Assessing the Accuracy of Legal Implementation Readiness Decisions

Aaron K. Massey, Ben Smith, Paul N. Otto, and Annie I. Antón

North Carolina State University

RE 2011—Trento, Italy

2 September, 2011

Problem Statement

Software engineers must be able to build systems that comply with laws and regulations.

This requires the ability to identify which / determine whether requirements have met or exceeded their legal obligations.

Our work seeks to better support software engineers in making these determinations!

Research Goal

Analyze empirical observations for the purpose of characterizing legal implementation readiness with respect to software requirements from the viewpoint of software engineers in the context of an EHR system that must comply with HIPAA regulations.

Outline

- Motivation
- Case Study
- Legal Requirements Metrics
- Analysis Methodology
- Results

Example LIR Requirement

Consider Requirement A:

iTrust shall generate a unique user ID and default password upon account creation by a system administrator.

[Traces to § 164.312(a)(1) and § 164.312(a)(2)(i)]

Relevant HIPAA Section:

(a)(1) **Standard: Access control.** Implement technical policies and procedures for electronic information systems that maintain electronic protected health information to allow access only to those persons or software programs that have been granted access rights as specified in § 164.308(a)(4).

(2) **Implementation specifications:**
(i) **Unique user identification (Required).** Assign a unique name and/or number for identifying and tracking user identity.

Example Non-LIR Requirement

Consider Requirement B:

iTrust shall allow an authenticated user to change their user ID and password so long as it remains unique.

[Traces to §164.312(a)(1) and §164.312(a)(2)(i)]

Relevant HIPAA Section:

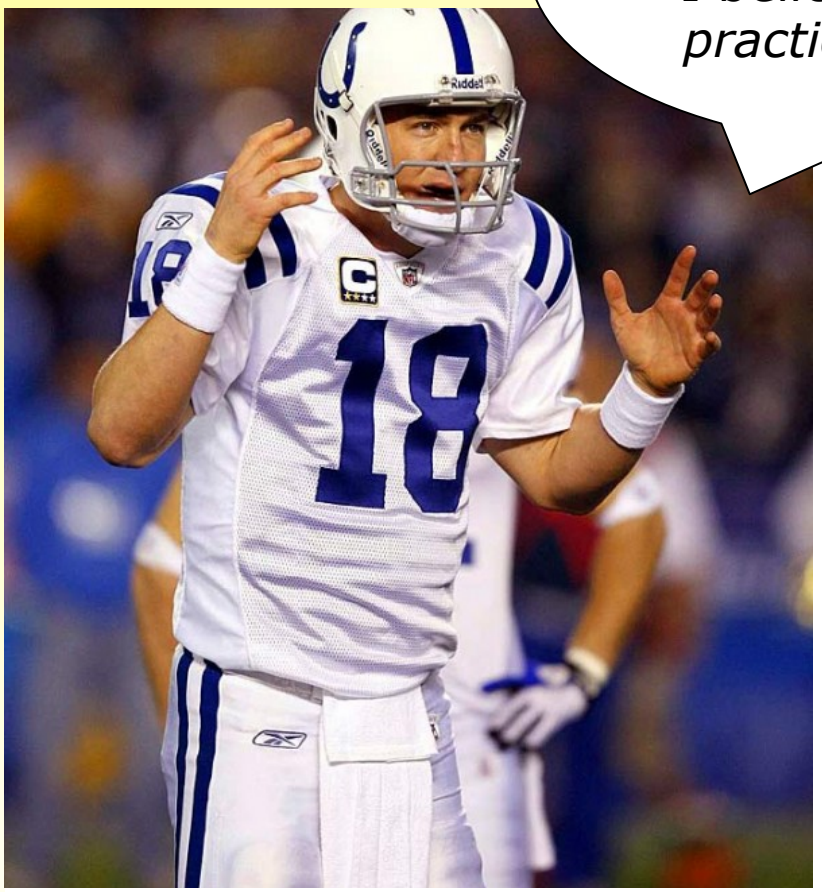
(a)(1) **Standard: Access control.** Implement technical policies and procedures for electronic information systems that maintain electronic protected health information to allow access only to those persons or software programs that have been granted access rights as specified in § 164.308(a)(4).

(2) **Implementation specifications:**
(i) **Unique user identification (Required).** Assign a unique name and/or number for identifying and tracking user identity.

