Expecting Targeted Advertisements? Characterizing Private Data Leakage in Fertility Tracking Apps

Yeeun Jo*, Mahnoor Jameel†, Camille Cobb‡ and Adam Bates§
*†‡§ *Siebel School of Computing and Data Science, University of Illinois Urbana-Champaign*
Emails: {yeeunjo2, mjameel2, camillec, batesa}@illinois.edu

***Abstract*—Users' perceptions of menstrual tracking privacy is a subject of extensive study, but little attention has been paid to the technical aspects of fertility tracking apps' data handling practices. We propose a measurement study of the fertility tracking app ecosystem, leveraging network and program analysis to explore apps' communications with third party ad networks. Selecting a corpus of fertility tracking apps from the Google Play Stores, we systematize and define user interaction paths to collect data on registration, menstruation journaling, and pregnancy tracking features. Our analysis of TLS-stripped network traffic, which is ongoing, has uncovered examples of apps transmitting fine-grained user data, such as pregnancy status and trimester, to ad networks. However, widespread nested encryption and obfuscation practices motivate a need for incorporating program analysis into the study.**

1. Motivation

Fertility tracking apps enable users to take daily notes and track premenstrual syndrome, pregnancy symptoms, moods, intercourse, and period flow. Some apps use this data to provide a personalized content stream designed to engage users with timely health insights and tips tailored to their unique needs, search questions in online Q&A forums, and share data with support groups. Since the launch of Clue in 2013 and Flo in 2015, these apps have become increasingly popular: for example, Flo Health's app portfolio sees nearly 75M + monthly users [1]. Users of these apps reveal highly sensitive and intimate information in their accounts, making them susceptible to violations of their health privacy by private corporations [2, 3, 4].

This extensive network of data recipients raises important questions about users' privacy and the flow of sensitive information. There are several key challenges that limit our understanding of information exposure from apps on fertility tracking and their privacy implications. First, prior work on fertility tracking privacy following the overturn of Roe v Wade has focused on user's perceptions and mental models [5, 6, 7] or developers' statements and policies [8, 9] rather than delve into the technical ground truth of app's data handling. We believe this oversight may be attributable to a gap in methodological expertise – one we hope to address in this study. Second, mobile advertising networks are designed to actively resist third party analysis, engaging in code and data obfuscation, nested encryption, and other practices to frustrate transpareny efforts. However, we hypothesize that the need for cooperation in this ecosystem between ad networks and app developers exposes an opportunity for independent investigators.

In this paper, we propose to address these challenges by conducting an in-depth analysis of information exposure from fertility tracking apps. Our research questions include:

RQ1: How does advertising manifest differently in fertility tracking apps given the intrinsically sensitive nature of user data?

RQ2: How, and to what extent, do fertility tracking app developers actively participate in targeted advertising practices?

RQ3: To what extent is private user data shared with ad networks and other third parties in fertility tracking apps?

An overview of the planned study is as follows. We begin by systematizing the fertility tracking app ecosystem in terms of feature categories, leveraging this taxonomy to define a data collection procedure based on representative user interactions. We then exercise these interaction paths in an instrumented environment that collect network and program trace data. We then analyze these traces for evidence of private data leakage.

In our ongoing analysis of the network traces, we developed tooling to identify active ad networks in each apps, recovered the purpose and function if different Ad Network API endpoints, and have begun to review URL and JSON data sent by the app to the network. Preliminary results have uncovered several instances of apps emitting user data – including general audience information, pregnancy status, and even current trimester – to ad networks. However, widespread nested encryption and obfuscation motivate the incorporation of program analysis approaches, which we are currently exploring.

2. Methods

As a first step towards understanding the implications of advertising in fertility tracking apps, we set out to generate a dataset of network traffic activity from a representative selection of Android apps from the Google Play store.

Interaction Session	Data Types
Registration	Personal Identification, (Reproductive State Indicators, Menstrual Activity)
Menstruation Journaling	Menstrual Activity, Physical Symptoms, Emotional and Behavioral symptoms
Fertility Journaling	Hormonal or Cycle-specific Physiological Indicators, Sexual and Relationship Data
Sleep Journaling	Lifestyle and General Health
Contraceptions Journaling	Sexual and Relationship Data, Physical Symptoms, Emotional and Behavioral symptoms, and Medical data
Pregnancy Journaling	Sexual and Relationship Data, Reproductive State Indicators, Pregnancy Tracking, and Medical data
Postpartum Journaling	Physical Symptoms, Emotional and Behavioral symptoms, and Medical Records
Pregnancy Loss Journaling	Reproductive State Indicators, Physical Symptoms, Emotional and Behavioral symptoms

TABLE 1: Interaction Sessions used in Data Collection.

2.1. App Selection

We primarily identified fertility apps through keywords searches on the Google Play Store using the following phrases: "Period Tracker," "Period Cycle", "Fertility Tracker" "Ovulation Tracker," "Pregnancy Tracker & Calendar." Our searches, conducted in April 2025, identified 254 total apps. We then considered the app for inclusion in the study based on the following criteria:

- Provides features related to menstruation, fertility, pregnancy, or menopause.
- Accepts as input user data related to (in)fertility (e.g., period cycles, pregnancy, symptoms).
- Includes calendar-based features, a proxy for self-tracking functionality.
- Is a live service that is actively maintained.

We applied these criteria to filter from our consideration apps that did not handle sensitive user data related to fertility, and thus do not have the same privacy concerns as fertility trackers. We deliberately chose not to exclude apps based on install count estimates or user ratings, because these metrics may not accurately reflect app quality or behavior and may introduce popularity bias into our dataset. In addition to Google Play search results, we included 4 additional apps: 1 app developed by a nonprofit organization Euki; 1 app that appeared in a web advertisement served to one of the authors claiming to be a privacy-preserving period tracker Stardust; and 2 apps associated with tracking through sensing identified based on one of the authors' prior experience with trackers.

