NC STATE UNIVERSITY

CSC 495.002 – Lecture 9
AI for Privacy: Privacy Norms

Dr. Özgür Kafalı

North Carolina State University
Department of Computer Science

Fall 2017

NC STATE UNIVERSITY

Agents and Reasoning

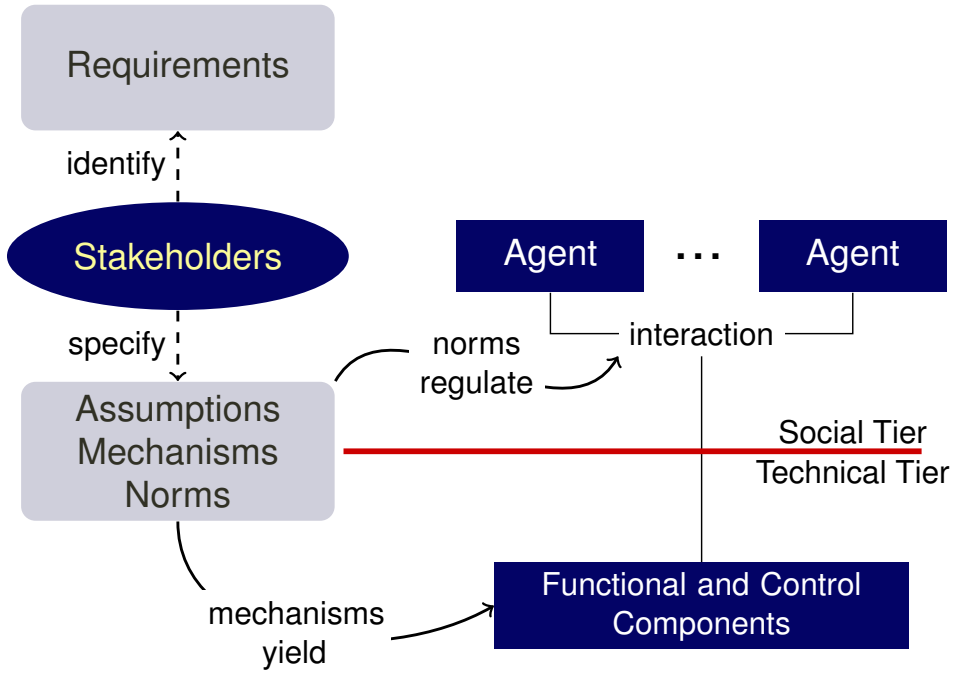
- Agents in pervasive healthcare
- Resolving multi-party privacy concerns via argumentation
- Negotiating privacy preferences

Problem Definition

- Imagine you are developing a healthcare application
- You designed a perfect role-based access control mechanism to regulate access to sensitive patient information
- But, you later observed nurses are sharing their passwords to access each other's accounts

- Cannot control everything with software features
- Provide flexibility to users (don't prevent everything)
- Need a social mechanism to regulate the interactions among users
- Hold users accountable for their actions

Sociotechnical Systems (STS)



Objectives

- Develop abstractions, models, and tools to help address legal and social aspects of security and privacy
- Build computational models of the social architecture
- Enable unified treatment of technical and social considerations

STS Example: Hospital Organization

- Roles: Physician, hospital, patient
- Assumptions: Physicians cannot authenticate when there is a power failure
- Mechanisms: Hospital software allows physicians to authenticate with valid passwords
- Norms: Physicians should not disclose patient information to outsiders

Exercise: Course Management System

- Roles?
- Assumptions?
- Mechanisms?
- Norms?

Contextual Integrity

- A conceptual framework to evaluate the flow of information between parties
- Norms change depending on context
- Previous example: Physicians should not disclose patient information to outsiders
- Are there any variations of this norm? If the context changes
- Physicians may disclose patient information to family members in emergencies

Barth et al. Privacy and Contextual Integrity: Framework and Applications. IEEE Symposium on Security and Privacy, pages 184–198, 2006

Formal Specification

- N(SUBJECT, OBJECT, antecedent, consequent)

- Type: $N \in \{$Commitment (C), Authorization (A), Prohibition (P)$\}$
- SUBJECT: Party that is [responsible for / beneficiary of] the norm
- OBJECT: Party that is [beneficiary of / responsible for] the norm
- antecedent: Preconditions that need to hold to activate the norm
- consequent: Action that needs to be [performed / avoided]