Example Legal Domain: Health Care

- Health Insurance Portability and Accountability Act (HIPAA) passed in 1996
  - Regulates security and privacy for both electronic and paper-based patient information systems
  - $25,000 fines per violation per year for non-criminal violations
- Health Information Technology for Economic and Clinical Health (HITECH) Act passed in 2009:
  - Updated civil and criminal penalties
  - New rules for disclosures of PHI
  - Data breach notification

In layman's



anning

4-time NFL MVP

Had neck surgery Spring and has been hounded by reporters about his recovery.

Reference: http://espn.go.com/blog/afcsouth/post/_/id/27143/mannings-stance-on-hipaa-for-it

Research Questions

1. Is there consensus among:
   a. subject matter experts about which requirements are LIR?
   b. graduate students about which requirements are LIR?
2. Can graduate students accurately assess which requirements are LIR?
3. Can we predict which requirements are LIR using attributes of those requirements?
4. Are the metric categories we have established valid measures of whether a requirement is LIR?
5. Can our legal requirements triage algorithm automate the process of predicting whether a requirement is LIR?

Case Study Design

- 32 graduate student participants over multiple study sessions (21 completed the study in the same room at the same time)
- Three subject matter experts
- Eight legal requirements metrics from three categories
- 31 requirements to analyze
- Session Outline:
  - Discussion of IRB Informed Consent Form
  - Introductory tutorial
  - 45 minutes to complete the study
- Most participants completed the study five or more minutes before the deadline

Case Study Participant Population

- 32 graduate-level software engineering students
  - No prior experience with legal compliance in software engineering
  - All had completed or were, at the time of the study, taking a course on software engineering
    - 150 combined minutes of lectures on requirements engineering
    - 75 minutes of lectures on regulatory and policy compliance

Case Study Materials

- Text of HIPAA §164.312
  - Familiarity *[BA08, MA10a, MA10b, MOH10, MOA09, MA09a, MA09b]*
  - Focuses on Technical Measures of protection
  - Self-contained
- Requirements Specification
  - 31 total requirements
  - Glossary
- Traceability Matrix

Subject Matter Experts

- Three experts in software engineering, relevant laws & regulations
  - Experienced working with HIPAA, legal compliance, and software engineering
  - Three software enginerers (one is also a lawyer)
- Consensus achieved using the Wideband Delphi technique
  - Made individual assessments of the requirements
  - Exchanged our assessments
  - Discussed areas of disagreement to arrive at consensus

Legal Requirements Metrics

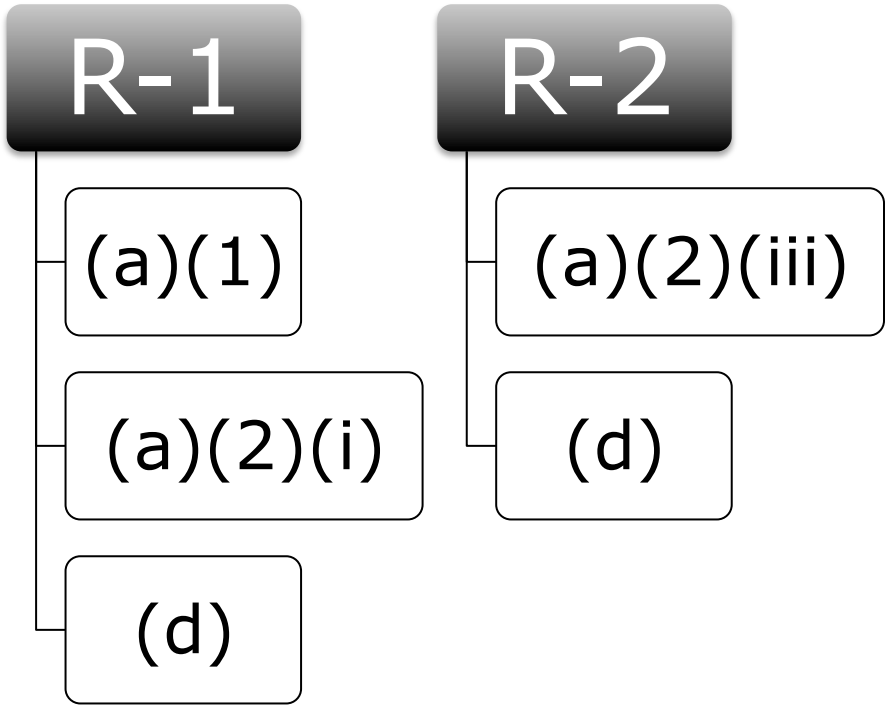
- Assume we have a mapping of req'ts to elements of a legal text *[Bre09, MOH09, MA09, CCG10]*
- Dependency Metrics:
  - *Potential dependencies between the requirements?*
- Complexity Metrics:
  - *Are current requirements too complex for implementation?*
- Maturity Metrics:
  - *Are requirements as simple as possible given the structure of the law?*

