Measuring Privacy Risks in Mobile Apps at Scale

June 4, 2024

Lisa LeVasseur, Research Director
Bryce Simpson, Safety Researcher/Auditor
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.
- 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?
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/](https://sdpc.a4l.org/)

Identified 1722 apps
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 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
SDK risk score dependent on

- **SDK function**
  - Risk impact $\times$ 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

<table>
<thead>
<tr>
<th>SOME RISK</th>
<th>HIGH RISK</th>
<th>VERY HIGH RISK</th>
<th>NOT SCORED</th>
</tr>
</thead>
</table>

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.
### Scoring App Privacy Risk

<table>
<thead>
<tr>
<th>SOME RISK</th>
<th>HIGH RISK</th>
<th>VERY HIGH RISK</th>
<th>NOT SCORED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>School login required</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Paid app</td>
</tr>
<tr>
<td>Presence of one (1) registered Data Broker SDKs</td>
<td>Presence of a dangling domain</td>
<td>Presence of MaxPreps Suspicious permission behavior.</td>
<td>Broken app</td>
</tr>
</tbody>
</table>
# Scoring App Privacy Risk

<table>
<thead>
<tr>
<th>SOME RISK</th>
<th>HIGH RISK</th>
<th>VERY HIGH RISK</th>
<th>NOT SCORED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Presence of advertising (any)</td>
<td>Presence of one (1) or more registered Data Broker SDKs</td>
<td>Presence of MaxPreps</td>
</tr>
<tr>
<td></td>
<td>School login required</td>
<td>Paid app</td>
<td>Suspicious permission behavior.</td>
</tr>
<tr>
<td></td>
<td>Presence of up to two (2) of the following data aggregator platforms (SDKs or NW traffic): FB, Amazon, Twitter, Adobe</td>
<td>Broken app</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Scoring App Privacy Risk

<table>
<thead>
<tr>
<th>SOME RISK</th>
<th>HIGH RISK</th>
<th>VERY HIGH RISK</th>
<th>NOT SCORED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of at least one (1) SDK that is High Risk or Very High Risk</td>
<td>Presence of advertising (any)</td>
<td>School login required</td>
<td></td>
</tr>
<tr>
<td>WebView Use</td>
<td>Presence of one (1) or more registered Data Broker SDKs</td>
<td>Paid app</td>
<td></td>
</tr>
<tr>
<td>Presence of up to two (2) of the following data aggregator platforms (SDKs or NW traffic): Apple, Google</td>
<td>Presence of one (1) or more of the following data aggregator platforms (SDKs or NW traffic): FB, Amazon, Twitter, Adobe</td>
<td>Broken app</td>
<td></td>
</tr>
<tr>
<td>Presence of a dangling domain</td>
<td>Presence of MaxPreps</td>
<td>Suspicious permission behavior.</td>
<td></td>
</tr>
</tbody>
</table>
Key Learnings
App Scores (1357 Apps)

- **High Risk**: 240, 18%
- **Some Risk**: 51, 4%
- **Do Not Use**: 1066, 78%

---

**Scores Distribution**
- 1357 Apps
- 18, 51
483 (28.1%) of all apps were NOT EdTech
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
How good of a proxy are SDKs for *actual* data sharing?

- **40%** of companies identified in SDKs were seen in network traffic

<table>
<thead>
<tr>
<th></th>
<th>Avg # Expected Companies</th>
<th>Avg # Expected Companies in NW Traffic</th>
<th>Avg # Unexpected Companies in NW Traffic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Webview – With</td>
<td>4.95</td>
<td>1.89</td>
<td>12.63</td>
</tr>
<tr>
<td>Webview – Without</td>
<td>4.28</td>
<td>1.43</td>
<td>2.57</td>
</tr>
<tr>
<td>Advertisements – With</td>
<td>5.58</td>
<td>2.11</td>
<td>23.95</td>
</tr>
<tr>
<td>Advertisements – Without</td>
<td>4.47</td>
<td>1.6</td>
<td>5.01</td>
</tr>
<tr>
<td>Behavioral Advertisements – With</td>
<td>5.44</td>
<td>2.14</td>
<td>33.73</td>
</tr>
<tr>
<td>Behavioral Advertisements – Without</td>
<td>4.58</td>
<td>1.64</td>
<td>5.53</td>
</tr>
<tr>
<td>All Tested Apps With 1+ SDK Companies</td>
<td>4.67</td>
<td>1.7</td>
<td>8.39</td>
</tr>
</tbody>
</table>
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 4\textsuperscript{th} score tier
- Score 1\textsuperscript{st} party and 3\textsuperscript{rd} 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