Commitment

- Informally, describes “what you should do”

- Example: A physician is committed to the hospital to operating upon patients in an emergency

- Formally, $\boxed{C(\text{PHYSICIAN}, \text{HOSPITAL}, \text{emergency}, \text{operate})}$

Authorization

- Informally, describes “what you can do”
- Example: A physician is authorized by the hospital to access the patient’s electronic health records (EHR) if the patient gives consent
- Formally, $\boxed{A(\text{PHYSICIAN}, \text{PATIENT}, \text{consent}, \text{view\_EHR})}$

Prohibition

- Informally, describes “what you should not do”
- Example: A physician is prohibited by the hospital from disclosing a patient’s protected health information (PHI) to others
- Formally, $\boxed{P(\text{PHYSICIAN}, \text{HOSPITAL}, \text{true}, \text{disclose\_PHI})}$

Exercise: Norm Specifications

- A physician may prescribe drugs to the patients or schedule their next visit after a routine visit

  A(PHYSICIAN, HOSPITAL, visit, prescribe $\vee$ schedule_visit)

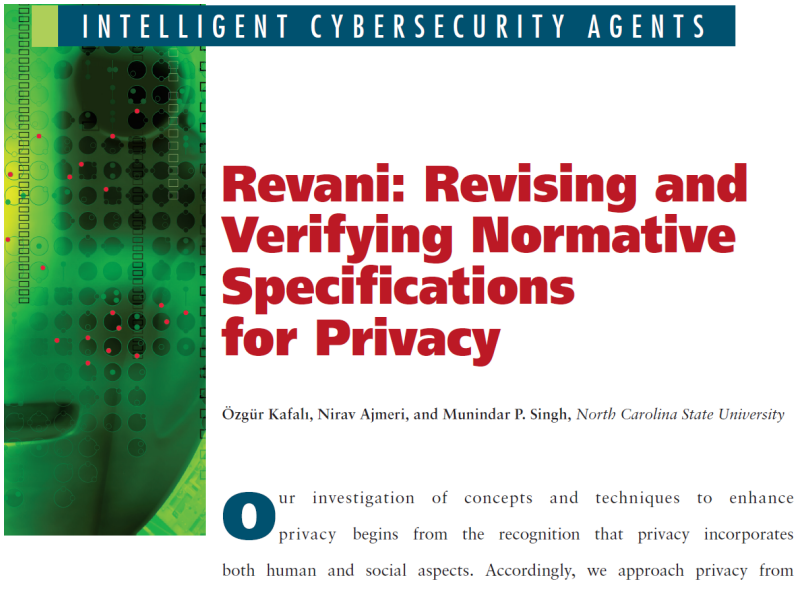
- Hospital workers must log out of a public computer as soon as they finish viewing EHR of patients

  C(WORKER, HOSPITAL, public_computer $\wedge$ view_EHR, logout)

- A nurse should not disclose patient information to patient's family unless there is consent from the patient or it's an emergency

  P(NURSE, HOSPITAL, $\neg$consent $\wedge$ $\neg$emergency, disclose_family)

Normative Specifications for Privacy



Kafalı et al. Revani: Revising and Verifying Normative Specifications for Privacy. IEEE Intelligent Systems, 31(5):8–15, 2016

Why Do We Need Norms?

- Think about flights
- What can you not do on a flight?
- How do you ensure people don't smoke?
- Install smoke detectors in restrooms [Technical solution]
- Don't temper with the smoke detector! [Norm]

Exercise: File Sharing System

- Assume you're collaborating on a project proposal
- You're using Google Drive to share the proposal documents among your colleagues

- What are the functional requirements?
- What is the sensitive information? How do you ensure privacy?
- What are the norms?

Exercise: Specification of Norms

- "Parents can exercise individual rights, such as access to the medical record, on behalf of their minor children."

  A(PARENT, HOSPITAL, representative(PARENT, MINOR), access_EHR(PARENT, MINOR))

- "A covered entity may not disclose protected health information, except as the subject individual authorizes in writing."

