



Personal Data Leakage and Other Privacy Violations in the Era of Big Data Mining

Nicolas Kourtellis, Ph.D.

Telefonica R&D, Barcelona, Spain

Tuesday, December 18, 2:00 PM, ENB 313

Abstract: In this talk, I will review our recent efforts to study the leakage of online users' private data on the web via different advertising protocols such as the real-time bidding and cookie synchronization. Extracted from a large corpus of mobile user web activity, several findings will be presented, including what type of personal data are exposed, entities involved, factors influencing the intensity of leakage, etc. Based on these studies, our recent efforts to improve online user awareness for such privacy leakages with practical web tools will also be covered. Furthermore, an analysis of a large dataset of network activity logs of mobile users will be presented, in an effort to assess the potential of identity leakage of mobile users. In particular, spatio-temporal fingerprints are built on the network activity data, which can then be used to de-anonymize network users in time and space. We discuss factors influencing the potential leak and ways to reduce it.

Bio: Dr. Nicolas Kourtellis is a Researcher in the Telefonica R&D team, in Barcelona. Previously he was a Researcher in the Web Mining Research Group at Yahoo Labs, in Barcelona. He holds a Ph.D. in Computer Science and Engineering from the University of South Florida (2012), a MSc in Computer Science from the University of South Florida (2008), and a BSc in Electrical and Computer Engineering from the National Technical University of Athens, Greece (2006). His primary interests lie in the following dimensions: 1) user online privacy, leakage of personal data to the online advertising ecosystem and other entities, 2) analysis and characterization of online user behavior, with respect to different dimensions such as: abusive, hateful, aggressive and bullying behavior, fake news propagation, fringe online communities, etc., 3) system design for streaming data analysis and graph mining on distributed streaming processing engines. He has published more than 60 papers, and presented his work in top academic conferences and journals such as IEEE TKDE, IEEE TPDS, IEEE ICDE, ACM KDD, ACM WWW, ACM IMC, ACM/IFIP/USENIX Middleware, etc., as well as industry-oriented conferences such as Apache BigData in Europe and N. America. He has served in many program committees of top conferences and journals (e.g., WWW, KDD, CIKM, ACM TKDD, IEEE TKDE, IEEE TPDS, etc.).