Abstract

Privacy is a problem. Vast amounts of data about individuals is constantly accumulating in various databases— patient records, content and link graphs of social networks, mobility traces in cellular networks, book and movie ratings, etc.—and there are many socially valuable uses to which it can potentially be put. But, as Netflix and others have discovered [3, 22], even when data collectors try to protect the privacy of their customers by releasing anonymized or aggregated data, this data often reveals much more than intended, especially when it is combined with other data sources. To reliably prevent such privacy violations, we need to replace current adhoc solutions with a principled data release mechanism that offers strong, provable privacy guarantees

privacy allows us to reason formally about what an adversary could learn from released data, while avoiding many assumptions (e.g., what exactly the adversary might try to learn, or what he or she might already know) that have been the cause of privacy violations in the past. Early work on differentially private data analysis relied on manual proofs by privacy experts that the answers to particular queries were safe to release