  P(COVERED_ENTITY, HOSPITAL, $\neg$consent(COVERED_ENTITY, PATIENT), disclose_PHI(COVERED_ENTITY, _))

Research Questions

- Specification: What are the necessary components to develop a computational model of an STS?
- Verification: How can we verify that an STS satisfies the functional, security, and privacy requirements of its stakeholders?
- Refinement: Supposing an STS fails to satisfy its requirements, how can we propose refinement so that its refined specification satisfies the requirements?

STS Components: Assumptions

- Example: Physicians cannot authenticate when there is a power failure
- Formally, $\langle \neg$authenticate, power_failure$\rangle$
  or,
  $\neg$authenticate $\leftarrow$ power_failure

STS Components: Mechanisms

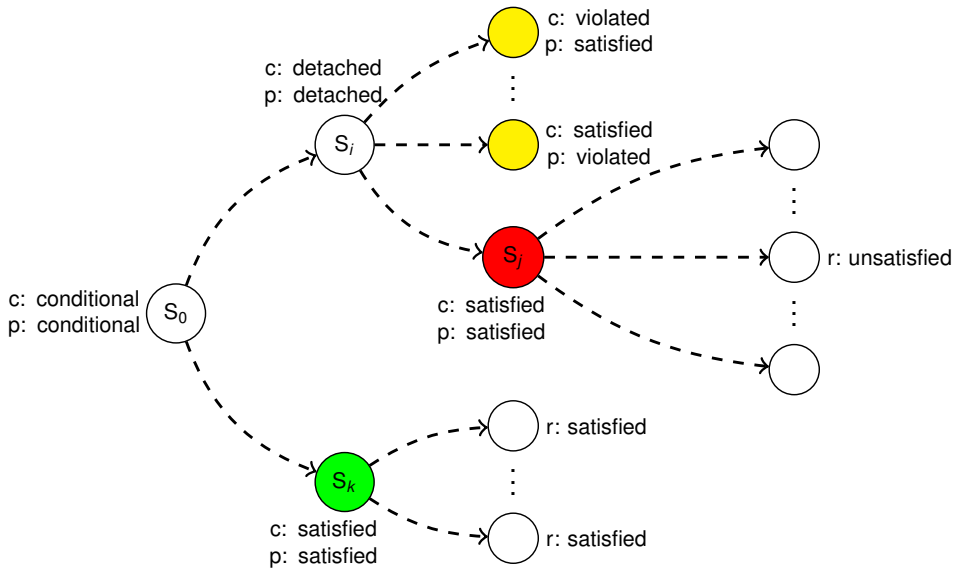
- Example: Hospital software allows physicians to authenticate with valid passwords
- Formally, *m*(enabler, add, delete)
- *m*(enter_password, {authenticate}, { })

Requirements in Temporal Logic

- Express stakeholder requirements as Computation Tree Logic (CTL) formulas
  - A branch quantifier, all (A) or exists (E), over branches emanating from the current point
  - A linear temporal operator, describing properties of a single branch (next (X), eventually (F), always (G), and until (U))

- Examples:
  - Physicians should always be able to access patients' EHR
    In CTL: AF view_EHR
  - Physicians should never disclose patients' PHI
    In CTL: AG ¬disclose_PHI

