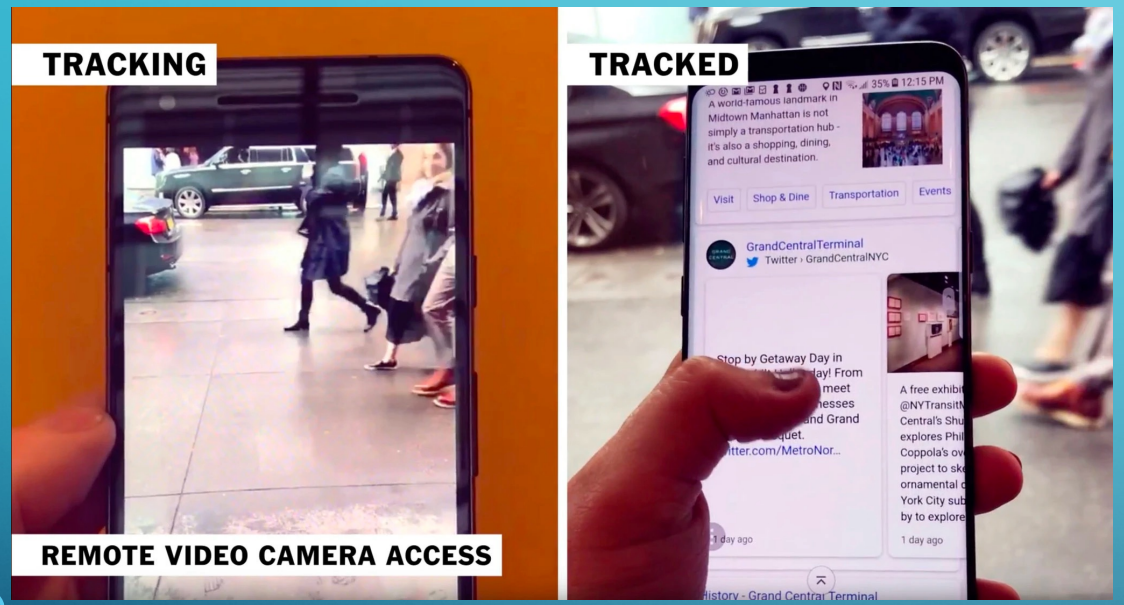
TOWARDS STALKERWARE DETECTION WITH PRECISE WARNINGS

YUFEI HAN

