

Principedia

A Database of Privacy Incidents

Dr. Pradeep Murukannaiah and Dr. Jessica Staddon

Privacy Research

- When did awareness of cyberbullying increase?
- How many public privacy incidents involved [insert large Internet company] last year?
- What types of privacy incidents are increasing in frequency according to publicly available data?

How would you go about answering such questions today?

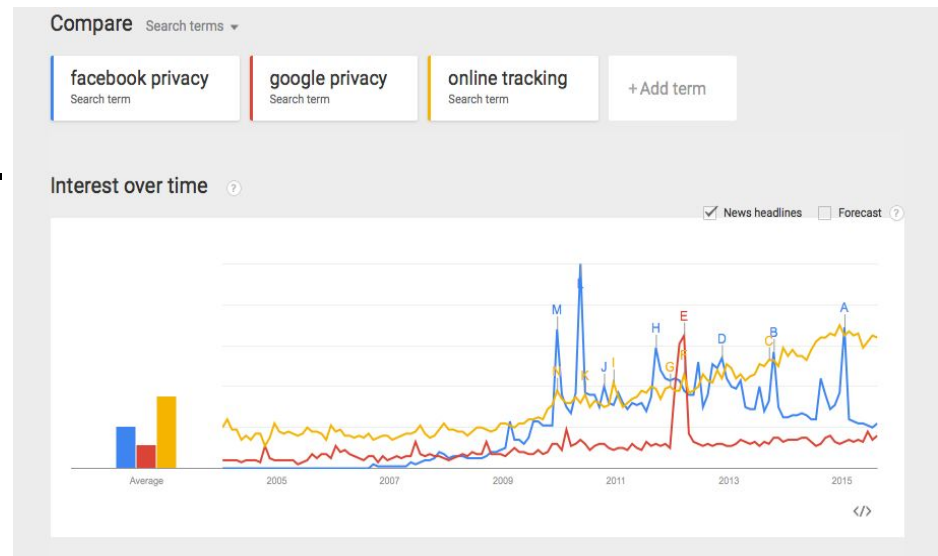
Privacy Research

- How can we answer privacy-related research questions?
 - There is no easy solution!
- Why do we need to answer such questions?
 - Technical solutions
 - Training/education
 - Social research
 - Legal scholarship in privacy

Principedia

Goal: Build the first comprehensive database of privacy incidents

- Crowdsourced, moderated database (Wikipedia style)
- Tagged by entities involved, root cause, etc.
- Automated tracking of analytics
 - Like Google Trends but for privacy incidents



An Example Privacy Incident

Apple fights court order to unlock accused terrorist (one of the San Bernadino shooters) iPhone and provide the FBI access to the unlocked phone.

