Measuring Privacy Risks in Mobile Apps at Scale

June 4, 2024

Lisa LeVasseur, Research Director
Bryce Simpson, Safety Researcher/Auditor
Saish Mandavkar, Quantitative Analyst/Researcher

INTERNET SAFETY LABS

PEPR '24

Internet Safety Labs

- Not for profit technology watchdog since 2019.
- We measure and expose safety risks in technology.
- Our programs:
  - **Safety Labels, Tools & Research**
    - Safety Labels (https://appmicroscope.org)
    - Safety Benchmarks
    - Responsible Disclosures
    - Software Safety Standards Panel
  - **Policy Advocacy**
  - **Safety Audits**

2022 US K-12 EdTech Safety Benchmark

2022 EdTech Benchmark

- Started with 2021 analysis of “School Utility Apps”
  - **Focus on SDKs as proxy for 3rd party data sharing**
- Wanted to look at EdTech more broadly
  - **And factor in observed network traffic.**
- Constructed US-wide EdTech Benchmark.
- First effort to measure privacy risks at scale.

Key Research Questions

- How much risky data sharing is happening in EdTech Apps?
- What impact do certifications, pledges, and privacy agreements have on app safety?
- What demographic patterns exist relative to risky tech behaviors?
- Is there a difference between SDK vs. observed network traffic?

2022 EdTech Benchmark

- National sample of 663 schools in 50 states plus DC.
  - **13 schools in each state:**
    - 4 elementary
    - 4 middle
    - 4 high
    - 1 private school
- Independent audit of school and district websites to identify recommended and required technologies
  - **Also used the Student Data Privacy Consortium resource https://sdpc.a4l.org/**
- Identified 1722 apps

2022 EdTech Benchmark (cont'd)

- Privacy Data collection on apps
  - **88,000+ data points**
  - **1357 apps' worth of network traffic**
- Technology behavior information collection on all schools
  - **29,000+ data points**

Outputs

- Tools
  - Tableau Summary
  - App Microscope
  - **Risk Dictionaries (Company, SDK, Subdomains)**
- Research Findings
  - **Findings Report 1: Overall app safety findings**
  - **Findings Report 2: School Tech Practices & 3rd Party Certifications**
  - **Findings Report 3: Demographic Analysis**
- Recommendations for EdTech Stakeholders

Methodology

App Data Collection

- Metadata Collection
  - **SDKs**
    - AppFigures
  - **App permissions, app age rating, last updated, etc.**
    - Google Play Store
    - Apple App Store
  - **Privacy policy information**
- Manual Testing
  - **Looking for:**
    - Presence of ads, including behavioral ads
    - Use of Webview
    - In-App Permission Requests
  - **Network traffic collection**
    - Charles Proxy (iOS)
    - PCAPdroid (Android)

Scoring App Privacy Risk

- Assess privacy risks in apps through assessing risk of app “ingredients” [SDKs] and observed behaviors
  - **SDK risk scores**
  - **Observed risky behaviors**
    - Presence of aggregator platforms
    - Presence of advertising or behavioral advertising
    - Use of Webview
- Stoplight Range of Scores
  - **Some Risk (least risk)**
  - **High Risk**
  - **Very High Risk (highest risk; was “Do Not Use”)**
  - **Not scored**

Scoring SDK Privacy Risk

- SDK risk score dependent on
  - **SDK function**
    - Risk impact X risk likelihood
    - Impact based on data accessed
    - Likelihood based on data monetization practices
  - **Publisher of SDK -> Company Risk Score**
    - Data Broker?
    - Otherwise monetize data?
    - Data breaches / fines / legal actions?

Scoring App Privacy Risk

SOME RISK	HIGH RISK	VERY HIGH RISK	NOT SCORED

Scoring App Privacy Risk

SOME RISK	HIGH RISK	VERY HIGH RISK	NOT SCORED
			School login required
			Paid app
			Broken app