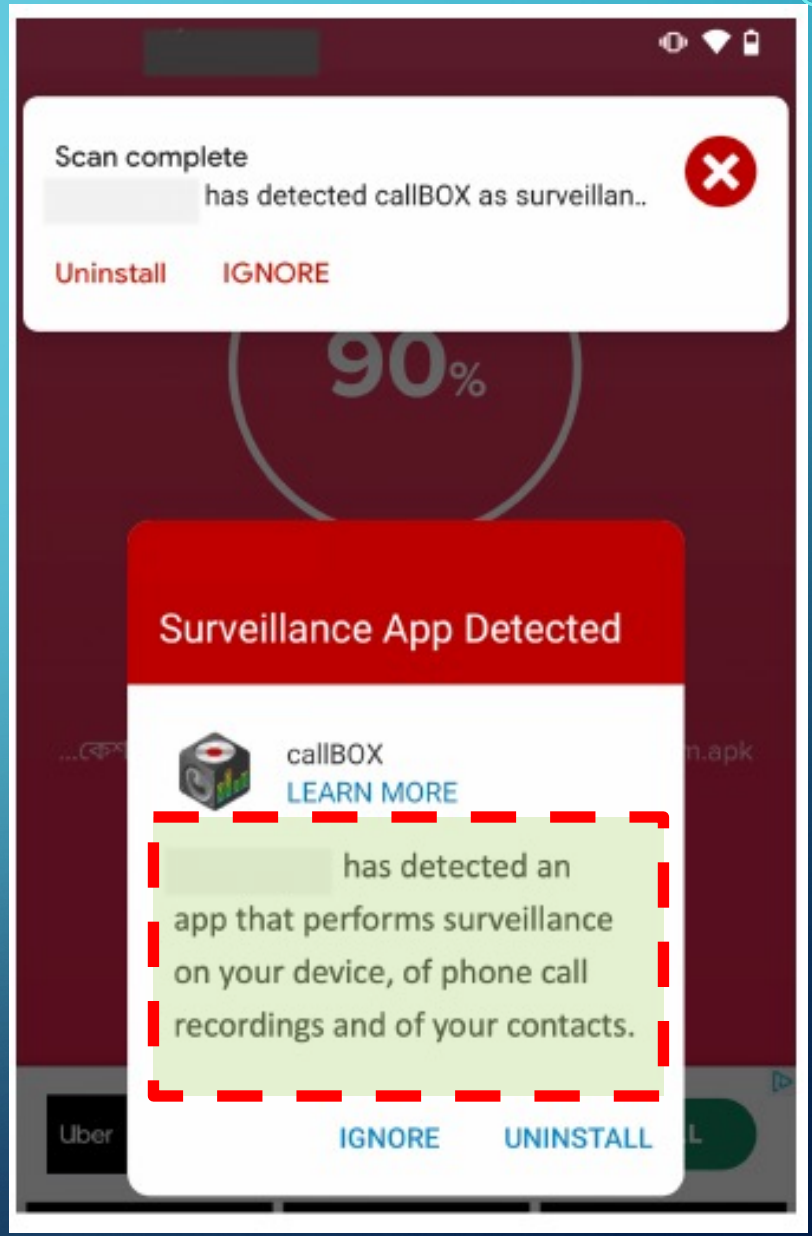
KEVIN A. ROUNDY & ACAR TAMERSOY

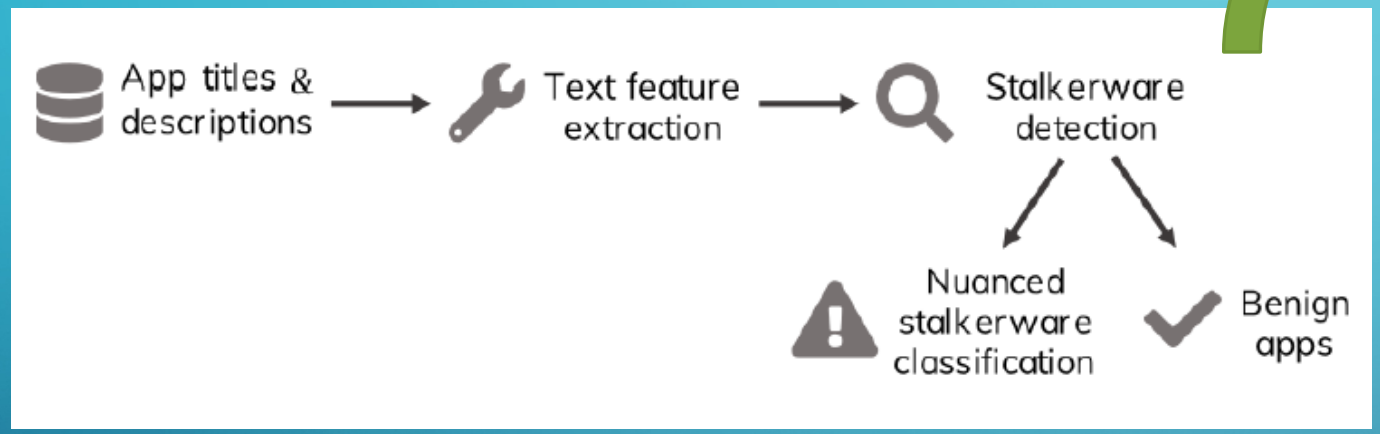
WHAT IS STALKERWARE?