#02/2016,#Apple,FBI

#surveillance

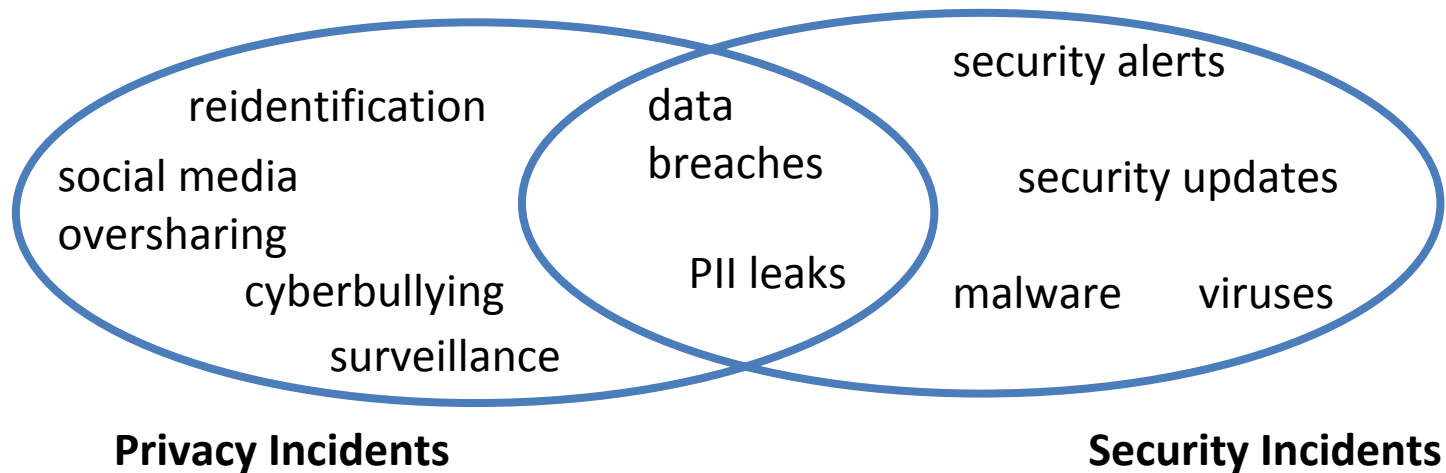
#California #CourtOrder



Image Source: <http://wccfttech.com/>

Privacy vs. Security Incidents

- Security incident databases exist, but **are they sufficient for privacy research?**




Principedia Prototype

<http://go.ncsu.edu/privacyincidents>

go.ncsu.edu/privacyincidents/

Privacy Incidents Database

Incidents Contact Team



Snapchat's claims of the ephemerality of snaps found to be misleading.

#5/2014 #World #SnapChat #UnexpectedProductBehavior #Citizens

Research Questions

- Definition and scope
 - How to define a privacy incident so as to capture all incidents of interest?
- Automation
 - How can we automate the process of populating the privacy incidents database?

DEFINITION AND SCOPE

Defining a Privacy Incident

A **privacy incident** is

- an instance of accidental or unauthorized collection, use or exposure of sensitive information,
- OR
- an event that creates the perception that unauthorized collection, use or exposure of sensitive information may happen,
- AND
- information is either being collected, used or shared in digital form

Testing Our Definition

- Do we have the right definition of an incident?
- What is a sound and intuitive taxonomy for privacy incidents?
- How well does Solove's taxonomy work for classifying incidents?
- Can **Amazon MTurk** be used to generate useful incident descriptions?

A User Study on Amazon MTurk

- Feed hand-curated news articles to AMT as a classification task and ask turkers to assign primary category with privacy being one option
- Ask turkers to assign keywords to incidents that are already in our database
- Present turkers with a taxonomy and ask them to classify incidents in our database accordingly

Definition Testing HIT

Provide a hand-curated set of articles that illustrate our privacy incident definition in several different ways as well as some negative examples

- Is this article about a privacy incident?
 - [radio button: yes, no, I'm not sure]

Solove Categorization Test HIT

Provide a hand-curated set of articles that illustrate our privacy incident definition in several different ways - no negative examples.

- Which of the following best describes the privacy incident reported in this article?
 - [radio button: information collection, information processing, information dissemination, invasion, other]

Solve Subcategorization Test HIT

Continuation of the categorization task...

(Assume that a participant chooses *information collection* as the primary category)

- Which of the following best describes the information collection incident reported in this article?
 - [radio button: surveillance, interrogation, other]

Taxonomy Test HIT

Provide a hand-curated set of articles that illustrate our privacy incident definition in several different ways - no negative examples

- Which of the following best describes the privacy incident reported in this article? Please select as many as apply.
 - [checkbox: personal information leak, cyberbullying, revenge porn, child privacy, targeting, location tracking, data collection, defamation, unexpected sharing, unexpected collection, hipaa, privacy policy, cyberstalking, surveillance, other]

Open-Ended HIT

Provide a hand-curated set of articles that illustrate our privacy incident definition in several different ways - no negative examples

- Does this article describe a privacy incident?
 - [yes, no, I'm not sure]
- If yes, in 1-2 sentences, please describe why it is a privacy incident

Our goal is to gain insights into the attributes that MTurk users think are indicative of an incident. This task may suggest a new taxonomy.

Odd One Out HIT

Provide each Turker a set of 3 news articles illustrate our privacy incident definition in several different ways - no negative examples

- Do you feel that two of these three incidents are more similar to each other than to the third incident? If so, which of these incidents is least like the other two?
 - ☐ Incidents 1 and 2 are similar, incident 3 is different
 - ☐ Incidents 1 and 3 are similar, incident 2 is different
 - ☐ Incidents 2 and 3 are similar, incident 1 is different
 - ☐ All incidents are equally similar in my opinion

Our goal is to gain insights into how turkers cluster incidents. This task may suggest a new taxonomy

AUTOMATION

Recognizing Privacy Incidents

- Privacy incidents may have differentiating features
 - Sentiment, entities, keywords, authors
- Many document will mention privacy, but the term “privacy” itself is not indicative of an incident

Can we train a classifier for efficient, large-scale identification of privacy events from a variety of inputs such as news articles, blog posts and social media?

