Online Tracking of Kids and Teens: Gaming Sites Not as Bad as Others

Natalija Vlajic, Marmara El Masri, Pinghua Yan
Dept. of Electrical Engineering & Computer Science
York University, Toronto, Canada
vlajic@cse.yorku.ca, marmara@my.yorku.ca, yan1993@cse.yoru.ca

Abstract—**The recent news of large-scale online tracking campaigns, which gave way to systematic misuse of the collected user-related data, have left millions of people deeply concerned about the state of their online privacy as well as the state of the overall information security in the cyber world. While most to-date revelations pertaining to user tracking are related to Web-sites and social media intended for adult online users, relatively little is known about the prevalence of online tracking in Web-sites geared towards children and teens.**

The goal of our work was to examine several representative Web-sites found in Alexa's top 50 'Kids and Teens' category for potential signs of user tracking. The key findings of our work are both unexpected and alarming. Namely, we found overwhelming evidence of wide-spread and highly covert user tracking in five examined Web-sites that appeared to be 'general and/or educational' in its content and nature. Moreover, the majority of the discovered tracking was in direct conflict with Children's Online Privacy Protection Act (COPPA), as it was performed without parental consent and by third-party advertising and tracking companies. On the other hand, and to our surprise, there was very little evidence of user tracking in the other five 'gaming' Web-sites examined in our study. Given the significance of these first preliminary results, we are currently expanding our study to a much larger pull of Web-sites intended for online users under the age of thirteen.

Keywords — online tracking, user privacy, COPPA

I. ONLINE TRACKING: PROS, CONS, AND COPPA

A. Pros and Cons of Online Tracking

The term *online tracking* refers to the process of recording, measuring and analyzing the behavior of individual human users while browsing the WWW. On one hand, from the financial and performance perspective, this process is seen as one of the key drivers for optimization and growth of online-based businesses and organizations. (Namely, understanding how customers behave and what they want has always been the key to success of any business in any industry.) On the other hand, from the perspective of individual users, online tracking is considered to be a major threat to online privacy, as it can lead to extraction and leakage of sensitive personal information. Furthermore, the recent incidents involving the tracking of Facebook users have shown that the negative implications of online tracking can now easily cross the boundaries of the cyber-domain and impact the operation and stability of the actual real-world [1].

B. COPPA

The Children's Online Privacy Protection Act (COPPA) is a US federal law that was enacted in October 1998. COPPA legislation was put in place to regulate (i.e., limit) the collection, use or disclosure of personal information by commercial Websites and online services (including mobile apps) directed to children under the age of thirteen as well as operators of general audience Web-sites or online services with actual knowledge that they are collecting, using or disclosing personal information from children under thirteen [2]. In particular, COPPA requires these Web-sites and services to post a complete privacy policy, notify parents directly about their information collection practices, and get verifiable parental consent before collecting personal information from their children or sharing it with others [3].

Unfortunately, ever since COPPA was passed, the on-the-ground implementation of this legislation was slow and problematic, with sporadically reported incidents of Web-sites and services either avoiding or violating many of COPPA provisions [4]. In the most recent of such incidents, the Disney company (often considered to be the 'gold standard' when it comes to 'customer experience' and 'user trust') has been alleged of violating COPPA in 42 of their online apps [5].

Fig. 1 Article on COPPA-violation claims against Disney [5]

II. COMMON ONLINE TRACKING TECHNIQUES

The most common techniques of online user tracking, together with their most significant pros and cons, include [6]:

- *IP based tracking* allows tracking of users on multiple Web-sites belonging to different entities, and it cannot be prevented by simple change(s) in browser settings. However, this form of tracking does not work well if multiple users share the same computer or use IP-anonymization.
- *Cookie based tracking* requires that a small piece of server-generated data be explicitly stored on the user's device and then be returned whenever the user revisits the given server/site. While generally very effective, this form of tracking can be obstructed if the user disallows or cleans cookies from his browser.
- *Cache (i.e., Etag) based tracking* avoids storing explicit tracking data in the client's memory, but instead relies on server-generated data/objects that get implicitly and automatically stored in the browser's cache (such as Etag which get exchanged by means of HTTP headers). This form of tracking can also be obstructed if the user regularly clears his browser's cache.
- *Session based tracking* does not require that any server-generated data get stored in the browser's memory or cache. Instead, the server inserts user/session-id into all relevant URLs sent to the user. Unfortunately, the main drawback of session based tracking is that it is very 'visible' and cannot be easily obscured from the user, unlike the other three methods.

Given that each of the above described tracking techniques has its own set of drawbacks, most real-world Web-sites deploy two or more of these approaches combined.

III. Experimental Objectives and Obtained Results

The main objectives of our research were twofold:

Objective 1)   Study several representative Web-sites found in Alexa's top 50 'Kids and Teens' category for potential signs of user tracking. In particular, at this first stage of our research, we have concentrated our attention to Web-sites from two specific categories – general/education sites and gaming sites, as enlisted in Table 1.

