

Abstract

By their very nature, keyboards process everything that users input into their devices, including private messages, passwords and payment details. Despite this, their privacy is often overlooked, and in fact, not much is understood about the privacy considerations of even the most popular keyboards available on Android. This dissertation addresses this issue and investigates the data collected and shared with back-end servers by some of the most popular keyboard applications, namely Google's Gboard, Microsoft's SwiftKey and an open-source alternative, AnySoftKeyboard. Providing users with a deeper understanding of the underlying privacy considerations of these keyboards will allow them to make more informed decisions about their consent to the ongoing data collection.

The privacy investigation is performed by collecting the network traffic generated by each keyboard, followed by a significant amount of reverse engineering work required to decrypt and decode the data, whose content is then analysed for the presence of sensitive information.

It is found that the two proprietary keyboard applications both collect and share large amounts of telemetry data. Gboard provides an opt-out from this data collection; however, SwiftKey does not. The data logged by these keyboards is largely similar. It includes timestamped log entries containing specific hardware information, the length of individual words entered, languages used, and the application's name in which the keyboard was opened. Neither keyboard has been observed to collect or share the input content or even track the frequency of individual characters. Most notably, the logs of both keyboards include unique identifiers, namely the Android ID and the Google Advertising ID, corresponding to Gboard's and SwiftKey's logs, respectively. This allows the data to be linked to a specific handset and potentially a user's identity, putting their anonymity at risk. Additionally, it is possible to infer certain personality traits about users based on the timestamped application usage information found in both keyboards' logs. It is also found that the open-source AnySoftKeyboard lives up to its reputation as a privacy respecting keyboard. It is not observed to collect or share any telemetry at all, proving that excessive telemetry collection is indeed a choice made by developers and not a requirement.