We study Oculus VR (OVR), the leading platform in the VR space, and we provide the first comprehensive analysis of personal data exposed by OVR apps and the platform itself. We developed OVRseen, a methodology and system for collecting, analyzing, and comparing network traffic and privacy policies on OVR, as well as for checking the consistency between the two. On the networking side, we captured and decrypted network traffic of VR apps, which was previously not possible on OVR, and we extracted data flows, each defined as \(\langle\text{app/platform}, \text{data type}, \text{destination}\rangle\). Compared to the mobile and other app ecosystems, we found OVR to be more centralized, and driven by tracking and analytics, rather than by third-party advertising. We show that the data types exposed by VR apps include personally identifiable information (PII), device information that can be used for fingerprinting, and VR-specific data types. Next, by comparing the actual data flows found in the network traffic with the statements made in the apps’ privacy policies, we discovered that approximately 70% of OVR data flows were not properly disclosed. Furthermore, we extracted the purpose of data collection from the privacy policies and found that 69% of data flows were sent for purposes unrelated to core functionality; further, we also found apps that do not provide notice or obtain consent. The additional information about \(\langle\text{consistency, purpose, consent}\rangle\) relates to the transmission principle in the CI framework, and can extend the original \(\langle\text{app/platform}, \text{data type}, \text{destination}\rangle\) tuple, so as to better characterize the appropriateness of these data flows. We believe that our approach and analysis generalizes to other platforms beyond VR, where the CI tuple can be used as the basic conceptual building block for auditing data collection practices using a combination of network traffic and privacy policy analysis.

ACM Reference Format:

1 OVERVIEW

Among VR platforms, Oculus VR (OVR) is a pioneering, and arguably the most popular [6, 9]. VR technology enables a number of applications, including recreational games, physical training, health therapy, and many others [20]. Similarly to the other Internet-based platforms (e.g., mobile phones [4, 5], IoT devices [1, 7], and smart TVs [14, 28]), the Virtual Reality (VR) platform introduces privacy risks and some of these risks are unique to VR devices. For example, VR headsets and hand controllers are equipped with sensors that may collect data about such as user’s physical movement, body characteristics and activity, voice activity, hand tracking, eye tracking, facial expressions, and play area [11, 13, 15], which may in turn reveal information about our physique, emotions, and home.

In our upcoming USENIX Security 2022 paper [26], we present the first large scale, comprehensive measurement and characterization of privacy aspects of OVR apps and platform, using a combined analysis of (1) the network traffic generated by Oculus VR apps and platform, (2) the corresponding privacy policies, and (3) the consistency between network traffic data flows and policy statements. An overview of our approach is depicted on Fig. 1. We characterize and compare how sensitive information is collected and shared in the VR ecosystem, in “theory” (i.e., as stated in the apps’ and platform’s privacy policies) as well as in practice (as exhibited in the actual network traffic generated by the OVR apps and platform).

The Contextual Integrity (CI) tuple is at the core of our analysis. We were able to extract “data flows”, defined as \(\langle\text{app/platform}, \text{data type}, \text{destination}\rangle\) from the network packets sent by the OVR apps and platform. We were also able to extract the same information from the statements in the corresponding privacy policies of the apps and platform. Next, we were able to check the consistency between the two, i.e., whether the actual data flows extracted from network traffic agree with the corresponding statements made in the privacy policy. Furthermore, we were also able to extract the purpose of data collection from the privacy policies: independently, we also extracted the purpose from the network traffic by checking the destination domains, i.e., whether they are advertising-and-tracking services (ATS) domains based on well-known blocklists. Finally, we found that not all of the apps provided a privacy policy, and even when they did, the apps themselves did not always implement “notice and consent”. The additional information about \(\langle\text{consistency, purpose, consent}\rangle\) relates to the transmission principle in the CI framework, and can extend the commonly studied \(\langle\text{app/platform}, \text{data type}, \text{destination}\rangle\) tuple, so as to better characterize the appropriateness of these data flows. The “subject” is typically the user of the device or the app, and can be implicit (denoted as “-”) in the CI tuple.

We believe that our approach and analysis generalize to other platforms beyond VR [26] and their apps, including website, mobile, smart TV, IoT, etc. The CI tuple, \(\langle\text{app/platform}, \text{data type}, \text{destination}, -, \langle\text{consistency, purpose, consent}\rangle\rangle\), can be used as the basic data structure for auditing data collection practices using a combination of network traffic and privacy policy analysis.

