Backtracking Online Trackers: Third-Party Trackers in Health-Related Advertising

Carnegie Mellon University School of Computer Science

Brandon Pek with thanks to project advisor Timothy Libert

Problem

Juxtaposition between sensitive nature of personal health information and easy availability of such information through online tracking means, exploited for profit purposes in targeted advertising of specific users.

Failure of Health Insurance Portability And Accountability Act (HIPAA) enacted in 1996 to extend the safekeeping of patient medical information into digital space.

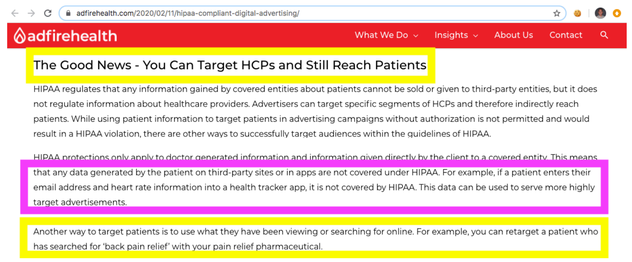

Figure 1: Adfire Health proudly markets its exploitation of HIPAA loophole

Further motivated by personally identifiable medical information from multiple platforms: healthcare apps, virus location tracing apps



Goals

1. Increase awareness of third-party trackers, the information they store and the applications of such data, in health-related information.
2. Spark discussion on what constitutes sensitive medical information.
3. Call for governmental regulation in digital health data protection.

Abstract

We explore advertising and online tracking technologies in ethically controversial and personally-sensitive areas of healthcare. We begin by analyzing the social context and lack of regulation in healthcare advertising that allow for tracking of potentially concerning health conditions like mental illness or sexually transmitted diseases. We then explore the failures of self-regulatory advertising associations like the Network Advertising Initiative and member violations with respect to healthcare tracking. We proceed to collect cookie data from a range of healthcare websites, from medicare hospitals to county-level health departments, and backtrack against the unregulated players that have placed tracking cookies in these health-related websites. We discuss some of the more concerning actions and companies uncovered over the course of this study, explore the future direction of including population surveys on what society deems sensitive healthcare information, and hope to inspire policy-level enforcement changes with the evidence uncovered.

Methodology

Top-Down: Analyze effectiveness of self-regulation

Inspect all 97 members of Network Advertising Initiative (NAI), a self-regulatory advertisement industry group for violations of its two proposed policies regarding health data:

(1) Explicit user consent must be obtained for sensitive health information

(2) Any member engaging in health advertising must publicly disclose all health-related audience segments (health topics used to classify users).

Bottom-Up: Backtrack cookies to uncover hidden, unregulated firms

Crawled 63584 cookies from 4664 health-related web pages formed from medicare hospitals, county, state, federal-level health departments.

Crawled 504870 cookies from 9277 most visited web pages to form general corpus.

Formulated a scoring algorithm to find specifically health-related cookies, cookies that had high frequency in health pages but not general pages:

$$cookieScore = healthFrequency \cdot \ln\left(1 + \frac{1}{corpusFrequency + 1}\right)$$

Backtracked the higher scoring cookies to their owners to uncover smaller and hidden players in healthcare tracking.

Tracking Trackers

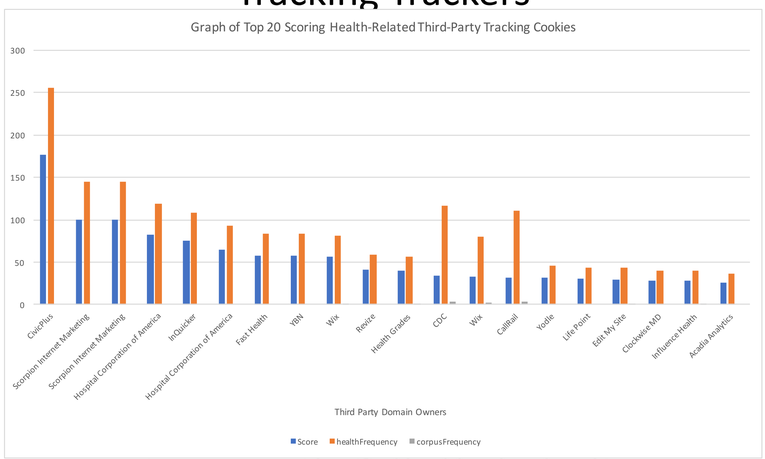

Figure 2: Top 20 identified health-related cookies

Traced back to health-related advertising firms like Scorpion Internet Marketing, HealthGrades, CallRail and Yodle.

Generally offer, real-time tracking for targeted advertising and analyzing network traffic and searches for search engine optimization.

In the Absence of Regulation - Yodle Case Study

- User testimonials coming from sensitive medical fields like therapists, home care for chronic illnesses, and marriage counselors.
- Allude to the sensitive nature of user data that Yodle is storing.
- A cause for concern: Yodle tracking cookies appears in almost 0.94% of sampled medicare hospital website domains.

Failures of Self-Regulation

Many members of the Network Advertising Initiative (NAI) fail the 2 proposed self-regulated conditions relating to health advertising.

Most blatantly: Adfire Health for violating (2) and proudly flaunting capabilities of obtaining user medical information from third-party tracking in Figure 1.

Potentially allowed by NAI's vague definition of what constitutes sensitive medical information, allowing for inefficient self-policing:

> *"The NAI acknowledges that these are subjective considerations and that no one factor is determinative. Therefore, any member company that conducts a reasonable analysis of a health condition and determines that it does not meet the factors of a sensitive health segment will not be in violation of the Code even if other stakeholders, including, but not limited to, the NAI compliance team, arrive at a different conclusion."*
>
> *– NAI Code of Conduct 2020.*

Most reputable company: Google for engaging in health advertising but violating (2), not publicly disclosing health-related audience segments.

Most concerning: PulsePoint for its massive tracking reach (Fig 3), its violation of (1) in not having an explicit opt-in user consent, and the engagement in targeted advertising of concerning audience segments like mental disorders, STDs, addiction, cancer and many more.

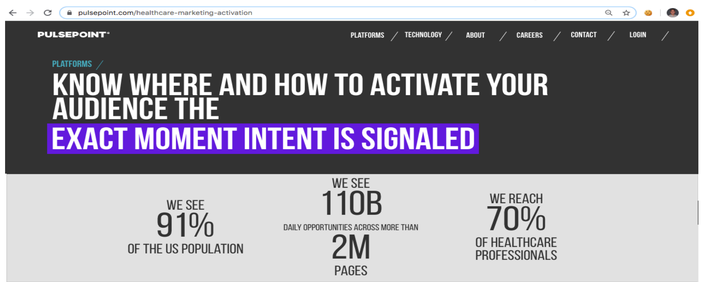

Figure 2: Extent of third-party tracking of sensitive medical information by PulsePoint, which does targeted advertising in concerning fields like cancer, STDs, mental disorders.

Conclusion

Health-related tracking under regulated proper use has potential for good in promoting community-wide health, fighting fake news related to false medical remedies, reducing market inefficiencies.

We have shattered the façade that medical information is confidential, from online side-channel vulnerabilities.

Future studies will include a census survey to explore what society deems sensitive medical information, more potential in this field for advanced cookie clustering to detect intentionally hidden players.