Privacy Incident Classifier

Semi-automated collection of privacy incidents

- Crawl articles from various sources
 - News APIs, RSS feeds, tweets from key privacy-related persons, and so on
- Classify articles as **privacy incident** or **not privacy incident**
- Verify (expert) privacy incidents and add to the database
- Train the classifier on weekly basis

Training a Classifier

- **Positive Articles:**
 - Articles currently present in the database
- **Negative Articles:**
 - Articles gathered randomly from NY Times
 - Articles collected with security keyword from NY Times
 - Articles mentioning the term “privacy,” but are not privacy incidents

Data for Classification

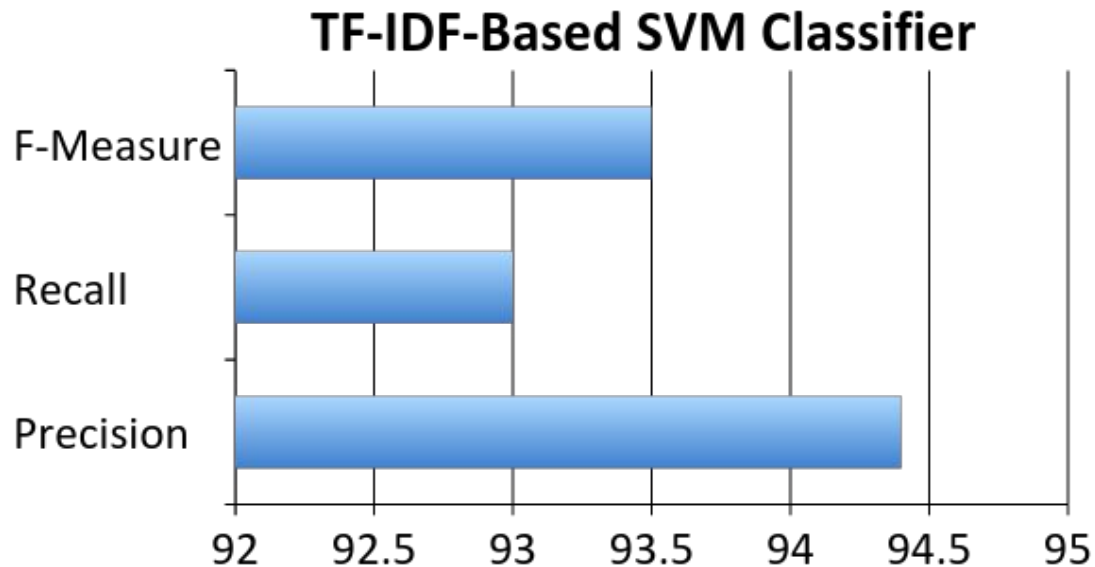
- **Textual content**
 - Text scraped using the news url (with html tags removed using Python BeautifulSoup library)
- **Metadata**
 - Type of the news material, newsdesk, author, word count, location, date, and so on

Natural Language Processing Steps

- **POS Tagging:**
Retain nouns, verbs, adverbs, or adjectives
- **Stemming:**
Employ Porter Stemmer to stem the words
- **TF-IDF Computation:**
Calculate the term frequency and inverse document frequency for each stemmed word for every document