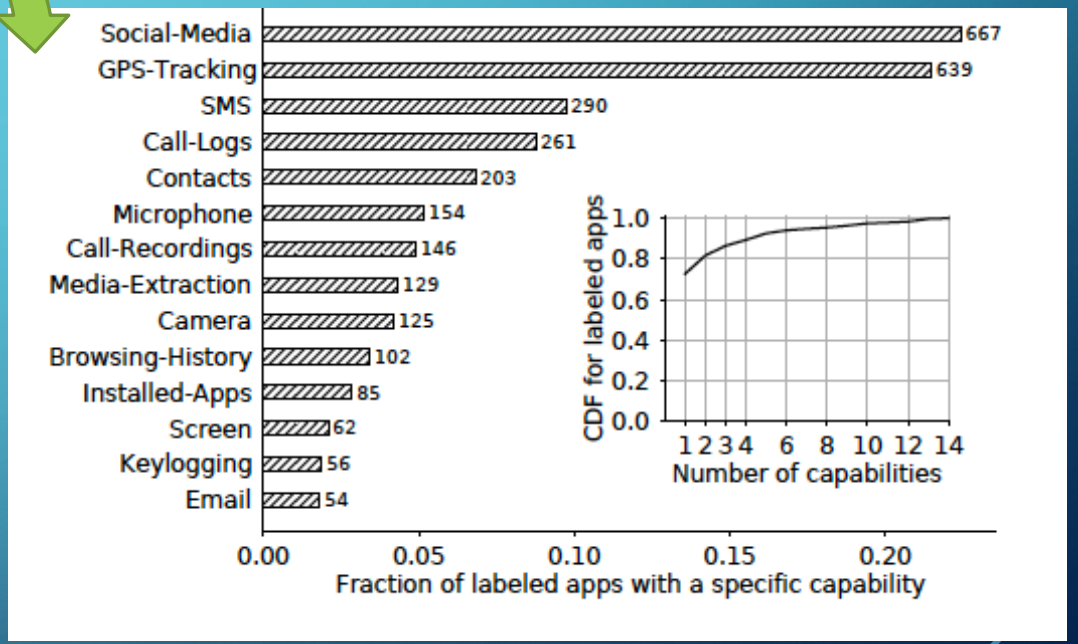
- Impersonation attacks and providing surveillance capabilities.
  - Location / GPS tracking
  - Keylogging, Call logs copying
  - Phone call recording / SMS copying
  - Remote control of victims' mobile devices
- These apps can be used to abuse, stalk, harass and spy on others, frequently to enable intimate partner violence.
  - In Florida, a man killed his wife and two children several days after he installed surveillance software on her phone and used it to see her texts and photos, https://www.nytimes.com/2018/05/19/technology/phone-apps-stalking.html

- **DOSMELT: Don't Scare Me Like That**
  - **A system providing nuanced detection of stalkerware apps**
  - Goal is to provide warnings that are as informative as possible to enable intelligent reactions on the part of the victims
  - Uses self-descriptions of stalkerware apps. Empirical results confirmed that self-descriptions are informative enough for accurate detection





Steps taken by DOSMELT to perform nuanced classification of stalkerware.

