Poirot: Private Contact Summary Aggregation

Johes Bater[1], David Pujol[1], Chenghong Wang[1], Yanping Zhang[1], Matthew Lentz[1,2], Ashwin Machanavajjhala[1], Kartik Nayak[1], Lavanya Vasudevan[3,4], and Jun Yang[1]

[1]Department of Computer Science, Duke University
[2]VMware Research
[3]Family Medicine and Community Health, Duke University
[4]Duke Global Health Institute

COVID-19 is a contagious respiratory disease that is known to spread rapidly through person-to-person contact. It has resulted in 200K deaths (as of September 2020) in the US alone. One of the key measures for curbing the spread of COVID-19 is physical distancing between individuals. During the initial phase of the spread, many state governments announced lockdowns, and offices and schools were temporarily shut down. While these measures helped stabilize the infection rate, subsequent repeals (or weakening) of these measures resulted in an increased infection rate once again. Organizations, cities, and municipalities that want to reopen but they need data about the adherence to physical distancing to inform their decisions on measures and interventions to stop the disease's spread. However, the necessary data to track physical distancing are very sensitive as they would reveal the location trajectories of individuals and their social interactions with others. Thus, there is a tension between quantifying physical distancing in populations and ensuring individuals' privacy.

The goal of our work, Poirot, is to collect necessary information about individuals' physical interactions in a privacy-preserving manner to provide actionable information to decision-makers and the individuals themselves. Our work measures physical interactions through the number of contact events between individuals; such measurements are directly related to the disease's spread and thus a crucial enabler to decision-making. To ensure privacy, we collect aggregate statistics of contact events and release them so that it cannot be linked back to any individual.

Here's how Poirot works at a high level: (1) users install the app on their smartphones, (2) the app detects contact events via a Bluetooth protocol, (3) each day, the app computes aggregate summaries of the contact events and uploads secret-shared (encrypted) versions to a collection of servers, (4) the servers compute aggregate statistics over all users (using secure multiparty computation and differential privacy to guard the privacy of individual users). We emphasize that the goal of Poirot is different from that of an automated contact tracing system: while the latter focuses on finding contacts of infected individuals, Poirot is only concerned with the number of contact events independent of who the contacts are or their COVID status.

How is the information collected by Poirot useful? Universities, companies, counties, and states have introduced policies to open schools and businesses at a reduced scale. Ideally, they aim to minimize the number of contact events while still allowing some people to be at school/work. The aggregate contact statistics not only help assess the effectiveness of current policies but also serve as a feedback loop to inform new policies. By breaking down these statistics based on time, location, and broad individual attributes (e.g., essential worker), decision makers can introduce smarter, more targeted policies that offer more flexibility without sacrificing safety. For instance, such statistics can help determine office locations, cafeterias, and other public places that can be a "hotspot" for some days or some hours during the day. With this

information, we can better schedule the number of people visiting a given location at a given time, and better allocate sanitation resources and efforts. At the same time, the collection of aggregate statistics is also useful to individuals in helping them better understand and improve how they are adhering to social distancing policies. For example, individuals can compare their daily contacts to their personal history, to those of an average user (of a given category), or even to the targets set by public health officials

What kind of privacy guarantees do we provide? Our goal is to learn aggregate summaries of contact events between end users. These contact events are aggregated by the computing servers and eventually revealed to the decision makers (and the end users themselves).

A key goal of Poirot is to protect the privacy of end users while still learning population-level statistics. When users share summaries of their contact events with multiple computing servers, these values are anonymized and encrypted (or secret-shared). The encryption guarantees that no subset of servers can learn any individual's contact summaries. The servers then engage in a secure multiparty computation protocol to compute aggregate statistics. We further apply differential privacy to ensure that these statistics themselves do not disclose information about an individual (e.g., contact event in a location where very few individuals visit). Thus, an end user contributes valuable and actionable information without sacrificing their privacy.

Summary. We believe Poirot will be an effective solution to stem the spread of contagious diseases such as COVID-19 that are transmitted through person-to-person interaction. By measuring the degree of physical distancing between individuals, it provides actionable information to both decision-makers and individuals while preserving the privacy of all individuals involved.