COMPUTER AND CONTROL ENGINEERING

PNRR/SERICS - Protect-IT – Privacy on Internet: measurements and novel approaches

Funded By	MINISTERO DELL'UNIVERSITA' E DELLA RICERCA [P.iva/CF:97429780584] Politecnico di TORINO [P.iva/CF:00518460019]
Supervisor	MELLIA MARCO - marco.mellia@polito.it
Contact	MELLIA MARCO - marco.mellia@polito.it VASSIO LUCA - luca.vassio@polito.it
Context of the research activity	The project aims to investigate privacy issues on the Internet, and on possible alternatives for behavioural advertising on the web. Via passive and active data collection, we aim to define policies and mechanisms to quantify and monitor the amount of information services collect, and to define novel policies to balance it with privacy requirements for users. Progetto finanziato nell'ambito del PNRR M4C2, Investimento 1.3 - Avviso n. 341 del 15/03/2022 - PE0000014 Security and Rights in the CyberSpace (SERICS) - CUP E13C22001850001
	In the current web ecosystem, targeted or behavioural advertising lets providers monetize their content, by collecting and processing personal data to build accurate user profiles. This massive data collection has created tension between users and the ads ecosystem. Mozilla Firefox and Apple Safari have started battling third-party cookies by giving third-party cookies a separate cookie jar per site, so they cannot be used to track users across sites anymore. Other alternatives have been proposed to limit and control the amount of information users share with web services. This project consists of the design, implementation and testing of novel measurement methodologies to observe and quantify the amount of information online services collect, and on the definition of possible alternative solutions that result in more privacy friendly. The research will focus on the design of data collection methodologies that allow one to observe and quantify the amount of information web services collect when users browse the web. This includes both active measurements, e.g., web crawlers, and passive measurements, e.g., browser

Objectives	plugins that observe the users' browsing activity. We will investigate mechanisms that will offer privacy guarantees to users, e.g., via the extraction of client-side models that process the information locally rather than sharing the raw data with servers. At the same time, we will investigate the introduction of privacy guarantees, e.g., based on differential privacy or k-anonymity concepts. In the first phase, the candidate will study the state of the art of privacy methodologies in general and of the problem of data collection on the internet. He/she will design possible data collection methodologies and metrics to measure and quantify the current state. In the second phase, the candidate will design and integrate possible alternative solutions to mitigate and control the amount of information web services collect from users. These shall include privacy guarantees for example those provided by differential privacy or k-anonymity approaches. The project will involve a collaboration with partners in the Protect-IT project including the University of Brescia, of Naples, of Milano among others. -- Outline of the research work plan 1st year - Study of the state-of-the-art on privacy and the web, initial design of the data collection platforms and machine learning models to extract and quantify the amount of personal data exchanged with web services. 2nd year - Data collection campaign and measurement analysis - Propose and develop innovative solutions to the problems of privacy on the web. - Propose machine learning approaches to offer privacy guarantees and extract local models to avoid the sharing of raw data. 3rd year - Tune the developed techniques and highlight possible strategies to deploy them in the wild. - Application of the strategies to new data for validation and testing. -- References: - Nikhil Jha, Martino Trevisan, Luca Vassio, and Marco Mellia. 2022. The Internet with Privacy Policies: Measuring The Web Upon Consent. ACM Trans. Web 16, 3, Article 15 (August 2022), 24 pages. https://doi.org/10.1145/3555352 - N. Jha, M. Trevisan, M. Mellia, R. Irarrazaval and D. Fernandez, "I Refuse if You Let Me: Studying User Behavior with Privacy Banners at Scale," 2023 7th Network Traffic Measurement and Analysis Conference (TMA), Naples, Italy, 2023, pp. 1-9, doi: 10.23919/TMA58422.2023.10198936. - Nikhil Jha, Martino Trevisan, Emilio Leonardi, and Marco Mellia. 2024. Re-Identification Attacks against the Topics API. ACM Trans. Web Just Accepted (June 2024). https://doi.org/10.1145/3675400 -- List of possible venues for publications Security venues: IEEE Symposium on Security and Privacy, IEEE Transactions on Information Forensics and Security, ACM Symposium on

	Computer and Communications Security (CCS), USENIX Security Symposium, IEEE Security & Privacy; AI venues: Neural Information Processing Systems (NeurIPS), International Conference on Learning Representations (ICLR), International Conference on Machine Learning (ICML), AAAI Conference on Artificial Intelligence, ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD), European Conference on Machine Learning and Knowledge Discovery in Databases (ECML/PKDD); Computer networks venues: Distributed System Security Symposium (NDSS), Privacy Enhancing Technologies Symposium, The Web Conference (formerly International World Wide Web Conference WWW), ACM International Conference on Emerging Networking EXperiments and Technologies (CoNEXT), ACM Internet Measurement Conference, USENIX Symposium on Networked Systems Design and Implementation (NSDI);

Skills and competencies for the development of the activity	- Solid programming skills (such as Python, Torch, Spark) - Excellent Machine Learning knowledge - Knowledge privacy preserving analytics - Knowledge of Networking and security