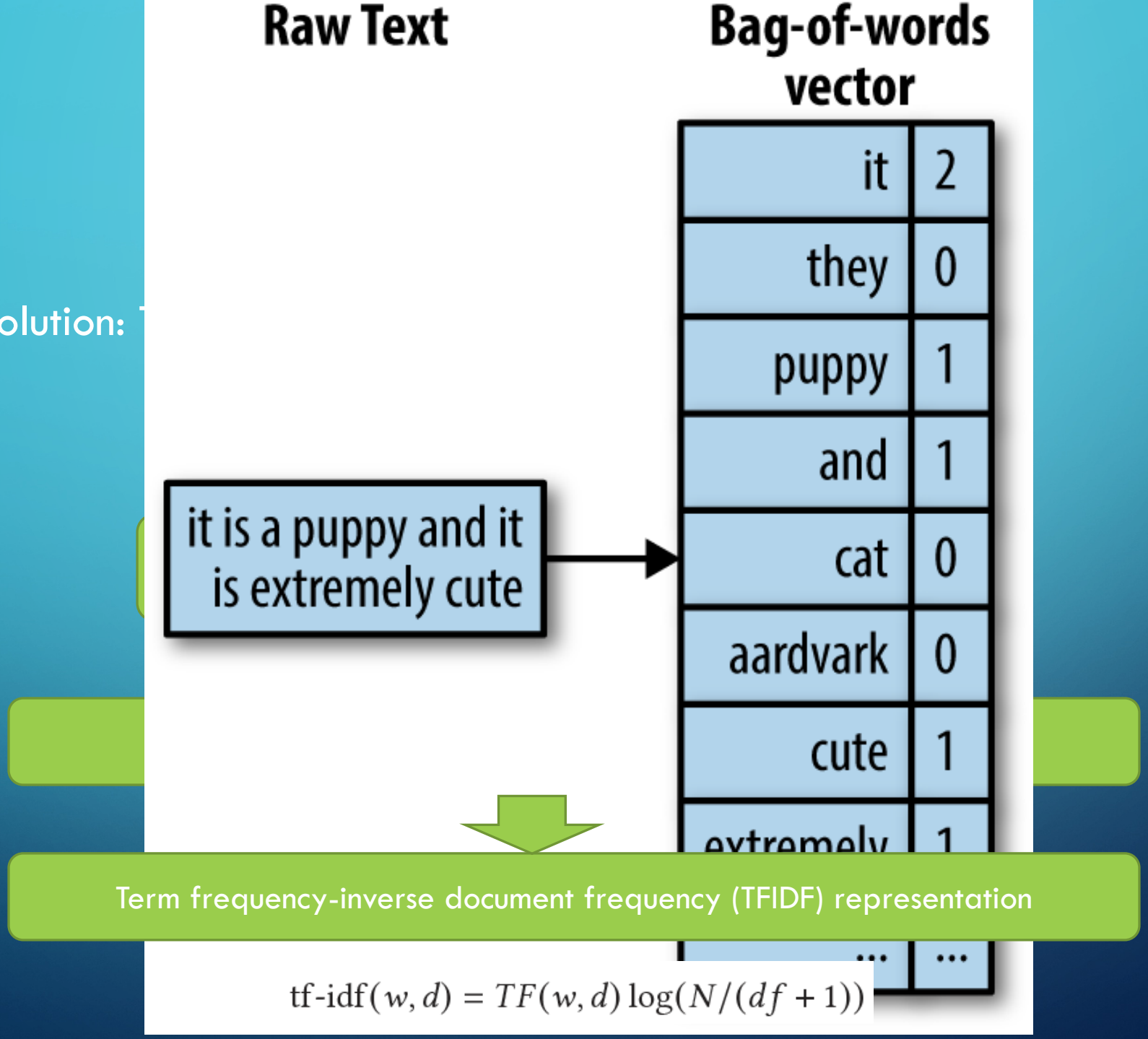


- Existing challenges
  - **Dynamic taint tracking remains useful but insufficient**. Stalkerware apps are likely to require an initial configuration step as the app needs to know to whom the extracted data should be forwarded. Thus, dynamic analysis is far less likely to succeed on stalkerware.
    - Even when dynamic taint tracking does detect information leakage, we still cannot classify the app as stalkerware, because it could be generic spyware instead or a legitimate app that collects data from the device for benign purposes.
      - A spyware app with a privacy-invasive ad library that collects location data either for use in targeted ads or to aggregate it and sell the data to third parties.
      - A genuine stalkerware app that tracks location and shares it covertly with third parties.