Scoring App Privacy Risk

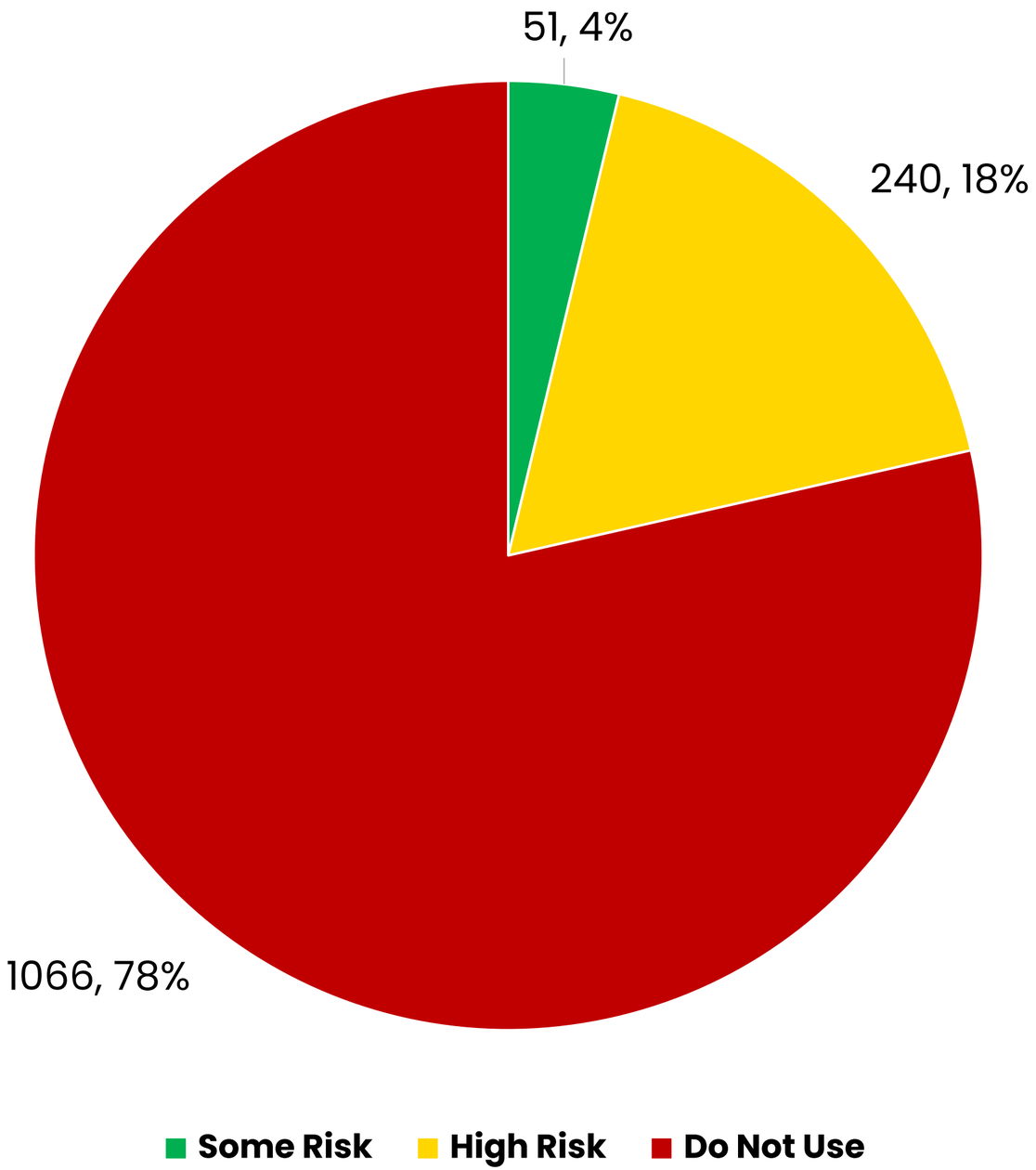
SOME RISK	HIGH RISK	VERY HIGH RISK	NOT SCORED
		Presence of advertising (any)	School login required
		Presence of one (1) or more registered Data Broker SDKs	Paid app
		Presence of one (1) or more of the following data aggregator platforms (SDKs or NW traffic): FB, Amazon, Twitter, Adobe	Broken app
		Presence of MaxPreps	
		Suspicious permission behavior.	

