NC STATE UNIVERSITY

Toward Extraction of Security Requirements from Text

Hui Guo[†], Özgür Kafalı[‡], Anne-Liz Jeukeng[§], Laurie Williams[†], and Munindar P. Singh[†]

Goal

Aid security analysts in producing security and privacy requirements that incorporate knowledge from breach reports through the development of an information extraction and analysis framework that combines human intelligent (crowdsourcing) with automated methods.

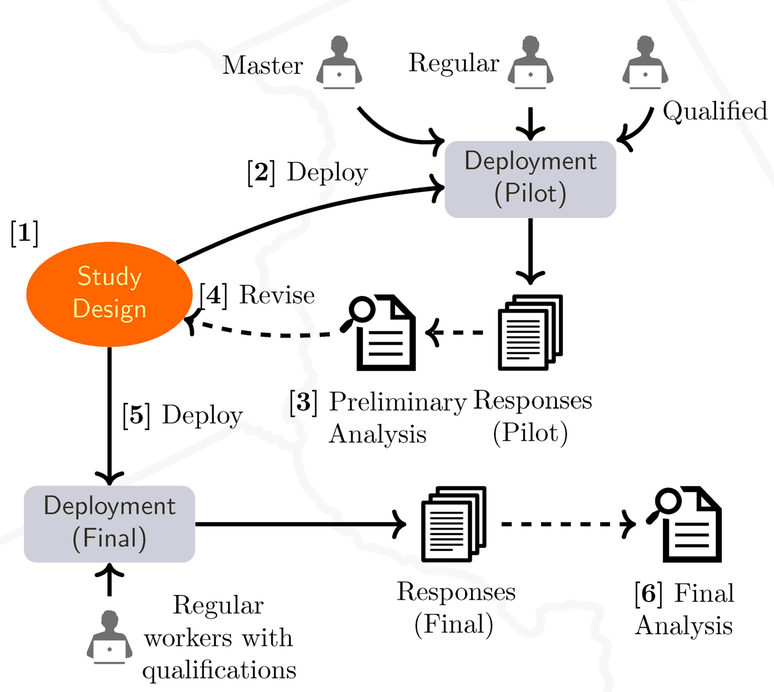
Research Questions

RQ$_1$: Manual extraction. How can crowdsourcing be applied to the extraction of norms from textual artifacts, and what factors affect its performance?

RQ$_2$: Automated extraction. Can automated extraction methods be augmented by crowdsourcing results?

RQ$_3$: Breach reporting. Do breach reports that are more concise and more structured yield higher-quality information extraction?

Study Overview



Contributions

- A crowdsourcing methodology for extracting normative elements from regulations and breach reports and its evaluation (RQ$_1$)
- A curated dataset with evaluated worker responses and preliminary results of automated methods using this dataset as a training set (RQ$_2$)
- Demonstration of the need for concise and structured breach reporting (RQ$_3$)

Future Directions

- Experiments on different strategies to improve worker responses
- Automated tools to preprocess sentences towards simpler and more straightforward tasks
- Heuristics to compose norms automatically from parts of sentences

Breach Report Example

1. An unencrypted portable data drive was lost by a pharmacy resident of the Arnold Palmer Hospital, a part of the covered entity (CE).
2. The drive contained the protected health information (PHI) of 586 individuals, including names, birth weights, gestational age, admission and discharge dates, medical record numbers, and some transfer dates.
3. The missing drive also stored personal items, a research study proposal, and two spreadsheets containing limited information on 586 babies who were part of a study.
4. The CE provided breach notification to HHS, the media, and to the parents of the affected individuals because they were all minors.
5. Substitute notice was posted on the CE's website.
6. The CE updated its policies and procedures for its data loss prevention system and added controls.
7. The CE retrained the resident involved in the loss of data and provided additional information to all employees and medical staff members regarding the use of portable data devices through education and published articles.
8. OCR obtained assurances that the CE implemented the corrective actions listed above.

Crowd Responses

Task	Response
Action	The portable drive should have been better safe-guarded, including using data encryption
Who	The pharmacy resident
Condition	When handling patients' data it should always be encrypted and handled with the utmost concern
Whom	Arnold Palmer Hospital and some its patients who were involved in a study
Action	Train employees about data loss, data protection
Who	the covered entity
Condition	When PHI is involved . . . take the action
Whom	patients, the covered entity, the employee involved

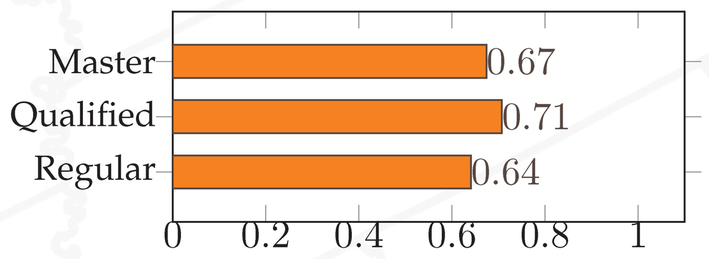
Results

Type: p: Prohibition
Subject: EMPLOYEE (Sent. 1)
Object: COVERED ENTITY (Sent. 1)
Antecedent: portable devices contain PHI (Sent. 1)
Consequent: lose portable devices (Sent. 1)

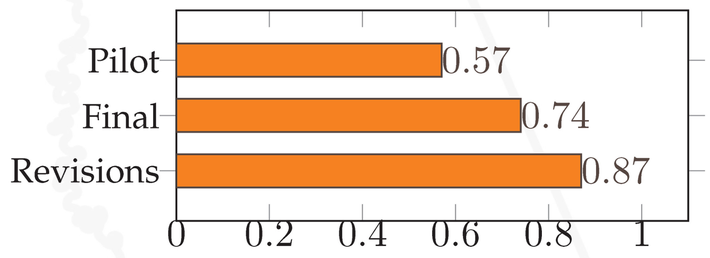
Type: c: Commitment
Subject: COVERED ENTITY (Sent. 7)
Object: PATIENTS (Sentence 2)
Antecedent: TRUE (at all times)
Consequent: train employees on data loss, data protection (Sent. 7)

Response Quality

Average responses quality across worker groups in the pilot:



Average responses quality for a collection of reports in the pilot and in the final, as well as for their revisions in the final:



Automated Extraction

Following the breach, the CE canceled access passwords for patient data, and changed patient data software to a server based system that is password protected and encrypted.

- c(CE, NONE, devices/systems contain PHI, encrypt devices/systems)
- c(CE, NONE, TRUE, implement access control mechanism on system with PHI)
- c(CE, NONE, TRUE, implement safeguard on devices/systems with PHI)

A former employee stole a printout of a patient listing created in January 2015 that was hanging in the locked medical records room and used the information to send letters to several patients.

- p(EMPLOYEE, CE, TRUE, disclose PHI to an unauthorized party)
- c(A THIRD PARTY, CE, PHI is impermissibly disclosed, return or destroy PHI)
- p(EMPLOYEE, CE, documents contain PHI, disclose document to an unauthorized party)

References

[HHS, 2003] HHS. 2003. Summary of the HIPAA privacy rule. (2003). HHS. http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/.

[HHS Breach Portal, 2016] HHS Breach Portal. 2016. Notice to the Secretary of HHS Breach of Unsecured Protected Health Information Affecting 500 or More Individuals. (2016). HHS. https://ocrportal.hhs.gov/ocr/breach/.

5TH ANNUAL
HOT TOPICS *in the* **SCIENCE OF SECURITY**
APRIL 10-11, 2018 | RALEIGH, NORTH CAROLINA