- Existing challenges
  - Previous solutions proposed to use keyword-based searches like “catch my cheating girlfriend” on the web and app marketplaces to identify candidate stalkerware apps. **However they do not address the problem of detecting how stalkerware compromises the privacy of an individual to enable the creation of more informative warnings**

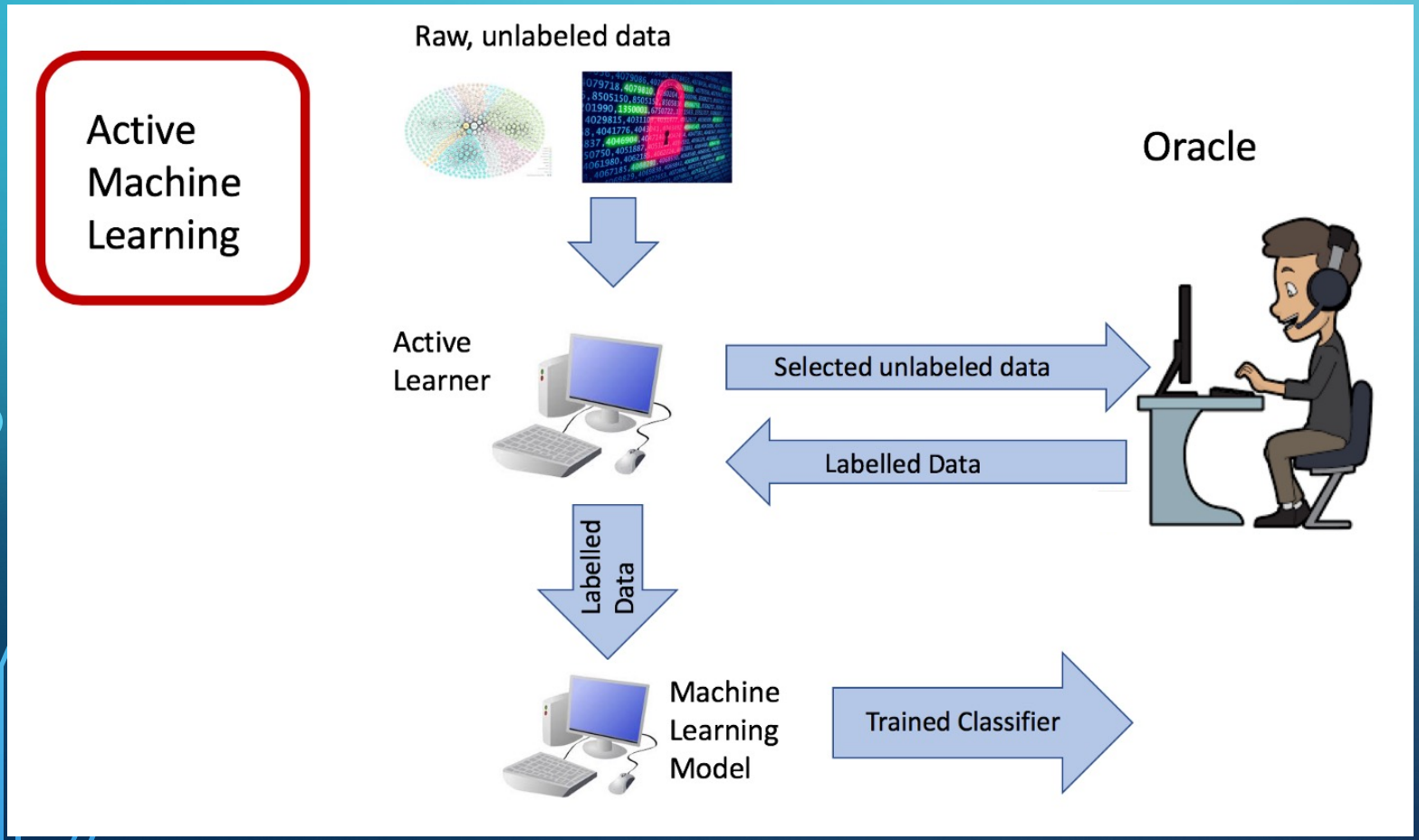
- Existing challenges
  - **Manually coding stalkerware apps costs (tagging surveillance functions) is expensive**
    - No stalkerware datasets exist that identify individual stalkerware capabilities
    - Stalkerwares are sufficiently rare that randomly sampling Android apps would be unlikely to turn up any meaningful number of stalkerware apps to label