Scoring App Privacy Risk

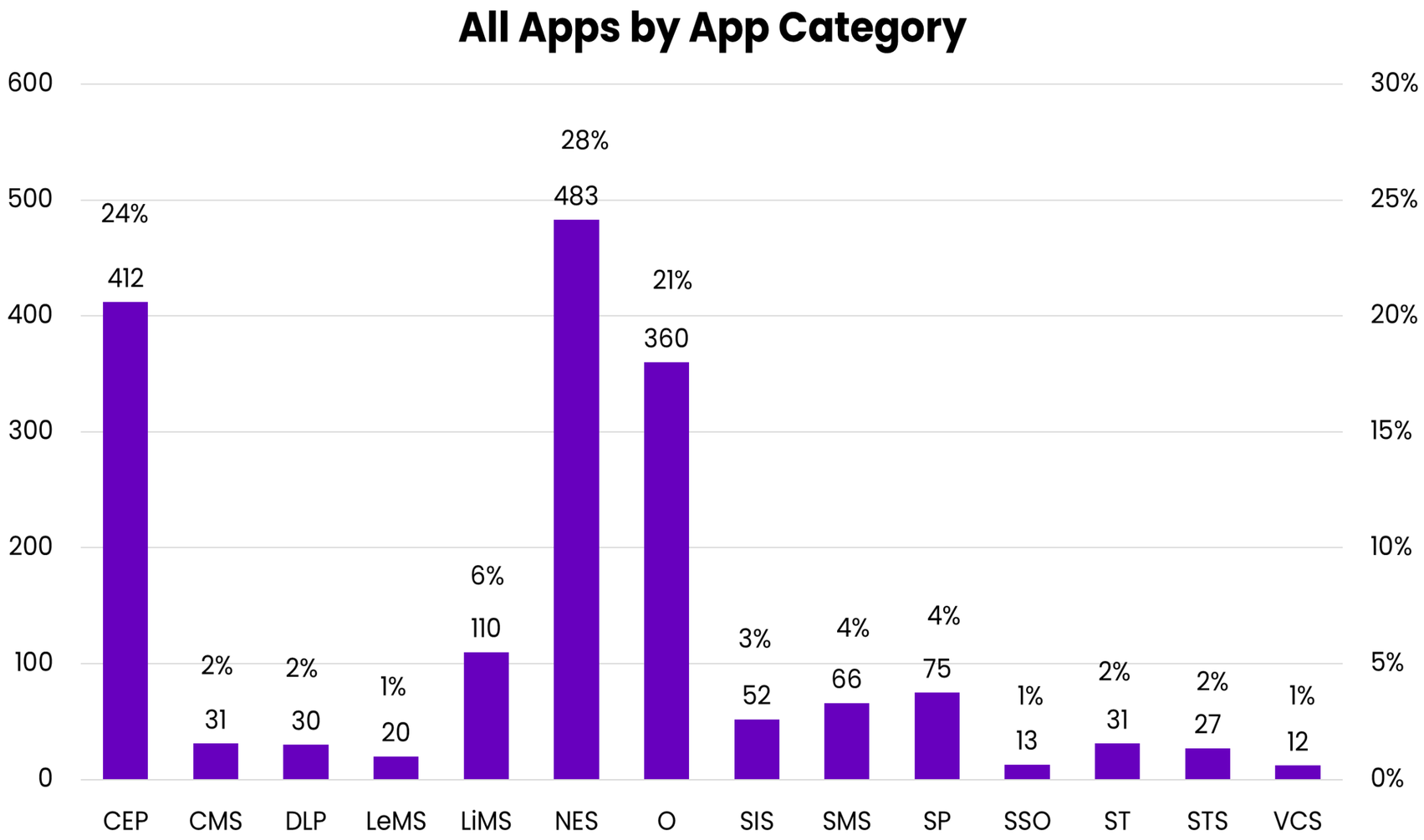
SOME RISK	HIGH RISK	VERY HIGH RISK	NOT SCORED
	Presence of at least one (1) SDK that is High Risk or Very High Risk	Presence of advertising (any)	School login required
	WebView Use	Presence of one (1) or more registered Data Broker SDKs	Paid app
	Presence of up to two (2) of the following data aggregator platforms (SDKs or NW traffic): Apple, Google	Presence of one (1) or more of the following data aggregator platforms (SDKs or NW traffic): FB, Amazon, Twitter, Adobe	Broken app
	Presence of a dangling domain	Presence of MaxPreps	
		Suspicious permission behavior.	

Key Learnings

App Scores (1357 Apps)



483 (28.1%) of all apps were NOT EdTech



This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Measuring Privacy – Network Traffic

- How to assess large volumes of network traffic to detect privacy risks?
  - **We catalogued all the subdomains observed in the network Traffic.**
  - **8,168 unique subdomains**
  - **3,211 unique domains**
  - **Scored riskiness by subdomain**
    - Subdomain Risk Dictionary

Measuring Privacy - SDKs

- How good of a proxy are SDKs for *actual* data sharing?
  - **40%** of companies identified in SDKs were seen in network traffic

	Avg # Expected Companies	Avg # Expected Companies in NW Traffic	Avg # Unexpected Companies in NW Traffic
Webview - With	4.95	1.89	**12.63**
Webview - Without	4.28	1.43	2.57
Advertisements - With	5.58	2.11	**23.95**
Advertisements - Without	4.47	1.6	5.01
Behavioral Advertisements - With	5.44	2.14	**33.73**
Behavioral Advertisements - Without	4.58	1.64	5.53
All Tested Apps With 1+ SDK Companies	4.67	1.7	8.39

What Worked

- Independent data collection worked.
- Privacy risk scoring was/is useful.
- Having entire vertical data gives necessary context.
- Programmatic scoring of companies, SDKs, and subdomains.
- Assessing behavior NOT what maker says in privacy policy / terms of service.

Impacts

- Tons of awareness
  - **52 news stories written since Dec 2022**
- **9689** safety label views since Oct 2023
- **2016** views of Tableau dashboard
- One state board of education

What We'd Do Differently

- More tightly define data collection.
- Iterate on data collection and results processing.
- Process the data as close to collection as possible.
- Collect full payloads (network traffic).
- Measure twice, cut once on sampling methodology.
- Track data element sharing.

Scoring Changes

- Factor network traffic into App scores.
- Adding in a 4th score tier
- Score 1st party and 3rd parties separately
  - **Factor in Parent Company Risk Score**
- Programmatic scoring of Apps
- More rigor around “Aggregator Platforms” as highest risk platforms

Call to Action

- Use the app safety labels as measures of privacy risk
  - **Other risks coming in the future!**
  - **Let us know what's useful and what's missing**
- Join the Software Safety Standards Panel to help us define what goes in safety labels
- Safetypedia….

Thank You!

Lisa.LeVasseur@InternetSafetyLabs.org

Bryce.Simpson@InternetSafetyLabs.org