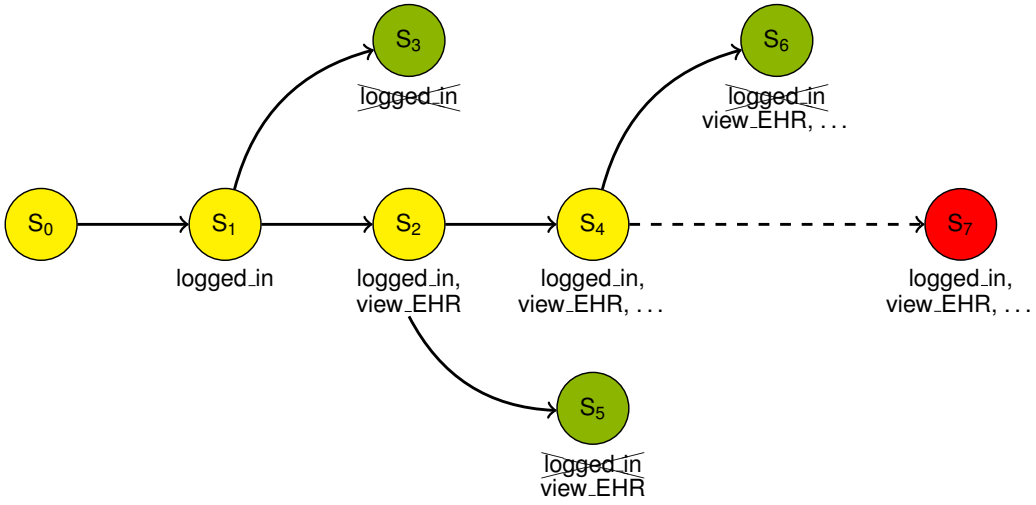
Verification Setting



Verification Example

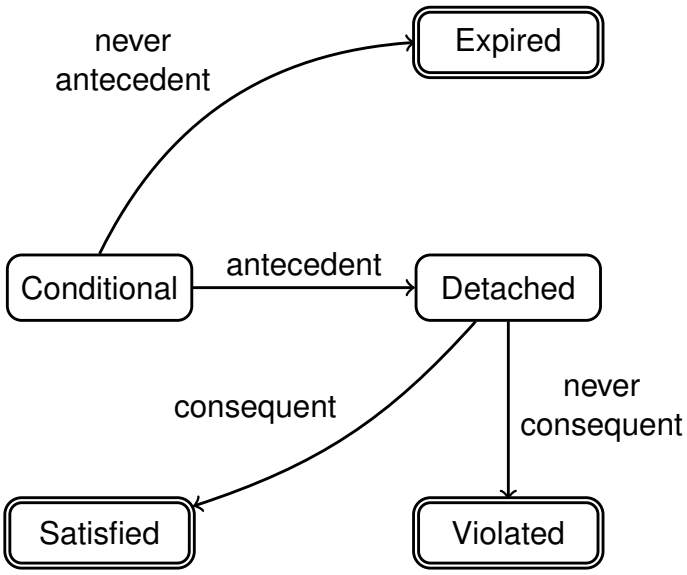
- Open sessions must be closed after reviewing EHR

  AG (view_EHR $\rightarrow$ AF $\neg$logged_in)



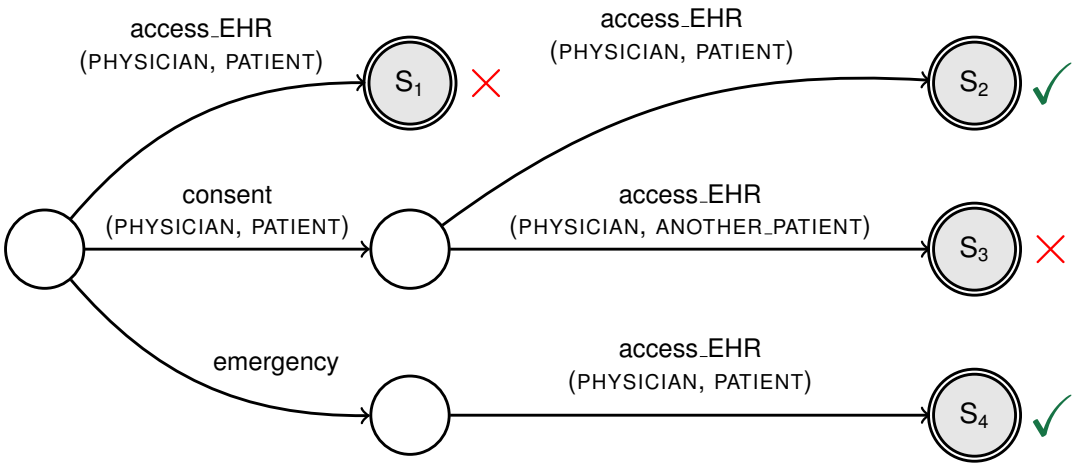
Norm Violations

C(SUBJECT, OBJECT, antecedent, consequent)



