



Politecnico
di Torino

ScuDo

Scuola di Dottorato - Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Electrical, Electronics and Communications Engineering
(36th cycle)

Privacy on the Web: Algorithms, Tools and Measurements

By

Nikhil Jha

Supervisor(s):

Prof. Marco Mellia, Ph.D., Supervisor
Dr. Martino Trevisan, Ph.D., Co-Supervisor

Politecnico di Torino
2024

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Nikhil Jha
2024

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Privacy on the Web: Algorithms, Tools and Measurements

Nikhil Jha

The continuous and ubiquitous collection and exchange of data is at the foundation of the current Web ecosystem. Many of these data may be of sensible kind, being extracted by various means from users' online activity — often with limited awareness of the process by the users themselves. Users' data are useful for a variety of online and offline entities: advertisers can exploit knowledge on users' preferences to show them advertisements tailored on their needs on the webpages they visit, marketers collect users' data to find valuable information for their business activities. One of the most prominent ways to collect users' data are third-parties cookies. They are small text files installed on users' browsers by entities contacted on the Web, containing identifiers that allow third-parties to follow the users along their navigation of the Web.

The push of the industry towards an ever-increasing amount of collected data collides with the right to privacy that should be guaranteed to Web users: collected data can be used to infer private information about the users, by cross-checking obtained data with other, public sources. In the past years, legislators have tried to give users a larger control over the data that are extracted from their use of the Web. This has led to the proliferation of Privacy Banners, that inform the users of the agents and the purposes of the collection.

In this dissertation, we will discuss various aspects on implementing and measuring privacy on the Web: we will start from the role of Privacy Banners in the current Web ecosystem. We will study how users interact with the Banners, and how crawling techniques that aim at taking measures of key metrics in the Web must take into consideration Privacy Banners in order for their empirical estimates to be accurate and close to real-world experience. Moreover, we will discuss the Topics API, a possible solution that goes beyond third-party cookies, in an effort to re-balance the trade-off between data utility and data privacy.

Finally, we will also introduce a study on the privacy properties of z -anonymity, a data anonymization property and algorithm suited for streaming data anonymization.

We compare it with the well-known k -anonymity property, and evaluate the utility loss needed to obtain desired levels of privacy.