Internship proposition: Assessing the trustworthiness of online identities

Keywords: security, privacy, trust, data collection and analysis, machine learning, inference, statistics, Sybils, fake news
Lab: Laboratoire d'Informatique de Grenoble (LIG), Grenoble, France (head: Eric Gaussier)
Team in the lab: SLIDE (head: Sihem Amer-Yahia)
Advisor: Oana Goga (CNRS & Univ. Grenoble Alpes, LIG) oana.goga@mpi-sws.org https://people.mpi-sws.org/~ogoga/

Project Description:
The lack of *strong identities*, i.e., secure and verifiable identities that are backed by a certificate from a trusted authority (e.g., passport or social security number) has been a long-standing problem on the Web. While strong identities could provide better security for Web services they failed to achieve mass adoption because they significantly raise the sign-on barrier for new users and raise privacy concerns – users cannot be anonymous. Consequently, most Web services today provide *weak identities* that are not backed by any certificate. Unfortunately, weak identities are heavily exploited by malefactors to create multiple fake identities with ease. Such fake identities are traditionally known as Sybils [1] and are typically used to inundate services with spam, spread fake news, or post fake reviews. Not only do these problems impact our daily activities, but they also deeply affect the economies and political life of our countries. The goal of this project is to investigate methods to assess the trustworthiness of online identities and content (e.g., *check whether a particular restaurant review is fake or real* or *whether an individual is interacting with a legitimate user or an attacker*).

More precisely the project consists in investigating whether we can quantify the trustworthiness of online identities in monetary terms using the price of identities in black markets [2]. Popular black-market services like Freelancer and Taobao [6] allow job postings promising creation of fake identities with different levels of grooming. With the availability of such job posting data, the idea is very simple: Given that the advertised black-market price for an identity groomed to a certain level is X, and that the expected utility that can derived from the activities of this identity is Y , if X > Y , we can expect that a rational attacker will have no incentive to use such an identity for malicious activities. The second step is to measure the extent to which we can increase the accountability of weak identities if we link the (potentially anonymous) identities of users across multiple systems.

Throughout the project the student will be able to familiarize himself with the different ways online systems can be exploited by attacker and possible countermeasures, learn to collect data from online services and apply and analyze this data.

Requirements:
Strong coding skills**.** Experience in working with data is a plus.

References:
[1] John R. Douceur. The sybil attack. In *IPTPS,* 2002.
[2] Gang Wang, Manish Mohanlal, Christo Wilson, Xiao Wang, Miriam Metzger, Haitao Zheng, and Ben Y. Zhao. Social turing tests: Crowdsourcing sybil detection. In *NDSS*, 2013.
[2] AnsleyPost,VijitShah,andAlanMislove.Bazaar:Strengtheninguserreputationsinonlinemarketplaces.In *NSDI*, 2011.
[3] Kurt Thomas, Damon McCoy, Chris Grier, Alek Kolcz, and Vern Paxson. Trafficking fraudulent accounts: The role of the underground market in twitter spam and abuse. In *USENIX Security,* 2013.
[4] Haifeng Yu, Michael Kaminsky, Phillip B. Gibbons, and Abraham Flaxman. Sybilguard: Defending against sybil attacks via social networks. In *SIGCOMM*, 2006.