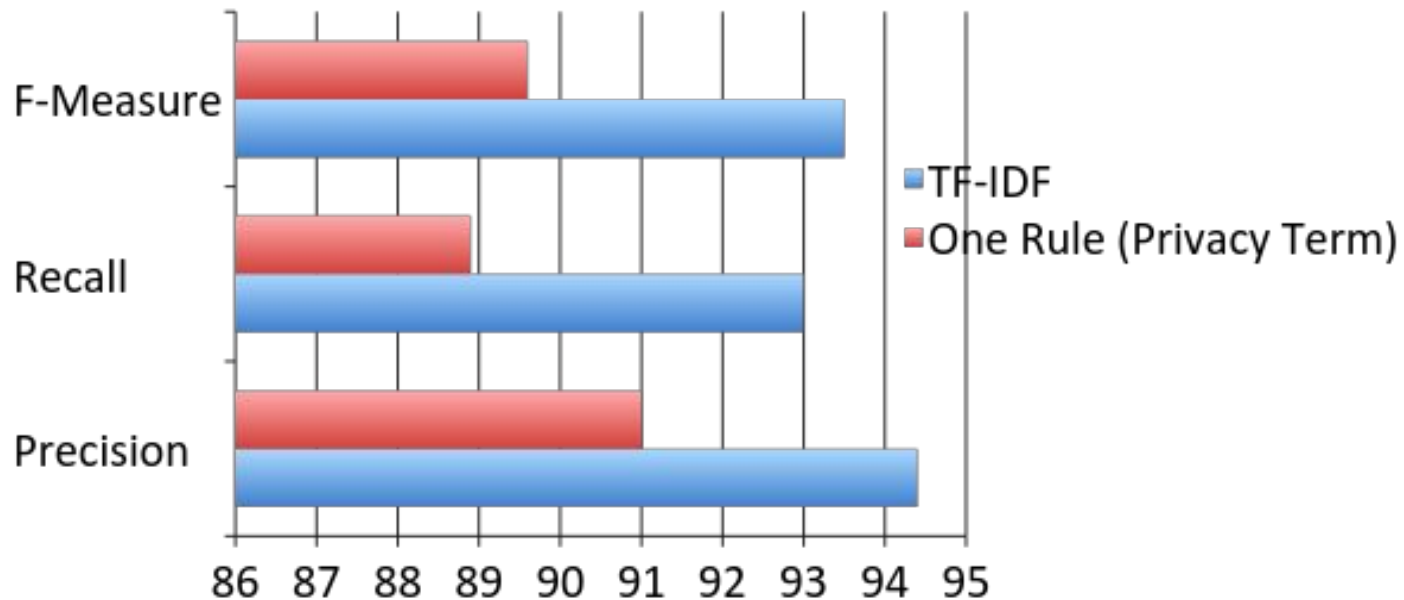
Preliminary Results

- Privacy incident vs. Not privacy incident classification



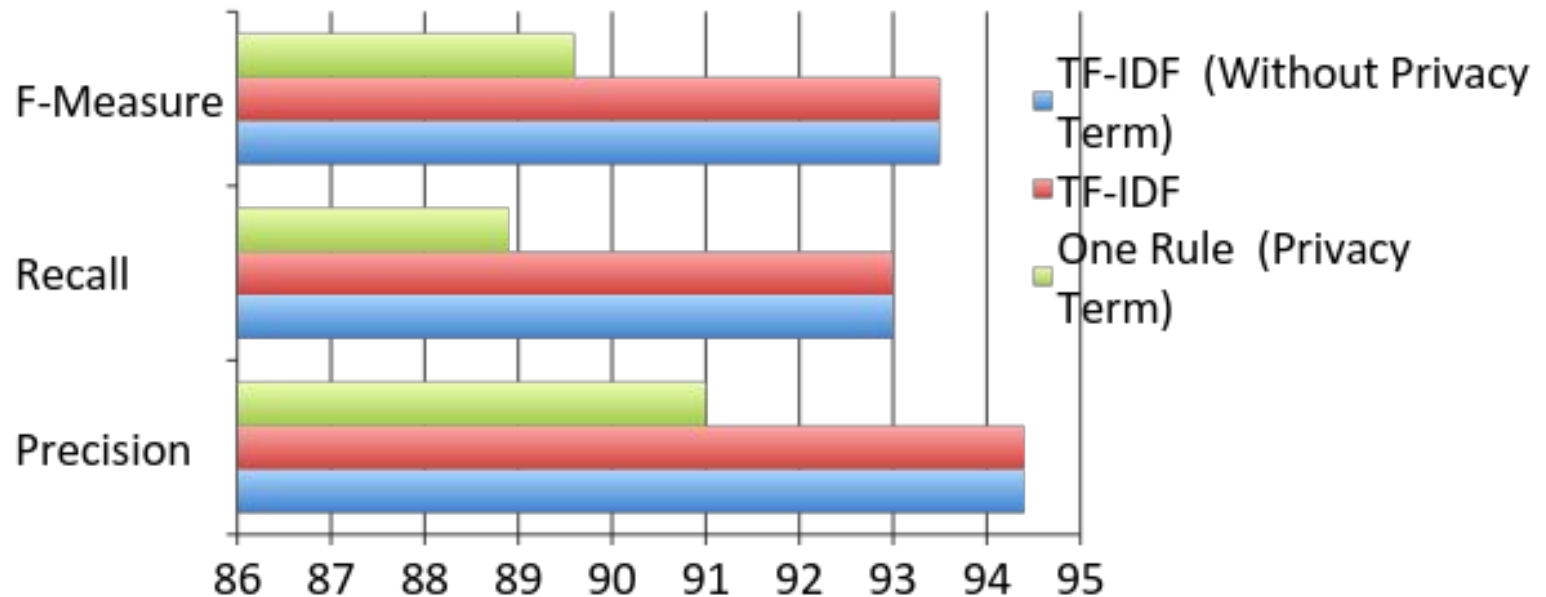
Preliminary Results

- Privacy incident vs. not privacy incident classification



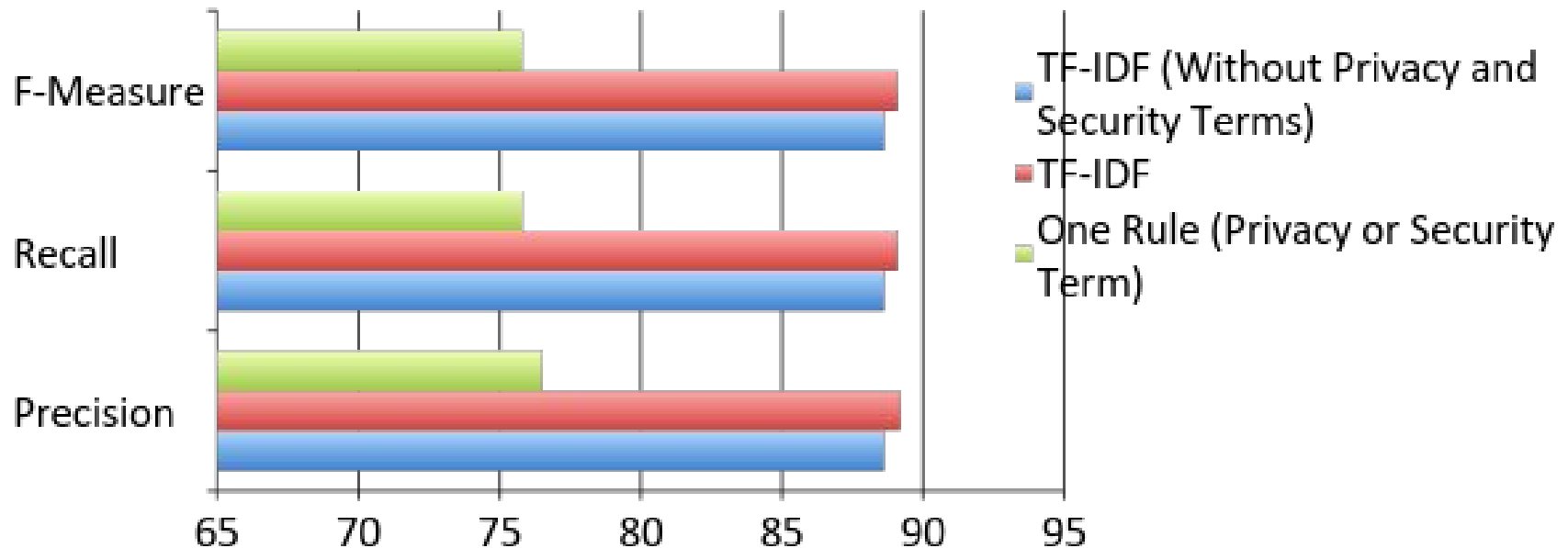
Preliminary Results

- Privacy incident vs. not privacy incident classification



Preliminary Results

- Privacy incident vs. security incident classification



Incident Classification: Directions

- Curating better training and test sets
 - Include documents mentioning privacy, but are not privacy incidents
- Multilabel classification
 - Documents involving both privacy and security incidents
- Entity recognition
 - Incident type, parties involved, perpetrator, victims, etc.
- Employ the **crowd** or **experts** for manual verification depending on the confidence of the classifier's predictions

Collaborators

- Sarvesh Rangnekar, Kaustubh Sant, Shiqian Xu, and Yuxu Yang, NCSU
- Dr. Heather Richter Lipford, UNC Charlotte
- Dr. Bart Knijnenburg, Clemson University

We are looking for more collaborators!

THANK YOU