- Our solution:



Remove common stopwords with the NLTK package

- Our solution: Learn-by-predict based active learning paradigm

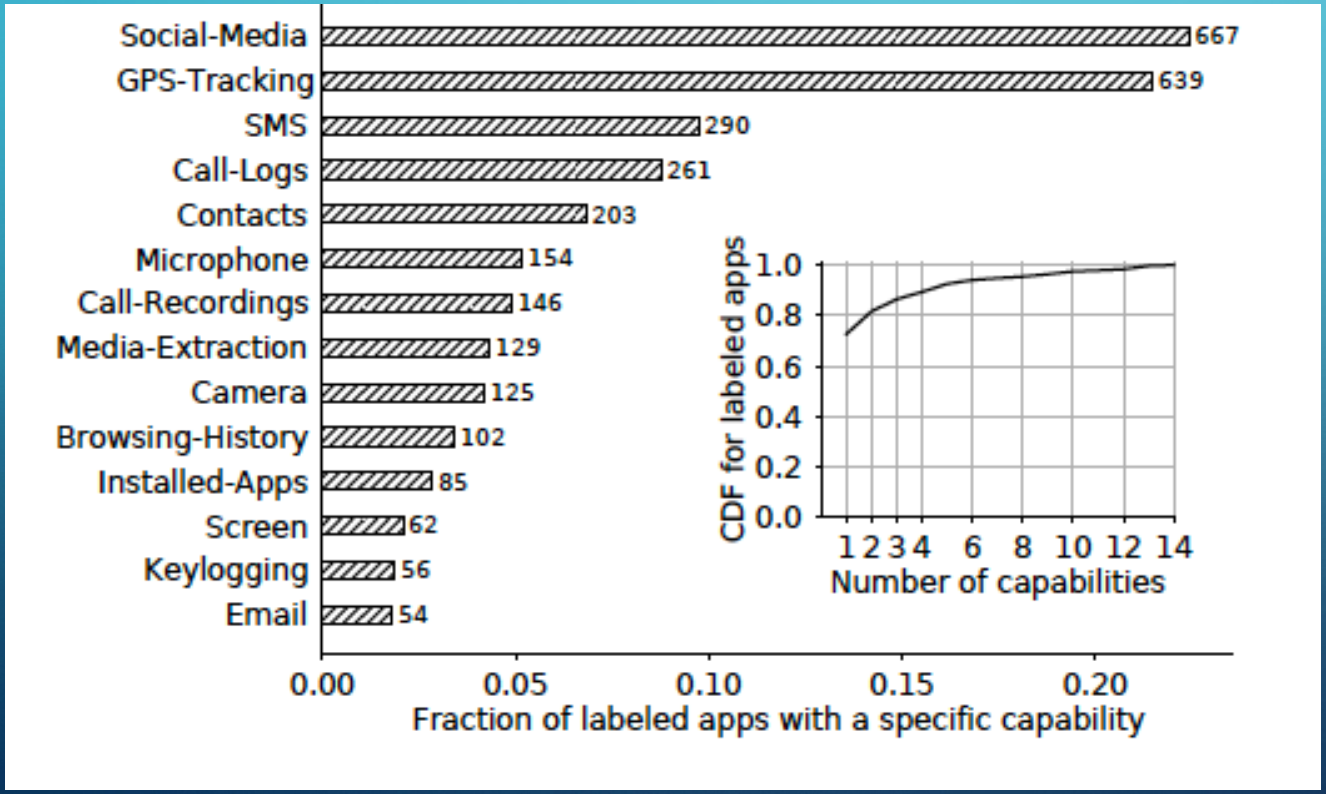


Step.1 We allow DOSMELT to access only a small amount of labeled stalkerware and non-stalkerware apps. These examples are used to initialize the classifier.

Step.2 In each round, DOSMELT selects the unlabelled instances with confidence score larger than 0.9 as the next apps to be labelled by human analysts, which are added to the training set in the subsequent round.

$$x^* = \arg\max_{x_i} P_\theta(\hat{y}_i = 1|x_i)\ell_f(x_i, y_i)$$

- Our solution: A multi-label classification system to identify surveillance functions hosted by stalkerware apps



Over 73% of stalkerware apps support at least one surveillance capability, there are many apps with multiple capabilities (**a multi-label problem**).

DOSMELT can report multiple surveillance capabilities hosted by one stalkerware app at the same time (**a multi-label classification system**).

- Experimental data collection
- We partnered with Nortonlifelock and obtained two anonymized reports that solely consisted of Android application identifiers (app id for short) between 2019 and 2020. Using the vendor's data to indicate apps that were in use on mobile devices, we then queried APKPure and Google Play Store with the app ids we possessed.
- There are two researchers from our team to manually label the examples in three rounds:
  - In each of the first two rounds, they independently coded 150 randomly chosen apps. Specifically, for each stalkerware capability, they noted whether or not the app under review possessed the capability, based on an examination of **the app title and app description.**
  - In the third and final round, both researchers coded another random set of 100 apps to test agreement. Both researchers achieved a Krippendorff's alpha of 0.86 in this final round, which indicates strong inter-rater reliability.
  - In total, there are this process led to the labeling of **4,839 apps**, where **1,462 of them are stalkerware apps with at least one surveillance capability** from the taxonomy of surveillance capabilities. The rest are benign apps.

- Resul
- Expe

Round	Accura RF	G
0	0.763	0
1	0.900	0
2	0.960	0
Baseline	0.975	0

Table 2: Stalker
rounds in Dosmi
Boosted Trees (

Keyword	Counts Stalk.	Non-stalk.	Keyword	Counts Stalk.	Non-stalk.
keymonitor	12	0	intercept	5	1
whatweb	11	0	transcript	5	1
biometrics	7	0	memos	2	1
calendar	30	8	infrared	2	1
clone	20	3	keystrokes	2	1
database restore	6	0	hider	2	1
whatsmessage	6	0	dialing	12	5
phonefind	6	0	chatting	24	13
spy	6	0	gps	128	70
espiar	5	0	message	207	157
read_phone_state	5	0	locate	39	37
stealer	3	0	email	6	4
monitored display	2	0			

Table 6: Top 25 ranked keywords for stalkerware detection derived using Random Forest (*RF*).

Number of labeled instances
409
615
830
All labeled instances

lity classification re-
sing Random Forest
(*Extra-Tree*).

- **Achievement**: DOSMELT is the first system to detect individual stalkerware surveillance capabilities. It can be a good complement to existing static and dynamic analysis techniques and to reputation systems for detecting stalkerware.
- **Improvement:**
  - DOSMELT works best for apps that have at least at one point been hosted on the Google Play app store. Obtaining descriptions for stalkerware that hosts its self-descriptions on its own website is very difficult to automate.
  - Adversarial attacks on the self-description texts.
  - Stalkerware detection methods applied to non-English apps.

Please reach me out via yufei.han@inria.fr