Agent-based Simulation of Privacy Behaviors

- Design and implement simulations for user sharing behaviors of sensitive content
- Use a dataset for content sharing platforms such as Facebook
- Develop agents for simulation:
  - Agents will act based on user content sharing behaviors reported in the literature
  - Agents' sharing intentions will be compatible with Westin's privacy category distribution among the general public
- You may use an agent development environment such as JADE to implement agents
- Design various sharing scenarios, develop hypotheses, and report sharing and violation statistics

JADE: http://jade.tilab.com/

Facebook Dataset



Online Social Networks Research @
The Max Planck Institute
for Software Systems

WOSN 2009 Data Sets

Data from our WOSN 2009 paper is available from the links below. Each of the data sets has been anonymized to protect the privacy of the users themselves. Included is information about the evolving link structure from the networks as well as the communication between users via the wall feature.

Note that we are unable to release any non-anonymized data.
We are aware that properly anonymizing online social network data is very challenging. Clever schemes have been found to break seemingly well anonymized data sets (e.g., the NetFlix data set). For the data we make available, we use a "best effort" anonymization. We do not offer any strong guarantees and we suspect that our anonymization scheme can likely be broken by clever comparisons to other real-world data. We encourage people to help bring problems and fixes to our notice, should they find any.

- *List of links*

  These files contain a list of all of the user-to-user links from the Facebook New Orleans networks. All links are treated as directed, even though they are undirected on Facebook.

  Format: Gzipped ASCII. Each line contains two anonymized user identifiers, meaning the second user appeared in the first user's friend list. Finally, the third column is a UNIX timestamp with the time of link establishment (if it could be determined, otherwise it is '\N').

  Data: Facebook Links (10.4MB)

- *List of wall posts*

  These files contain a list of all of the wall posts from the Facebook New Orleans networks.

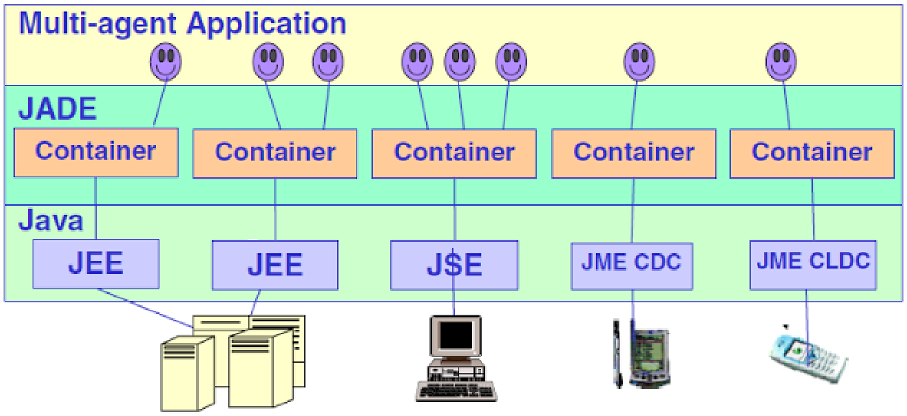
  Format: Gzipped ASCII. Each line contains two anonymized user identifiers, meaning the second user posted on the first user's wall. The third column is a UNIX timestamp with the time of the wall post.

  Data: Facebook Wall Posts (6.8MB)

Useful Datasets

- Alan Mislove’s OSN datasets:
  - http://socialnetworks.mpi-sws.mpg.de/data-wosn2009.html
  - http://socialnetworks.mpi-sws.mpg.de/data-imc2007.html
  - http://socialnetworks.mpi-sws.mpg.de/data-wosn2008.html
  - http://socialnetworks.mpi-sws.mpg.de/data-www2009.html
- Any other relevant dataset you might find online (e.g., Twitter)

JADE Agent Development Framework



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 Deliverables

- Working demo of the simulation environment (no visualization required)
- A set of user sharing behaviors and associated agent implementations
- Results reported with tables and plots (hypotheses validation)

Systematic Investigation of Privacy Policies and Laws

- Investigate privacy policies and international privacy laws among various countries such as the US, EU, and China
- Develop a systematic and repeatable methodology to identify conflicting clauses
  - For example, one policy allows sharing of sensitive user information in certain situations, whereas another policy prohibits
  - Represent privacy policies and laws in formal logic
  - Develop a set of conflict patterns using the logic representation
- Design interfaces that enable interaction with a user to confirm/reject conflicts

