overview

- framing privacy and security decisions
- privacy and IoT
- privacy on mobile devices
FRAMING PRIVACY AND SECURITY DECISIONS
nudges in the literature

- password meters
- social framing
- “strengthening”
- quantifying password weakness
Webmail

The password that you chose on the previous page is weak, please choose a stronger one below.

Worker ID:

Password: *

Confirm Password: *

Change Password
Password: *

This password is weaker than 86% of other users.
Tip: you can increase the strength of your password by adding random symbols or numbers to it. For instance, changing "foobar" to "foo%bar3" would dramatically increase its strength.
crack time

Password: *

It would take a hacker approximately 7 years to guess your password.
• consideration for future consequences (CFC)
• need for cognition (NFC)
• general decision-making style (GDMS)
  • intuitive
  • avoidant
  • dependent
  • rational
  • spontaneous
traits *do* predict nudge susceptibility:

- low numeracy participants created *weaker* passwords in response to quantitative feedback
- rational and high-NFC participants responded well to quantitative feedback
- dependent decision-makers created *weaker* passwords in social condition
validation

- three conditions (n=923):
  - control
  - random (password meter or crack time)
  - optimal (targeted to psychometrics)
$F(2, 920) = 5.201, \ p = 0.006$

$N = 239, 460, 224$, respectively

Planned contrast optimal vs. control+random showed the mean increase was by about 10% or $0.2 \text{ SDs}$ (Cohen's $d$); $p = 0.001$
the effect concentrated on the crack-time nudge, in which not optimizing resulted in worse passwords
PRIVACY AND IOT
privacy failures

systems that regulate data based on <data type> and <recipient> are bound to fail.
privacy as **contextual integrity**

privacy violations occur when contextual norms are violated:

- I expect medical data to be used by my doctor for treatment
- I do not expect medical data to be used for marketing purposes
applying CI in practice

inferring context is hard

• We cannot know exactly how data will be used

proxies can help

• knowledge of recipient
• data properties (e.g., source, permissions, etc.)
• what the app does
• what else was happening on the device
smart TV study

large survey on data collection and sharing – and protections
  • Exploring differences by data type/format and by recipient

wide variation in assumptions about data collection and flows
  • Most people are against data being repurposed
  • …but assume it will happen regardless!

people believe legal protections exist to prevent egregious violations
  • (they don’t)

study was cited in lawsuit against Vizio for selling customers’ viewing data without their knowledge or consent.

study of current users

When do users expect to be recorded?

What do they expect will happen to that data?

How often do current devices record inappropriately?
History
History shows your voice interactions with Alexa. Tap a line to see details, hear recordings, provide feedback, or delete recordings. Learn more.

alexawhat time is it
Today at 12:45 PM on Bedroom

alexaset volume six
Today at 12:45 PM on Bedroom

what time is it
Today at 12:45 PM on Bedroom

alexa
Today at 12:45 PM on Bedroom

stop
Yesterday at 10:26 PM on Bedroom

alexa
Yesterday at 10:26 PM on Bedroom

alexa set volume four
Yesterday at 9:59 PM on Bedroom

next
Yesterday at 9:56 PM on Bedroom

alexa
Yesterday at 9:56 PM on Bedroom

alexa next
Yesterday at 9:46 PM on Bedroom

next
Yesterday at 9:41 PM on Bedroom
Yesterday

- Assistant
  Said turn this off
  6:16 PM • Details

- Assistant
  Said what's the best time to go trick or treating
  6:03 PM • Details

- Assistant
  Said turn up the volume
  6:03 PM • Details

- Assistant
  Said what time is the sunset
  5:57 PM • Details
1. Who is this a recording of?

- This is a recording of me.
- This is a recording of someone else in my household.
- This is a recording of a guest.
- This is a recording of the TV, music, or other pre-recorded audio.
- This is a recording of noise/gibberish.
2. Did [person speaking] address the device, or was this recording an accident?
   - [ ] [person] was speaking to Alexa.
   - [ ] It was an accident.

3. Do you remember making this request?
   - [ ] Yes
   - [ ] No
   - [ ] I'm not sure
PRIVACY ON MOBILE DEVICES
tools for observing data flows

**Instrumented Android**: Access to sensitive resources (e.g., location, call logs, network state, various identifiers, etc.)