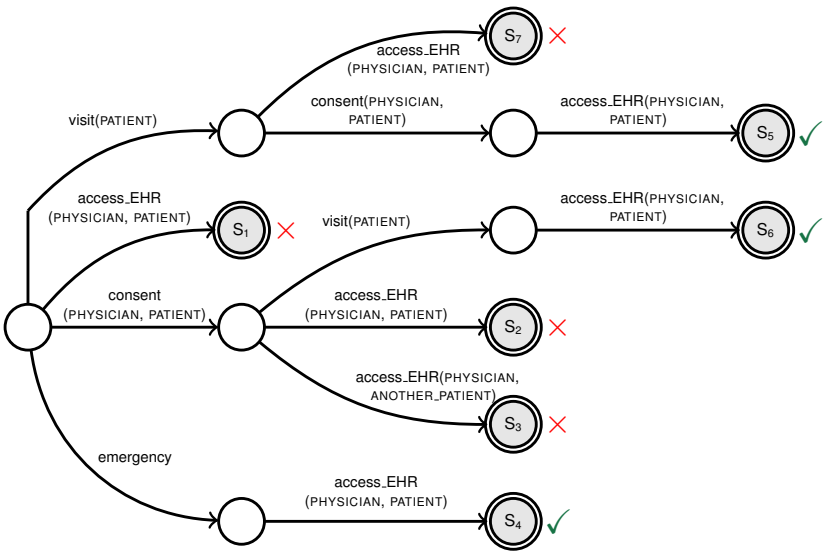
Exercise: Norm Violations I

- P(PHYSICIAN, HOSPITAL, ¬consent(PHYSICIAN, PATIENT) ∧ ¬emergency, access_EHR(PHYSICIAN, PATIENT))



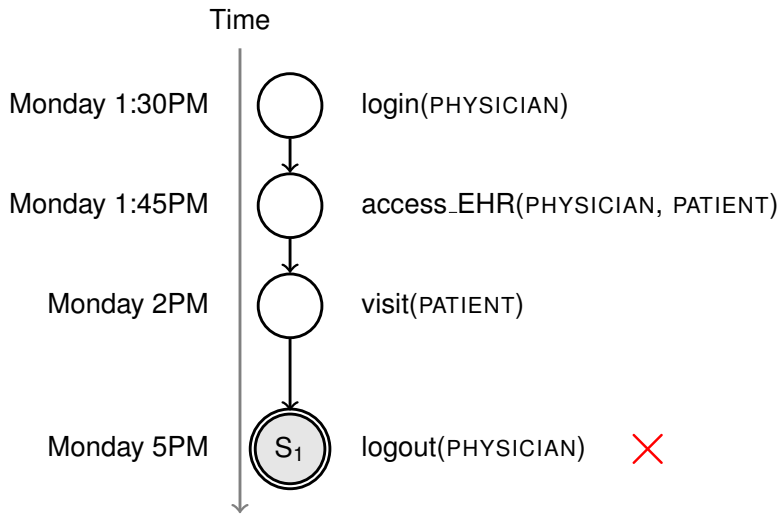
Exercise: Norm Violations II

- P(PHYSICIAN, HOSPITAL, (¬consent(PHYSICIAN, PATIENT) ∨ ¬visit(PATIENT)) ∧¬emergency, access_EHR(PHYSICIAN, PATIENT))



Norm Deadlines

- *C*(PHYSICIAN, HOSPITAL, access_EHR(PHYSICIAN, PATIENT), logout(PHYSICIAN, one_hour))



Exercise: Monitoring Logs

```
% Monday
happens(login(drBob), 8).
happens(access_EHR(drBob, john), 9).
happens(logout(drBob), 10).
happens(give_consent(drBob, john), 16).
happens(give_consent(drBob, kate), 18).
% Tuesday
happens(login(drBob), 32).
happens(access_EHR(drBob, john), 33).
happens(visit(drBob, john), 34).
happens(logout(drBob), 35).
% Wednesday
happens(login(drBob), 56).
happens(access_EHR(drBob, kate), 60).}
happens(logout(drBob), 64).}
```

- misuse(access_EHR(drBob, john), 9) due to no consent
- misuse(access_EHR(drBob, kate), 60) due to no visit
- misuse(logout(drBob), 64) due to no logout

