Analysis without Tears

The authors explain the new needs for policy guidelines and techniques to maintain privacy when analyzing users' data in this study. Personal data obtained not only from social media and internet services, but also from sensors in smart homes, automobiles, cities, and health gadgets, is becoming increasingly invasive as digital monitoring expands. Something will have to give. There are both technological and socioeconomic reasons for this to change, and authorities and industry have acknowledged this. Failures in privacy endanger personal and corporate wealth and safety. Theft of credit card information, identity, and trade secrets is a real and present concern, and it is becoming more so as the surveillance society's "attack surface" expands. The amount of detail accessible allows for inference about persons and institutions in ways that are neither acknowledged nor necessarily intended.

The authors state that the overarching objective of research into privacy-preserving data analysis is to create strategies that extract the most usefulness from a dataset while respecting the privacy of the persons represented in it. However, there are different views of what privacy means in this circumstance. First and foremost, one must recognize that if you are a member of a certain population and an analysis of that group is made public, your privacy has been violated and there is nothing you can do — or could have done — about it. However, there are other critical privacy challenges in data analysis that technology may assist in addressing. Two elements for which we have in theory sufficient technical solutions are privacy of stored data, i.e., encryption of data at rest (on disk), and privacy of data while it is transferred, i.e., encryption of data in transit.

We may utilize numerous strategies to compute on private data owned by mutually untrusted parties, including secure enclaves based on a Trusted Execution Environment, homomorphic encryption, multiparty computation, edge computing, and customised systems that combine several of the techniques. Finally, while the approaches described above may be used to construct a statistical model in a privacy-preserving manner, that is, without releasing any unneeded information, they do not solve the issue of measuring how much information is released by such a model.