Research Statement

Mohammad Nabil ALAGGAN, Ph.D.

December 22, 2014

I specialize in privacy and security with a special focus on peer to peer systems. Peer to peer systems constitute an appealing alternative in a world where scalability and privacy matter increasingly every day, as they give users the power to host their own data and collaborate to perform their own computation. Such systems, however, have different security and privacy aspects than centralized systems and require a study of their own.

While privacy and security is a huge topic, my contribution to it is mainly by studying differential privacy applications in interactive and non-interactive settings. Recent results are given as follows.

Interactive Private Similarity Protocol We studied a secure multi-party protocol for the computation of pairwise similarity, that does not reveal the similarity value if it is below a certain threshold. The protocol was designed in a way that addresses the unique set of challenges faced in a large scale dynamic P2P system. For instance, we addressed the privacy budget problem, which otherwise would have imposed a serious restriction on dynamic peer-to-peer systems, by setting a fixed upper bound on the number of interactions for similarity computations a peer can have by proposing a novel bidirectional anonymous channel, which prevents passive and malicious adversaries from linking different observations to each other, as it is not desired to prevent them from collection an unbounded number of observations. We also propose several methods for distributed noise generation, that are better suited for the kinds of random numbers needed, and are resilient to manipulation by one of the two parties.

Heterogeneous Differential Privacy Most of the proposed approaches to preserve privacy in personalization systems usually address the issue uniformly across users, ignoring the fact that users have different privacy attitudes and expectations (even among their own personal data). We propose to account for this non-uniformity of privacy expectations by introducing the novel concept of heterogeneous differential privacy. This notion captures both the variation of privacy expectations among users as well as across different pieces of information related to the same user. We also describe an explicit mechanism achieving heterogeneous differential privacy, which is a modification of the Laplacian mechanism by Dwork, McSherry, Nissim, and Smith. The basic idea underlying the mechanism is manipulating the sensitivity of the function using a linear transformation of the input domain. We prove that our mechanism protects the user's private items as well as the privacy vector representing his privacy expectations across all of his items.

Non-Interactive Private Similarity Protocol The need for a non-interactive protocol arises primarily from the privacy budget issue, but also serves other purposes such as computational efficiency and serving offline users. In particular, Non-interactive protocols are not subject to the privacy budget issue since they are computed only once. Moreover, a non-interactive protocol avoids the need to use cryptographic tools, allowing for more efficient execution and small communication cost; as its output may be cached. Another advantage of this non-interactive mechanism is that similarity computation may take place even when the user is offline.

For this purpose We introduce a novel privacy mechanism called BLIP (for BLoom-and-flIP). In brief, the profile of a user will be represented in a compact way, as a Bloom filter (a small probabilistic set data structure composed of a vector of bits) that will be perturbed through flipping some bits at random. The main objective is to privately estimate the similarity between two users using this perturbed Bloom filter representation. We showed that the utility of this mechanism is optimal.

BLIP also lifts the needs for the bidirectional anonymous channel, which strengthens the robustness of the peer-to-peer network. For instance, peers can log off and then log back in without worrying about losing connections to their personalized but anonymous neighbors. Moreover, peers can computer their similarity with other peers who are offline as long as their BLIPed profiles exist in the network.

The Meaning of the Privacy Parameter We analyzed of the protection offered by BLIP is provided with the objective of deriving an upper and lower bound for the value of the differential privacy parameter ε, for which it is difficult to grasp an intuition for. More specifically, we define a probabilistic inference attack, called the "Profile Reconstruction Attack", that can be used to reconstruct the profile of an individual from his perturbed Bloom filter representation, along with the "Profile Distinguishing Game" which measures whether an adversary can distinguish

a change in one item. An upper bound for ε is a value that makes one of these attacks succeed. The lower bound is both given by a theoretical bound on the deviation from the true answer, and empirically by finding the values of ε which provide a poor utility for a particular application.

The Dwork-Naor impossibility result (Dwork and Naor, 2010), states that for *any* privacy-preserving mechanism, if we do not restrict the auxiliary information accessible to the adversary, then a privacy breach (of size equal to the min-entropy of the utility) always occurs. Not only this motivated the notion of differential privacy, it also highlighted that the first step to understanding privacy (as lack of privacy breaches) is to understand what auxiliary information is accessible to an actual adversary and what the adversary can do with it. In this spirit, we investigated the meaning and implications of the value of the privacy parameter ε. In particular, we investigated three models for auxiliary information. The first model assumed no auxiliary information and is used a baseline (the profile reconstruction attack), while the second considered the knowledge of all the profile except one item (the profile distinguishing game), and the last one considered the correlations between bits while not knowing the value of any bit.

Future Directions

Non-interactive differential privacy is appealing because we do not have to worry about the privacy budget or multiple releases, which is a deep concern in peer-to-peer or distributed systems. A related mechanism–randomized response, is also appealing in large scale systems because each individual party can release its data in a private fashion without need for interaction with other parties. Unfortunately, ε-differential privacy in the non-interactive or randomized response model *must* introduce a large amount of noise (McGregor, Mironov, Pitassi, Reingold, Talwar and Vadhan, 2010). However, there are many relaxations to differential privacy that introduce less noise. Understanding the privacy implications of these relaxations and their applicability in large scale distributed systems is a promising research direction that I plan to pursue.

Another related research direction is to investigate privacy-preserving *mergeable* sketches for large scale data private data analysis applications, like online mobility monitoring, or traffic analysis. We have already made partial progress towards that target in an ongoing work, which investigates merging two flipped Bloom filters into a third one, representing the intersection of their underlying sets, with the goal of privacy-preserving traffic monitoring in mind.