**Haystack**: Network traffic, remote servers, HTTP/HTTPS payloads
automatic behavior detection

what data goes to which parties?

is location data collected?

what persistent identifiers are collected?
  • are they shared across apps?
observing app data flows

network flows are dynamically generated by apps at runtime

apps need to be run so that as much code gets executed as possible
generating data

solution 1: cheap labor

solution 2: Exerciser Monkey!
things we’re detecting

legal violations

incompetent uses of PII

deceptive practices
detecting COPPA violations

Apps targeted at children under 13 must:

- attain parental consent when collecting personal identifiers
- allow parents to limit data collection
- list all data recipients
- transfer data securely
- not use behavioral targeting
- post privacy policies
[potential] violations abound

7,076 *Family* apps tested

- 65% (4,575) transmit identifiers
- 56% (3,950) transmit hardware-based identifiers
  - IMEI, WiFi MAC, Android ID, GSF ID, or serial number
  - 74% (2,929) do so alongside the Android Ad ID (AAID)
- 8% (569) transmit PII
  - Phone number, email address, name, and/or location
parental consent

none of these are using **effective** age gating
(the monkey was able to click through)

in several cases, **identifiers are sent before the consent screen is shown**
Irwin Reyes*, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez, and Serge Egelman

“Won’t Somebody Think of the Children?”
Examining COPPA Compliance at Scale

Abstract: We present a scalable dynamic analysis framework that allows for the automatic evaluation of the privacy behaviors of Android apps. We use our system to analyze mobile apps’ compliance with the Children’s Online Privacy Protection Act (COPPA), one of the few stringent privacy laws in the U.S. Based on our automated analysis of 5,855 of the most popular free children’s apps, we found that a majority are potentially in violation of COPPA, mainly due to their use of third-party SDKs. While many of these SDKs offer configuration options to respect COPPA by disabling tracking and behavioral advertising, our data suggest that a majority of apps either do not make use of these options or incorrectly propagate them across mediation SDKs. Worse, we observed that 19% of children’s apps collect identifiers or other personally identifiable information (PII) via SDKs whose terms of service outright prohibit

1 Introduction

In the United States, there are few comprehensive privacy regulations. However, one notable exception is the Children’s Online Privacy Protection Act (COPPA), which regulates how mobile apps, games and websites are allowed to collect and process personal information from children under the age of 13 [22]. COPPA outright prohibits certain data collection practices, and requires parental consent for others. Of course, enforcement is a painstaking process, as investigations generally rely on manual examination of programs and websites to observe violations [83]. In this paper, we apply our Android dynamic analysis framework to automate the process of detecting potential COPPA violations.

Most current approaches to detecting suspicious applications activity on mobile platforms are static,
Lawsuit Claims Disney Is Violating COPPA, Tracking Kids in 42 Apps

*Disney believes the class action lawsuit “is based on a fundamental misunderstanding of COPPA principles.”*

By Matthew Humphries  August 10, 2017 7:00AM EST
How Game Apps That Captivate Kids Have Been Collecting Their Data

A lawsuit by New Mexico's attorney general accuses a popular app maker, as well as online ad businesses run by Google and Twitter, of violating children's privacy law.

By JENNIFER VALENTINO-DeVRIES, NATASHA SINGER, AARON KROLIK and MICHAEL H. KELLER  SEP'T. 12, 2018
Serge Egelman, a researcher with the International Computer Science Institute and the University of California, Berkeley, helped lead the study of nearly 6,000 children's Android apps. Jim Wilson/The New York Times
Android .apk Repository

Virtual Machines for Dynamic Analysis

Results Database

Website

Crowdsourced User Interactions

Simulated User Interactions

API
thanks!

external colleagues:
Nathan Good (Good Research/UCB)
Marian Harbach (Google)
Helen Nissenbaum (Cornell)
Eyal Peer (Hebrew U.)
Joel Reardon (U. Calgary)
Franziska Roesner (UW)
Florian Schaub (UMich)
Narseo Vallina-Rodriguez (IMDEA/ICSI)

students/visitors:
Amit Elazari (UCB)
Mitra Hosseini (UTSA)
Tomasz Kosinski (Chalmers U.)
Arunesh Mathur (Princeton)
Leysan Nurgalieva (U. Trento)
Abbas Razaghpanah (Stony Brook)
Madiha Tabassum (UNCC)

group members:
Noura Alomar
Julia Bernd
Alisa Frik
Maritza Johnson
Nathan Malkin
Irwin Reyes
Nikita Samarin
Primal Wijesekera
thanks!

Serge Egelman
egelman@cs.berkeley.edu
@v0max