2 OVRseen

Our methodology and system, OVRseen, is depicted on Fig. 1. It consists of two parts: network traffic and privacy policy analyses.

Network Traffic Analysis. We experimented with 150 popular, paid and free, OVR apps and we used the best known practices to explore them to generate rich network traffic. We use OVRseen to decrypt and collect network traffic generated by these apps. We then extracted data flows from the collected network packets. First, the sender of information is the VR app or platform. Second, we found 21 data types including personally identifiable information (PII such as device ID, user ID, android ID), device information that can be used for fingerprinting, and VR sensor data (e.g., physical...
movement, play area). Third, w.r.t. the recipient of the information, we extracted the destination domain and we further categorize it into entity or organization, first vs. third party w.r.t. the sending app, and ATS. We find that OVR exposes data primarily to tracking and analytics services, but not to advertising services (yet!); and that existing blocklists block only 36% of these exposures.

**Privacy Policy Analysis.** We use an NLP-based methodology for analyzing the privacy policies that accompany VR apps. OVRseen maps each data flow found in the network traffic to its corresponding data collection statement found in the text of the privacy policy, and checks the consistency of the two. Furthermore, it extracts the purpose of data flows from the privacy policy, as well as from the ATS analysis of destination domains. Consistency, purpose, and additional information (e.g., about notice and consent) relate to the transmission principle, which can help assess the appropriateness of the information flow [16]. Our methodology builds on, combines, and improves state-of-the-art tools originally developed for mobile apps: PolicyLint [2], PoliCheck [3], and Polisis [8]. We curated VR-specific ontologies for data types and entities, guided by both the network traffic and privacy policies. We also interfaced the different NLP models of PoliCheck and Polisis to extract the purpose behind each data flow.

Our network-to-policy consistency analysis reveals that about 70% of data flows from VR apps were not disclosed or consistent with their privacy policies. Furthermore, 38 apps did not have privacy policies, including apps from the official Oculus app store. Many app developers also tend to neglect declaring data collected by the platform and third parties. We also found that 69% of data flows have purposes unrelated to the core functionality, and only a handful of apps are explicit about notice and consent (see Fig. 2).

Due to lack of space, we defer the details of our methodology and results to [26], and the software and datasets to [27].

### 3 WORKS CITED

**Privacy of Various Platforms.** The research community has looked into privacy risks in various platforms, including Android [4, 5, 17, 19, 21–23], smart TVs [14, 28], and IoTs [12, 18]. Our work is the first to perform network traffic analysis on VR.

**Privacy Policy Analysis.** Privacy policy and consistency analysis in various app ecosystems [2, 3, 8, 25, 29–31] is becoming increasingly automated. Wang et al. applied similar techniques to check for privacy leaks from user-entered data in GUI [29]. We leveraged two state-of-the-art tools, namely PoliCheck [3] and Polisis [8], to perform data-flow-to-policy consistency analysis and extract data collection purposes—as detailed in the previous section.

**CI applications.** Most prior work on network traffic analysis characterized only destinations (first vs. third parties, ATS, etc.) and/or data types exposed without additional context for the CI-tuple. One exception is MobiPurpose [10], which inferred data collection purposes of mobile (not VR) apps, using network traffic and app features (e.g., URL paths, app metadata, domain name, etc.). The authors stated that their “purpose interpretation can be subjective and ambiguous,” while we extracted purpose explicitly stated in the privacy policies and/or indicated by the destination domain matching ATS blocklists. Shvartzshnaider et al. applied the CI framework to analyze privacy policies (not network traffic), with a case study of Facebook, and conducted manual inspection (not NLP) [24].

---

**Figure 1:** Overview of OVRseen. On one hand, we experiment with the most popular OVR apps and analyze their network traffic: we extract data flows \( \text{app/platform, data type, destination} \) and analyze them w.r.t. data types and ATS ecosystem. On the other hand, we also analyze the apps’ privacy policies: network-to-policy consistency analysis for each app using PoliCheck [3] and collection purpose extraction from the app’s privacy policy using Polisis [8]. The end result is that data flows, extracted from network traffic, are augmented with additional attributes related to the transmission principle, such as consistency with the statement made in the privacy policy and purpose of collection.

**Figure 2:** We consider the 370 data flows with consistent disclosures between network traffic and privacy policies. We also extract their purpose from the privacy policy and depict the augmented tuples: \( \text{app, data type, destination, purpose} \).
REFERENCES