Legal Texts are Hierarchical

a) Lorem ipsum dolor sit amet
b) Consectetur adipisicing elit
    1) Sed do eiusmod tempor
    2) Incididunt ut labore et dolore
        i. Magna aliqua
        ii. Ut enim ad minim veniam
    3) Quis nostrud exercitation
    4) Ullamco laboris nisi ut aliquip

Even if we know nothing about the meaning of the law, we can still extract some meaning from the structure.

Requirements Traceability



All HIPAA references are within § 164.312. For example, the reference (a)(1) refers to § 164.312(a)(1).

Legal Requirements Triage Algorithm

- Creates a value for each requirement based on the metrics
- Uses *k-means* clustering to group requirements into three sets:
  1. Legally Implementation Ready
  2. Needing refinement
  3. Non-legal

Statistical Models

- Created a logistic regression model for each data set against the consensus SME responses
- Each logistic regression model used 10-fold cross validation:
  - Partition data into 10 sets
  - Use 9 sets for training, and 1 for prediction
  - Repeat 10 times
  - Average the results for the final prediction model

Results: Consensus among subject experts

- **RQ 1(a):** Is there consensus among subject matter experts on which requirements are LIR?
- **κ** = 0.517 ($p < 0.0001$)
- Result: **Moderate agreement** among the experts about the requirements *prior to the discussion session*.
- Universal agreement on 19 of the 31 requirements

Subject Matter Expert Discussion Session

- Disagreed on 12 requirements
- Some disagreements resolved by one expert pointing out a legal concern that persuaded the other two experts the requirement needed further refinement
- Other disagreements were more complex…

Results: Consensus among participants

- **RQ 1(b):** Is there consensus among participants on which requirements are LIR?
- Result: **Slight agreement** about the requirements.
- **κ** = 0.0792 ($p < 0.0001$)
- Only somewhat better than "agreement" found in perfectly random responses.

Results: Assessment of LIR

- **RQ 2:** Can graduate students accurately assess which requirements are LIR?
- Used **50% as the cutoff for voting** on the status of requirements
- Result: **Students cannot accurately assess the LIR status of a requirement** and are more likely to miss requirements that are not LIR.
- Sensitivity = 0.875, Specificity = 0.2, and κ = 0.076 ($p < 0.0001$)

Results: Using attributes to predict LIR

- **RQ 3:** Can we predict which requirements are LIR using attributes of those requirements?
- Result: The logistic regression model built on our legal requirements metrics exhibited **fair agreement with the expert opinion.**
- Sensitivity = 0.625, Specificity = 0.80, and **κ** = 0.35 ($p < 0.0001$)
- Model is more likely to miss LIR requirements than non-LIR requirements.

The metrics can be useful!

Computer Science
NC STATE UNIVERSITY

the privacyplace.org

Results: Triage Algorithm Categories …

- **RQ 4:** How do the categories for the legal requirements metrics affect whether a given requirement is LIR?

Term	Coefficient Sign
Dependency	Negative
Complexity	Negative
Maturity	Positive

- If the coefficients of a logistisc regression function are ***negative***, then higher values mean the requirement is less likely to be LIR.
- If the coefficients are ***positive***, then higher values mean the requirement is more likely to be LIR.

Results: Triage algorithm vs. experts and participants

- **RQ 5:** Can we use our legal requirements triage algorithm to automate the process of predicting whether a requirement is LIR?
- Result: No, the algorithm did not perform well enough to support software engineers, but it performed better than the students!
- Sensitivity = 0.5, Specificity = 0.466, and κ = -0.03 ($p < 0.0001$)
- Future Work: the algorithm should be replaced by a statistical model.

Lessons Learned

- Software engineering graduate students are ill-prepared to make legal implementation readiness decisions with any confidence.
- Subject matter experts must be involved in legal compliance decisions.
- Legal requirements metrics show potential for quickly evaluating legal compliance for software requirements.

Acknowledgements

- This work was partially supported by **NSF ITR Grant #522931** and **NSF Cyber Trust Grant #0430166**.

Questions