

Einladung des Instituts für Technische Informatik  
zum **GASTVORTRAG am 18.12.2019 um 10:00 Uhr**  
im NXP Semiconductors Seminarraum, IE01090, Inffeldgasse 16/I, 8010 Graz

---

Title : **On the limits of Anonymizing Web Tracking Data by Generalization**  
Lecturer : **Prof. Thorsten STRUFE, KIT-Department of Informatics,  
Germany**

When : **10:00, December 18, 2019**

Where : NXP Semiconductors Lecture room (Inffeldgasse 16, 1st floor)

**Abstract:**

Cross domain tracking has become the rule, rather than the exception, and scripts that collect behavioral data from visitors across sites have become ubiquitous on the Web. The collections form comprehensive profiles of browsing patterns and contain personal, sensitive information. This data can easily be linked back to the tracked individuals, most of whom are likely unaware of this information's mere existence, let alone its perpetual storage and processing. As public pressure has increased, tracking companies like Google, Facebook, or Baidu now claim to anonymize their datasets, thus limiting or eliminating the possibility of linking it back to data subjects.

In cooperation with Europe's largest audience measurement association we use access to a comprehensive tracking dataset to assess both identifiability and the possibility of convincingly anonymizing such browsing data. Our results show that anonymization through generalization does not sufficiently protect anonymity. Reducing unicity of browsing data to negligible levels would necessitate removal of all client and web domain information as well as click timings. In tangible adversary scenarios, supposedly anonymized datasets are highly vulnerable to dataset enrichment and shoulder surfing adversaries, with almost half of all browsing sessions being identified by just two observations. We conclude that while it may be possible to store single coarsened clicks anonymously, any collection of higher complexity will contain large amounts of pseudonymous data.