After filtering based on application features, we were left with 258 fertility tracking apps in our candidate set. Of the remaining 258 apps, we selected 29 that represented a balanced mix of app categories, developer types, and install counts, ensuring coverage of both popular and niche applications. The median number of installs in our sample ranged between 50 and $100,000,000$, while the average user rating was 4.4. While offering overlapping features, the final set of apps also focus on different aspects of reproductive health: *fertility prediction apps* representing the conception-focused functions (n=5), *pregnancy support apps* representing the week by week development focused functions (n=12), *period apps* representing the period monitoring functions (n=8), and *hybrid apps* representing the multiple reproductive modes (n=4).

2.2. Profiling Ad Network Activity

Having identified a corpus of fertility tracking apps and provisioned a profiling environment, we now require a systematic procedure for eliciting advertising-related behavior from the apps. We also require the procedure to be applicable to all apps in our corpus, regardless of specific features, and to facilitate comparison between apps. Instead of bottom-up software analysis techniques, we adopt a top-down approach in which an experimenter manually interacts with the apps in a pre-determined procedure, summarized in Table 1. By specifying an interaction procedure for each fertility tracking feature, the network traces become easier to interpret and compare across apps. This also allows us to control the information provided to the app and test for its impact on the resulting ad network traffic.

For each app, we download the Android APK file, then installed them in the Waydroid Emulator. Waydroid was set to default configurations, except that the environment was modified to intercept HTTPS traffic, strip TLS, and record the packet captures to disk. Specifically, we used a rootkit to escalate privilege in the operating system,, installed a custom self-signed Certificate Authority (CA) certificate to Android's certificate store, then deployed `mitmproxy` with the new CA keypair.[1]

3. Preliminary Results

Thus far, we have recorded 7,829 requests to ad networks API's, spanning 29 ad networks. We have cleaned and pre-processed the data. systematizing the API endpoints 196 high-level functions and reviewed public documentation to determine there purpose – ad retrieval, impression tracking, static content, etc. We then manually reviewed the request-response pair for all non-static resource requests for evidence of private data leakage.

Our analysis thus far has identified a handful of fertility tracking apps that send user data to ad networks for the purposes of targeted advertising. The granularity of user data descriptions ranges from general audience descriptions to, in one case, reporting the user's pregnancy status and even their current trimester. Moreso than widespread data leakage, however, we have encountered widespread obfuscation – opaque data serialization routines, nested encryption, and dynamic code loading – that have underscored the limitations of a purely network-based approach. To resolve this, we are expanding our data collection and analysis methodology to incorporate software-based dynamic code instrumentation to hook calls to Ad Network SDK function calls in developer code.

1. Details at https://github.com/ddxv/mobile-network-traffic.

Acknowledgement

This paper results from the SPLICE research program, supported by a collaborative award from the National Science Foundation (NSF) SaTC Frontiers program under award 1955228. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of NSF. Any mention of specific companies or products does not imply any endorsement by the authors, by their employers, or by the NSF. We thank our anonymous reviewers for their helpful comments in revising this paper.

We investigate apps that appeared across a number of categories in the Google Play store including "Health and Fitness" (n=13), "Medical" (n=6), "Parenting" (n=9), "Lifestyle" (n=1), and "Education" (n=1). Twenty-one apps used a free-with-ads monetization models, one app used paid hardware–access model, while the remaining seven relied on in-app purchases. The inclusion and exclusion of the app samples does not indicate that these apps eliminate any needs of doctors or medicines, our aim is more about understanding the obfuscated or clearly presented presence of advertisement associated with data leakage in fertility tracking apps.

References

[1] Databricks, "Flo health accelerates ai innovation and personalizes care with databricks," 2025.

[2] N. Felizi and J. Varon, "Menstruapps - how to turn your period into money (for others)," https://chupadados.codingrights.org/en/menstruapps-como-transformar-sua-menstruacao-em-dinheiro-para-os-outros-2/, 2017.

[3] V. Rizk and D. Othman, "Quantifying fertility and reproduction through mobile apps: a critical overview." *Arrow for Change*, vol. 22, p. 13–21, 2016.

[4] L. M. Malki, I. Kaleva, D. Patel, M. Warner, and R. Abu-Salma, "Exploring privacy practices of female mhealth apps in a post-roe world," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–24.

[5] J. Cao, H. Laabadli, C. H. Mathis, R. D. Stern, and P. Emami-Naeini, ""i deleted it after the overturn of roe v. wade": Understanding women's privacy concerns toward period-tracking apps in the post roe v. wade era," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–22.

[6] Q. Song, R. Ma, Y. Kou, and X. Gui, "Collective privacy sensemaking on social media about period and fertility tracking post roe v. wade," *Proceedings of the ACM on human-computer interaction*, vol. 8, no. CSCW1, pp. 1–35, 2024.

[7] N. Mcdonald and N. Andalibi, ""i did watch 'the handmaid's tale": Threat modeling privacy post-roe in the united states," *ACM Transactions on Computer-Human Interaction*, vol. 30, no. 4, pp. 1–34, 2023.

[8] Q. Song, R. H. Hernandez, Y. Kou, and X. Gui, ""our users' privacy is paramount to us": A discourse analysis of how period and fertility tracking app companies address the roe v wade overturn," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–21.

[9] A. I. Hudig and J. Singh, "Intimate data sharing: Enhancing transparency and control in fertility tracking," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 2025, pp. 1–24.