Refinement

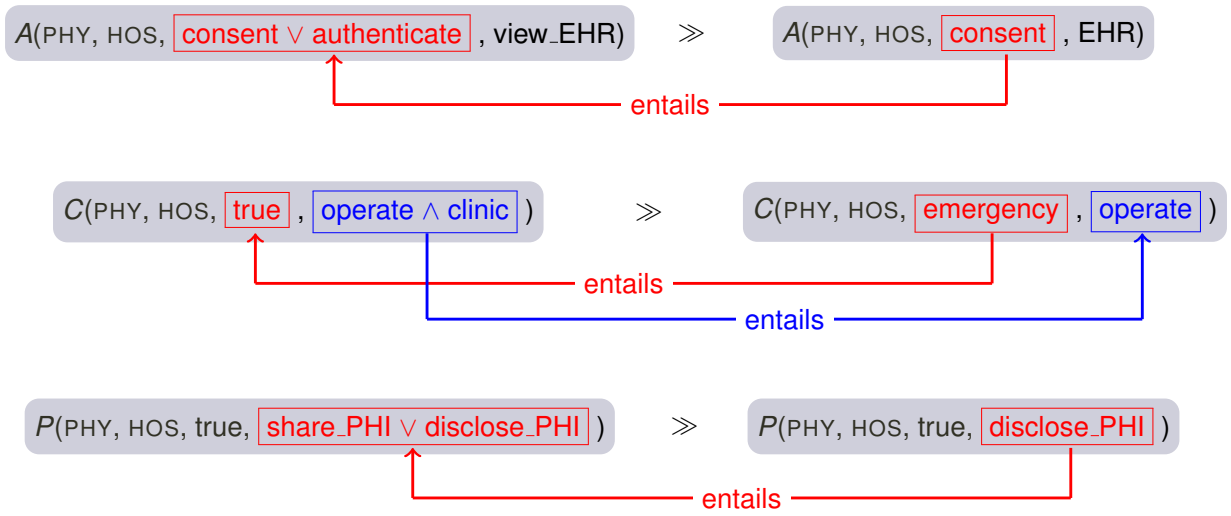
- *Refinement* of a norm: Generalization or specialization of its antecedent or consequent
- An iterative design process to refine norms of an STS specification
  - Takes as input a set of (unsatisfied) requirements
  - Each refinement is captured with a design pattern

Refinement Patterns

- Pattern: A general reusable solution to a commonly occurring problem
- Strengthening: Specify more strict norms
- Weakening: Relax norms
- Amendment: Combine strengthening and weakening
- Overseer: Assign a monitor to a given norm
- Operational: Refine mechanisms
- Sociotechnical: Transform between tiers