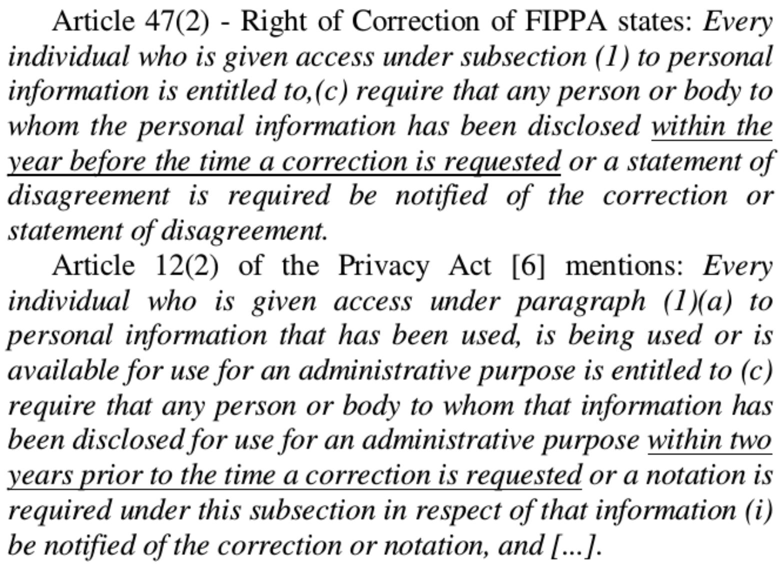
http://searchsecurity.techtarget.com/news/450420139/International-data-privacy-laws-create-inconsistent-rules

Policy Patterns

- Case 1: There is nothing in common between two statements
- Case 2: Two statements are similar to each other
- Case 3: One statement is complementary to the other statement
- Case 4: One statement is a subset of the other statement
- Case 5: One statement is stricter than the other statement
- Case 6: One statement contradicts the other statement

Ghanavati et al. Goal-oriented compliance with multiple regulations. International Requirements Engineering Conference, pages 73-82, 2014

Case 5: Stricter Statement

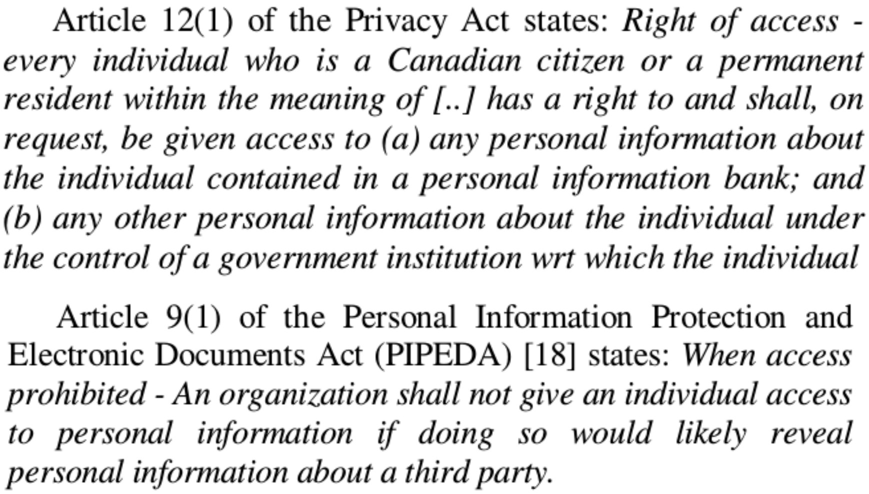


Article 47(2) - Right of Correction of FIPPA states: *Every individual who is given access under subsection (1) to personal information is entitled to,(c) require that any person or body to whom the personal information has been disclosed within the year before the time a correction is requested or a statement of disagreement is required be notified of the correction or statement of disagreement.*

Article 12(2) of the Privacy Act [6] mentions: *Every individual who is given access under paragraph (1)(a) to personal information that has been used, is being used or is available for use for an administrative purpose is entitled to (c) require that any person or body to whom that information has been disclosed for use for an administrative purpose within two years prior to the time a correction is requested or a notation is required under this subsection in respect of that information (i) be notified of the correction or notation, and [...].*

Ghanavati et al. Goal-oriented compliance with multiple regulations. International Requirements Engineering Conference, pages 73-82, 2014

Case 5: Contradicting Statements



Article 12(1) of the Privacy Act states: *Right of access - every individual who is a Canadian citizen or a permanent resident within the meaning of [..] has a right to and shall, on request, be given access to (a) any personal information about the individual contained in a personal information bank; and (b) any other personal information about the individual under the control of a government institution wrt which the individual*

Article 9(1) of the Personal Information Protection and Electronic Documents Act (PIPEDA) [18] states: *When access prohibited - An organization shall not give an individual access to personal information if doing so would likely reveal personal information about a third party.*

Ghanavati et al. Goal-oriented compliance with multiple regulations. International Requirements Engineering Conference, pages 73-82, 2014

Pros/cons

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

Specific Deliverables

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

Your Own Idea

- In case you have a project idea related to the topics of the course
- 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