Research Statement

Hieu Le
levanhieu.com

Protecting user privacy and improving awareness of data collection practices by online platforms are significant challenges. Personal data is routinely collected to personalize experiences and provide services. However, it is also used to manipulate buying behaviors (*e.g.,* ad targeting), spread misinformation (*e.g.,* political ads), and it can even be sold to undisclosed parties. My long-term career goal is to give people control over their personal data and experiences on online platforms. My research will advance this goal by developing frameworks and methodologies to enhance transparency, auditability, and control for users.

Overview

Transparency and Auditability. Personal data collection practices are often opaque to users and researchers, especially for emerging platforms. My research elucidates these practices by conducting large-scale network measurements of apps on smart TVs and Oculus VR. We employ specialized techniques to decrypt network traffic, extract data types that are exposed[1], and identify the companies that collect the data (and for which purposes), such as advertising and tracking (A&T) [9, 10]. We find that first, third, and platform parties all collect a multitude of data types, which makes identifying users possible. In addition, data is often collected for nonessential purposes (*e.g.,* marketing vs. functionality) and privacy policies are often vague about the data being collected. Thus, my research develops tools to audit these practices with their expected behavior. To do so, we utilize the contextual integrity framework to analyze the consistency between the network traffic and privacy policy [9]. Our research and findings have been presented to the Federal Trade Commission (FTC) at PrivacyCon 2021, and 2022 [7, 6], Consumer Reports, and privacy-focused industry players such as Duck Duck Go. Some of our research methodologies have been integrated into workshops that educate underrepresented college students about IoT privacy [8].

Personal Data Control. Beyond informing users about data collection practices, users should also be able to control how their data is collected. Millions of users worldwide depend on privacy-enhancing technologies (PETs), such as adblockers, to block A&T. However, the efficacy of PETs often rely on error-prone human trial-and-error processes. They do not scale well across millions of websites and apps, especially over time — human experts may update PETs almost hourly [5]. My research fortifies the effectiveness of PETs for the web by identifying their pain-points and developing frameworks and methodologies to automate them [4, 5]. In particular, we train a machine learning (ML) classifier to notify human experts when to update PETs (for websites when it is no longer effective) and develop a reinforcement learning (RL) framework (and tool) to automate those updates. As a result, my research reduces these manual human efforts while achieving comparable performance, thus, addressing problems with efficacy and scalability. Our research has been presented for the last four years at the Ad-Filtering Dev Summit, where companies that develop PETs (*e.g.,* browsers, adblockers) gather to disseminate emerging ideas for PETs and how they can influence policy to protect user privacy [2]. Some of our research has been customized and adopted by these companies [3, 5].

[1] *e.g.,* personally identifiable information (PII), metadata to fingerprint the user, and platform-specific data

Current Research

Robust and Scalable PETs for the Web. PETs, such as adblockers, are powered by filter rules. They are string-based patterns that block A&T requests but are manually created and maintained by human experts. Conversely, publishers and advertisers employ specialized techniques to circumvent adblockers. Thus, over time, filter rules become ineffective and must be updated. We conduct a longitudinal analysis to measure how frequently experts update rules by inspecting their GitHub history: they update rules almost hourly to combat circumvention. We build an AI-based tool, **CV-Inspector** [5], to reduce this human effort. It detects when websites have circumvented adblockers. We extract features from HTTP requests and HTML DOM modalities using differential analysis (*i.e.,* how a website naturally behaves *vs.* when an adblocker is installed) to train an ML classifier. Our evaluation shows that our model can accurately detect instances of adblock circumvention, reducing the human effort by 98%. This work was presented at the Ad-Filtering Dev Summit and customized for industry use in 2021.

Although CV-Inspector notifies experts when to update filter rules, it does not create or update them. Researchers have built AI-based tools that detect and block A&T to replace filter rules altogether or designed approaches that assist in their creation. However, these approaches rely on existing filter rules to label their ground truth, causing a *circular dependency*: experts must maintain both rules and ML models. In addition, they do not evaluate breakage automatically (*e.g.,* missing legitimate images and text). Thus, we present **AutoFR** [4], a framework based on reinforcement learning (RL). It automatically generates URL-based rules for a website to block ads while avoiding visual breakage. We formulate the problem using multi-arm bandits. The human is now a user that provides AutoFR with a website and a threshold that denotes how much they care about avoiding breakage. The RL agent is tasked with learning which rules effectively block ads without causing visual breakage beyond the given threshold and within a time limit. To implement a scalable tool, we represent how a site is loaded using *site snapshots*, a provenance-based graph of how resources such as JS code, images, and text are loaded. Now, instead of visiting a live site (which is slow), we read the graph into memory and apply the rule to infer its effectiveness (*i.e.,* blocked ads *vs.* images and text). This approach takes an average of 1.6 minutes to generate rules per-site. We use AutoFR on the Top–5K sites and find that it creates rules with comparable performance to the state-of-the-art filter list, EasyList. This work was presented at the Ad-Filtering Dev Summit in 2022.