https://en.wikipedia.org/wiki/Software_design_pattern

Norm Strength

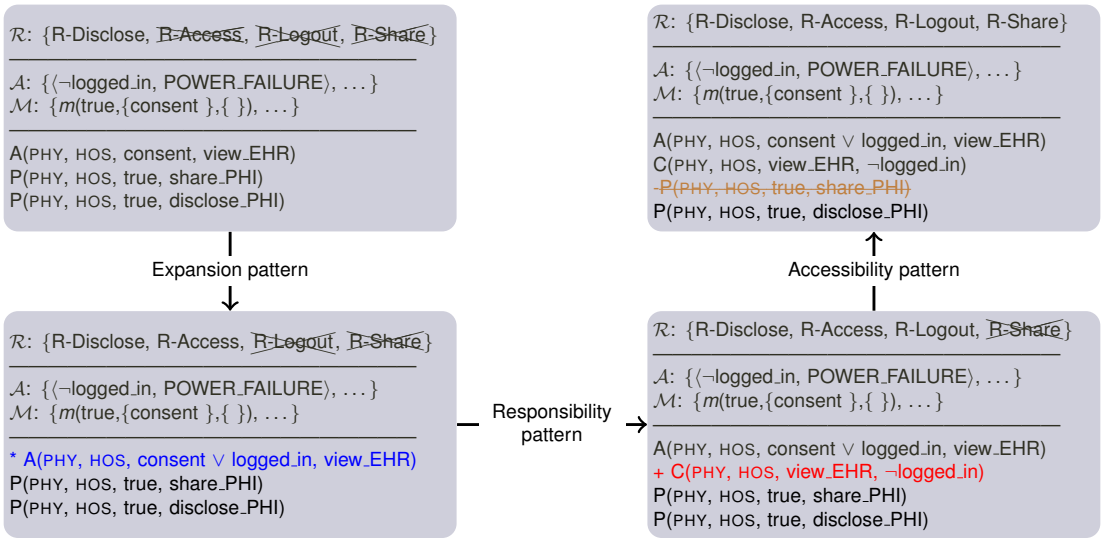


Sample Pattern

- Transform specifications between technical and social tiers
- Relaxing a mechanism may introduce security and privacy risks
- Specify a complementary commitment to mitigate security and privacy concerns
  - Physician is authorized to use PC for 15 minutes before session expires
  - Extend authorization's duration to two hours (technical tier)
  - Physician commits to logging off from computer (social tier)
  - Physician is accountable if commitment violated

Application of Patterns

R-Disclose: AG (¬ disclose_PHI)
R-Access: EF (view_EHR)
R-Logout: AG (view_EHR → AF ¬logged_in)
R-Share: AG (disaster → EF share_PHI)



How Much and When do Patterns Help?

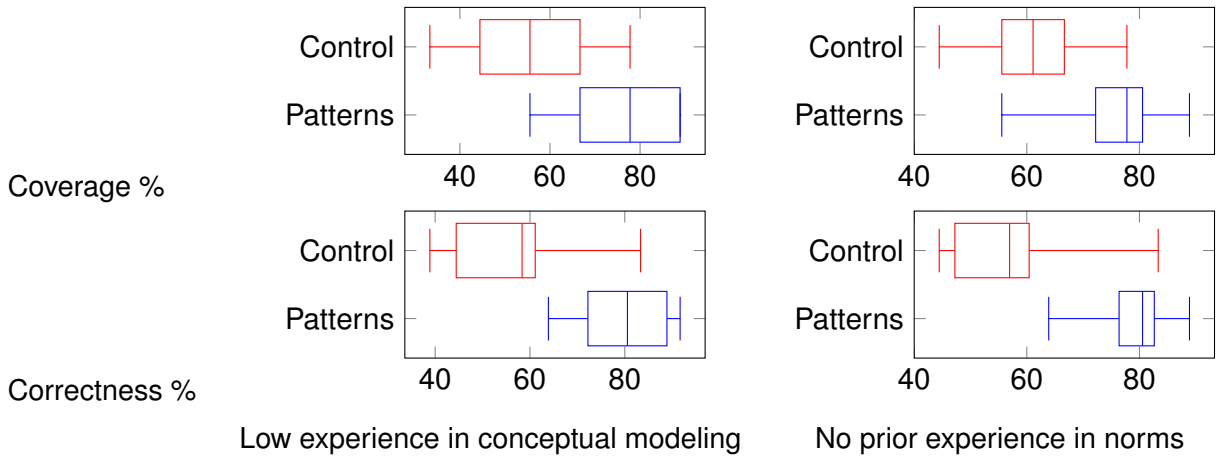
- Questions
  - Do patterns help design better STSs given the requirements?
  - Does prior industry experience or knowledge of norms affect quality of design?

- Preliminary study with 32 participants (computer science graduate students)
  - Control group (no patterns) vs treatment group (patterns), balanced in education and experience
  - After a learning phase, each group designs and refines an STS via norms
  - Three short scenarios on requirements and norms specification

Metrics

- **Coverage** of design: Fraction of norms in the oracle that are stated by the participants in each phase
- **Correctness** of design: Fraction of participant-stated norms that occur in the oracle for each phase
- **Time** to design: Time in minutes recorded by participants to complete each phase
- **Ease** of design: Subjective ratings provided by the participants via a post-study survey (Likert scale, 1–5)

Results



AI Knows Everything

- News article: http://www.thewire.com/technology/2012/07/confirmed-googles-siri-esque-personal-assistant-creepy/54117/
- Links are also on the course website

Things to Look For

- Root cause: What went wrong?
- If it was not intentional, what was the original aim?
- Affected parties
- Implications and similar problems
- Mitigation (using methods we have seen): Prevention, detection, recovery

- Take 10 minutes to look at the incident on your own

- Now discuss with your neighbor
- Also take a look at the summary report: https://drive.google.com/file/d/0B3m-I0YVAv0EcENHYUE3UmN2RTA/view