Analyzing Android Encrypted Network Traffic to Identify User Actions

Mauro Conti, Senior Member, IEEE, Luigi Vincenzo Mancini, Riccardo Spolaor, and Nino Vincenzo Verde

Structural unit	Move	Corresponding Sentences
Introduction section	**Move 1**	**The writers highlight the interest of their work.**
	Step 1	THE amount of sensitive data that users handle with their mobile devices is truly staggering. People continuously carry these devices with them and use them for daily communication activities, including not only voice calls and SMS, but also emails and social network interactions. A typical user gains access to her savings and checking account by using her smartphone. She installs and uses several apps to communicate with friends or acquaintances. Through her smartphone, she gets information about sensitive topics such as diseases, sexual or religious preferences, etc.
	Step 2	**Describing what is known about the research topic.** As a consequence, several concerns have been raised about the capabilities of these portable devices to invade the privacy of users actually becoming “tracking devices”. In this context, an important aspect is related to the possibility of

		continuously spying and locating an individual [3], [32], [35].
	Step 3	**Reviewing items of previews research to reinforce the importance of their research.** Solutions to identify and isolate malicious code running on smartphones [31], [37], [42] as well as to protect against attacks coming from the network [4], [11] might significantly reduce current threats to user privacy. While people become more familiar with mobile technologies and their related privacy threats (also thanks to the attention raised by the media, e.g., see the recent attention on NSA for supposedly eavesdropping foreign governments leaders such as Angela Merkel [35]), users have started adopting good practices that better adapt to their privacy feeling and understanding. However, many mobile apps use the Secure Sockets Layer (SSL) – and its successor Transport Layer Security (TLS) – as a building block for encrypted communications.
	Move 2 **Step1A+** **Step1C**	**Establishing a niche** **Criticism of the weak points of any previous work** **indicating the possible gaps regarding previous work** Unfortunately, we believe that even adopting such good practices would not

		close the door to malicious adversaries willing to trace people. Indeed, several attacks may violate the privacy of the user even when the adversary does not physically or remotely control the user device. In this paper, we consider a passive attacker that is able to sniff the network traffic of the devices from the network side. Obviously, if the network traffic is not encrypted, the task of such an attacker is simple: he can analyze the payload and read the content of each packet. Even when such solutions are in place, the adversary can still infer a significant amount of information from the analysis of the properly encrypted network traffic. For example, work leveraging analysis of encrypted traffic already highlighted the possibility of understanding the apps a user has installed on her device [36], or identify the presence of a specific user within a network [38].
	Step 1D	**presenting writer's work as a continuation of previous research topic** This work focuses on understanding whether the user profiling made through analyzing encrypted traffic can be enhanced to understand exactly what actions the user is doing on her phone: as concrete examples, we aim at identifying actions such as the user sending an email,

		receiving an email, browsing someone profile on a social network, publishing a post or a tweet. The underlying issue we leverage in our work is that SSL and TLS protect the content of a packet, while they do not prevent the detection of networks packets patterns that instead may reveal some sensitive information about the user behavior.
	Move 3 **Step 1B** **Step 1C**	**Description of the main feature of the study** In this paper (which is an extended version of the work in [12]), we propose a framework to infer which particular actions the user executes on some app installed on her mobile-phone. In particular, we assume that the traffic is encrypted and the adversary eavesdrops (without modifying them) the messages exchanged between the user's device and the web services that she uses. **Announcing the main findings** Our framework analyzes the network communications and leverages information available in TCP/IP packets (like IP addresses and ports), together with other information like the size, the direction (incoming/outgoing), and the timing. By using an approach based on machine learning, each app that is of interest is analyzed independently. To set up our

		system, for each app we first pre-process a dataset of network packets labeled with the user actions that originated them, we cluster them in flow typologies that represent recurrent network flows, and finally we analyze them in order to create a training set that will be used to feed a classifier. The trained classifier will then be able to classify new traffic traces that have never been seen before. We run a thorough set of experiments to evaluate our solution considering seven popular apps: Facebook, Gmail, Twitter, Tumblr, Dropbox, Google+ and Evernote. The results show that it can achieve accuracy and precision higher than 95%, for most of the considered actions. In the current version of the paper, we also add a discussion (not present in [12]) about the key idea underneath our traffic analysis approach. In particular, we examine in depth the concept of network flow and the metric to evaluate the similarity between them. We also report details of the machine learning techniques we leverage in our method. Furthermore, in addition to our previous work [12], we run a thorough comparison of our solution with three state of the art algorithms, showing that our solution outperforms them in all of the cases.
	Step 1D	**Indicating the RA structure**

		Organization: The rest of this paper is organized as follows. In Section II, we revise the state of the art around our research topic. In Section III, we introduce some background knowledge on machine learning and data mining tools used in our work. In Section IV, we present our framework describing all its different components. We present the evaluation of our solution for identifying user actions in Section V, where we compare with similar solutions as well. In Section VI, we discuss about possible countermeasures against the proposed attack. Finally, in Section VII we draw some conclusions and point out ways in which this work can be further extended.