Characterizing the Advertising and Tracking Ecosystems of Emerging Platforms. Privacy on emerging platforms, such as smart TVs and VR headsets, is not well-understood. For smart TVs, their increasing prevalence in households worldwide provides new opportunities for A&T. VR headsets, such as the Oculus VR by Meta, include a plethora of sensors that detect facial features, eye tracking, and environments, which can lead to exposing sensitive information, such as the user's age, gender, and economic status. To that end, our research involves developing methodologies to measure data collection practices within these emerging platforms to improve transparency for users [9, 10]. We conduct large-scale network measurement studies of smart TVs and the Oculus VR by employing dynamic code instrumentation, binary analysis, and an on-device VPN to collect the network traffic. We extract data types within the network traffic using string matching and regular expressions. We find that both platforms collect a wide variety of PII (*e.g.,* email, serial number), metadata that can be used for fingerprinting (*e.g.,* device OS, Unity version), and platform-specific data (*e.g.,* VR movement, VR play area). To understand whether the data was collected for A&T, we apply state-of-the-art filter lists (that contain domains of A&T) to the network traffic. For smart TVs, the A&T ecosystem is diverse, with organizations like Alphabet, Meta, and comScore. Conversely, we observe that the A&T ecosystem for Oculus VR is developing with a few companies that track users for social and analytics purposes. Our findings have been presented to the FTC in 2021 and 2022, Consumer Reports, and Duck Duck Go.

Future Directions

My future research continues the mission for the transparency, auditability, and control of data collection for users. I will expand my research scope to study privacy in emerging platforms, to foster human-AI interactions and trust for PETs, and to develop scalable methodologies to detect and control dark patterns.

1. Privacy in Emerging Platforms

Privacy for Extended Reality (XR) Platforms. XR platforms (*e.g.,* augmented, virtual, and mixed reality) introduce immersive experiences for users. However, they include new sensors that can learn and expose sensitive information about users, such as their age, gender, and economic status. I will leverage my expertise from the web, smart TVs, and Oculus VR to study privacy for XR. First, I will conduct network traffic measurements to examine the data types that are explicitly exfiltrated (*e.g.,* email, device ID) or ones that can be inferred (*e.g.,* the user's interest based on eye tracking) by XR apps and platforms [9]. Second, I will use the collected network traffic to audit its compliance with privacy policies and right-to-know disclosures (from companies) by using consistency analysis and natural language processing (NLP) approaches. Third, taking a user-centric approach, I will design and implement personalized privacy assistants for XR platforms to facilitate users in managing their privacy controls. For example, at the permission level, the assistant will help users semi-automatically configure app permissions at different granularities (*e.g.,* per-app, per-app category). More interestingly, I will explore privacy assistants for virtual reality. User interactions across different virtual worlds are analogous to mobility traces, which reveal sensitive information about the user. There are opportunities to design new privacy controls that are specific to the virtual world. Fourth, immersive experiences in XR can influence user behavior and emotions more easily, this is especially relevant to advertisements. Thus, I plan to investigate how XR advertising can harm users (*e.g.,* manipulate user buying behavior, manipulate emotions to spread misinformation) and work with committees, such as the Acceptable Ads Committee [1], to mitigate these potential harms by designing XR ad standards.

Robust and Scalable PETs. Mobile, smart TVs, and XR platforms are increasingly popular — mobile is even more popular than the web. On these platforms, users have little control over their privacy. They can resort to DNS-based blocking solutions, such as Pi-hole, which rely on human-created hostname-based filter lists to block A&T. However, these filter lists are often general purpose and can cause functionality breakage, *e.g.,* breaking smart TV apps [10]. There is a clear need to provide better PETs that are specialized for different platforms beyond the web. To that end, I will explore possible ways for users to protect their privacy through developing PETs by employing blocking and obfuscation approaches. For blocking techniques, I will extend the AutoFR framework [4] to create filter rules curated for different platforms that block A&T while considering both visual and functionality breakage. For obfuscation techniques, I will apply them to virtual reality, such as randomly visiting virtual worlds to hide user interactions across different worlds, or similarly, randomizing the avatar creation process per session to obfuscate user identity.

2. Human-AI Interaction for PETs

User Control of PETs. Users commonly install PETs that are created and maintained by companies with business agendas — some of which are not in the best interest of users. For instance, adblockers may whitelist ads from companies that have paid to be on that whitelist, or VPNs may not notify the user when the user's traffic is leaked. My research will establish PETs that are aligned with user interest.