General/Education Sites	Gaming Sites
thesaurus.com	blizzard.com
reverse.net	leagueoflegends.com
britannica.com	bulbgarden.net
howstuffworks.com	nintendo.com
citationmachine.net	ea.com

Table 1   List of Web-sites examined in our study

Objective 2)   Specifically focus on finding the signs of user tracking in **images** embedded/placed in the above enlisted Web-sites. Namely, from the perspective of online user tracking, images could/can be deployed as 'perfect trojan' since: a) today's users are accustomed to retrieving/viewing Web-pages that are graphically very rich, (i.e., contain a large numbers of embedded images often hosted at different Web-domains), b) during the rendering of a requested Web-page, all of its embedded images (regardless of their actual number or their respective domains of origin) are automatically and instantaneously retrieved by all commercial browsers operating in their default settings, c) most Web vulnerability scanning tools perform minimal if any scanning of images for potential signs of misuse, d) in order to prevent retrieval and displaying of images, a user would have to be sufficiently skilled and knowledgeable on how to appropriately modify the default settings of his browser.

The complete set of our obtained results can be retrieved from http://www.cse.yorku.ca/~vlajic/Tracking.pdf. The most important observations derived from these results include:

- The number of images found in the examined sites varied but did not appear to be related to the actual nature of the site - both the minimum number (14) and maximum number (113) of images were found in two of the gaming sites.

- However, as shown in Table 2 and 3, gaming sites appear much 'cleaner' in terms of the origin and purpose of the embedded images, as most of their images come from the actual 'host' domain and are intended to be visible (i.e., presented) to the user. On the other hand, we found that most images embedded in general/education sites originate from various 'other' domains (most of which are owned by advertising, tracking and Web-analytics companies) and are largely obscured (i.e., never presented) to the user.

- Most images coming from advertising, tracking and Web-analytics domains come with (i.e., carry) several different types of cookies as well dynamic-URLs, both being clear signs of the ultimate purpose of these images – user tracking.

- Sporadic incidents of (possible) user tracking by ETags were also identified in several of the examined Web-sites.

References

[1] https://www.theatlantic.com/technology/archive/2017/10/what-facebook-did/542502/

[2] Federal Trade Commission, "*Children's Online Privacy Protection Rule: A Six-Step Compliance Plan for Your Business*", June 2017.

[3] Federal Trade Comission, "*Protecting Children's Privacy Under COPPA: A Survey on Compliance*", April 2002.

[4] L. A. Matecki, "*Update: COPPA is Ineffective Legislation! Next Steps for Protecting Youth Privacy Rights in the Social Networking Era*", Northwestern Journal of Law & Social Policy, Volume 5, Issue 2, Article 7, 2010.

[5] https://www.pcmag.com/news/355509/lawsuit-claims-disney-is-violating-coppa-tracking-kids-in-4

[6] N. Vlajic, et al., "*Anonymity of TOR Users Demystified*", Proceedings of IEEE 2017 Symposium on Cyber Warfare, Cyber Defence, and Cyber Security (CSCI-ISCW), December 2017

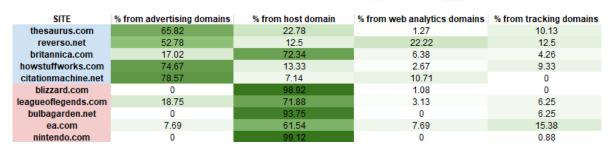Table 2   Percentage of images based on their domain of origin [LEGEND:]
0%   100 %

SITE	% from advertising domains	% from host domain	% from web analytics domains	% from tracking domains
thesaurus.com	65.82	22.78	1.27	10.13
reverso.net	52.78	12.5	22.22	12.5
britannica.com	17.02	72.34	6.38	4.26
howstuffworks.com	74.67	13.33	2.67	9.33
citationmachine.net	78.57	7.14	10.71	0
blizzard.com	0	98.92	1.08	0
leagueoflegends.com	18.75	71.88	3.13	6.25
bulbagarden.net	0	93.75	0	6.25
ea.com	7.69	61.54	7.69	15.38
nintendo.com	0	99.12	0	0.88

Table 3   Percentage of visible/invisible images from each particular category [LEGEND:]
Visible   Invisible
(100% visible)   (0% visible)

SITE	% from advertising domains	% from host domain	% from web analytics domains	% from tracking domains
thesaurus.com	3.85	100	0	12.5
reverso.net	2.63	100	6.25	N/A
britannica.com	12.5	97.06	0	0
howstuffworks.com	8.93	90	0	14.29
citationmachine.net	6.06	83.33	0	N/A
blizzard.com	N/A	100	0	N/A
leagueoflegends.com	0	100	0	0
bulbagarden.net	N/A	100	N/A	0
ea.com	0	100	0	0
nintendo.com	N/A	99.11	N/A	0