We will utilize a crowd-sourcing framework. For example, users provide their data in a privacy-preserving and automated manner, while trusted parties, such as researchers and nonprofit consumer agencies, use that data to maintain and update the efficacy of PETs. For full transparency, these PETs would be open-sourced. One possible implementation is to leverage AutoFR, which is already an open-sourced framework that can automatically create filter rules for blocking ads. It relies on site snapshots (*e.g.,* graph representations of how a site is loaded) with particular nodes annotated as ads. Users can contribute site snapshots of their visited sites and help annotate ads. A centralized server will collect them and utilize the snapshots to generate filter rules that are then deployed to users. We can employ techniques such as using proxies and encryption to keep the central server from linking the user with the visited site.

Auditability of AI-based Tools by Human Experts. Researchers have built AI-based tools to automate tasks such as detecting and blocking advertising and tracking, which reduces the human effort necessary to maintain PETs. However, they have not been widely adopted by industry due to mistrust of AI-based tools in terms of performance (*e.g.,* will they cause breakage) and maintainability (*e.g.,* how to update them when they are not effective). My research will build frameworks and automated tools coupled with user-friendly companion interfaces that allow humans to understand and audit decisions made by AI. Taking AutoFR as an example. Suppose human experts leverage AutoFR to create filter rules but do not trust its results. In that case, we can design interfaces that display the decision-making process, *e.g.,* showing the history of actions made by the RL algorithm, or exactly the ads, images, and text blocked by the generated rule. Importantly, this entails conducting user studies to understand which kinds of dialogs will be useful to human experts, which will guide the designs and implementation of auditing features. Overall, this work aims to build trust between humans and AI-based tools and push for industry adoption.

3. Dark Patterns

Scaling Detection and Transparency of Dark Patterns. Dark patterns are deceptive user interfaces within online platforms that can manipulate users into making choices that are not in their best interest. For example, it can manipulate users into agreeing with data collection from advertisers and publishers by providing only an "Accept All" button on a consent popup. Researchers have shown that dark patterns are prevalent across desktop and mobile modalities. However, they employ manual efforts to detect dark patterns. I will create methodologies to detect dark patterns to scale this line of research automatically. For the web, we can use NLP approaches to infer whether the language on forms is pushing users to select specific options. For mobile, we can build, *e.g.,* Unity plugins to enact the same actions that we would on the web to mobile apps. In addition, to improve the transparency of dark patterns for users, my research will develop tools that highlight and inform users of dark patterns in real-time (*e.g.,* browser extensions).

User Control of Dark Patterns. Current PETs that deal with dark patterns make choices for users entirely, removing agency from users. In the context of consent popups, a browser extension can automatically close all popups for the user upon visiting a website to reduce annoyances. However, this prevents users from being informed about data collection practices, negating the purpose of the information provided within consent popups. We propose developing privacy assistants that help users manage dark patterns. The assistant can highlight instances of dark patterns and recommend actions. For example, if a checkbox for a consent form is worded to confuse users to opt-in to data collection, the tool can warn users and recommend deselecting the checkbox. It can go a step further and rewrite or remove dark patterns to display neutral user interfaces so that users can make informed choices without dark patterns.

References

[1] Acceptable Ads. Sustainable and non-intrusive advertising. https://acceptableads.com/. (Accessed on 01/27/2022).

[2] eyeo. Ad-Filtering Dev Summit. https://adfilteringdevsummit.com/. (Accessed on 09/01/2022).

[3] eyeo GmbH. Where ad filtering meets profitability. https://eyeo.com/. (Accessed on 10/26/2022).

[4] Hieu Le, Salma Elmalaki, Athina Markopoulou, and Zubair Shafiq. AutoFR: Automated Filter Rule Generation for Adblocking. In *32nd USENIX Security Symposium (USENIX Security 23)*, Anaheim, CA, August 2023. USENIX Association.

[5] Hieu Le, Athina Markoupoulou, and Zubair Shafiq. CV-Inspector: Towards automating detection of adblock circumvention. In *The Network and Distributed System Security Symposium (NDSS)*, February 2021.

[6] PrivacyCon. OVRSeen Presentation 2022. https://www.ftc.gov/media/privacycon-2022-part-2. (Accessed on 10/26/2022).

[7] PrivacyCon. Smart TV Presentation 2021. https://www.ftc.gov/media/73491. (Accessed on 10/26/2022).

[8] ProperData. Privacy and IoT Research Exploration Workshop 2. https://properdata.eng.uci.edu/events/privacyiot-workshop-2/. (Accessed on 10/25/2022).

[9] Rahmadi Trimananda, Hieu Le, Hao Cui, Janice Tran Ho, Anastasia Shuba, and Athina Markopoulou. OVRseen: Auditing network traffic and privacy policies in oculus vr. In *31st USENIX Security Symposium (USENIX Security)*, Boston, MA, August 2022. USENIX Association.

[10] Janus Varmarken, Hieu Le, Anastasia Shuba, Athina Markopoulou, and Zubair Shafiq. The tv is smart and full of trackers: Measuring smart tv advertising and tracking. In *Proceedings on Privacy Enhancing Technologies*, volume 2, pages 129–154. Sciendo, July 2020.