Privacy-Preserving Contact Tracing of COVID-19 Patients

Leonie Reichert
Department of Computer Science
Humboldt University
Berlin, Germany
leonie.reichert@informatik.hu-berlin.de

Samuel Brack
Department of Computer Science
Humboldt University
Berlin, Germany
samuel.brack@informatik.hu-berlin.de

Björn Scheuermann
Alexander von Humboldt Institute
for Internet and Society
Berlin, Germany
bjoern.scheuermann@hiig.de

***Index Terms*—Secure Multiparty Computation, Contact Tracing, Privacy Enhancing Technologies, Health Data**

I. Introduction

The current COVID-19 pandemic shows that our modern globalized world can be heavily affected by a quickly spreading, highly infectious, deadly virus in a matter of weeks. It became apparent that manual contact tracing and quarantining of suspects can only be effective in the first days of the spread before the exponential growth overwhelms the health authorities (HA). Shutdowns of entire countries thus are a popular and drastic method to slow down infection rates in order to not overwhelm emergency capacities. While such shutdowns are effective, they also severely impact social and economical routines in the affected areas.

By automating tracing processes and quarantining everyone who came in contact with infected people, as well as arriving travelers, it should be possible to quickly loosen lockdown measures. Countries like China, Singapore and Israel hastily developed privacy-endangering schemes to computationally trace contacts using user-generated location histories or mass surveillance data [1]. There have been reports of de-anonymizations of infected South Korean citizens using the "anonymized" data set published by the state [2].

Citizen privacy is not a design goal in any of the known contact tracing schemes. While many countries require that infected people share their location history with the health authorities [2], it should not be required that everyone does. Nevertheless, society as a whole has an interest in resuming normal life as soon as possible. To approach this conflict of interests we propose a system for privacy-preserving contact tracing.

II. System Design

A. Secure multi-party computation

Secure multi-party computation (MPC) [3, Chapter 22] deals with creating protocols for joint computation on private, distributed data. It studies mechanisms to allow a group of n independent participants to collectively evaluate a function $y_1, \cdots, y_n = f(x_1, \cdots, x_n)$. Each participant holds a secret x_i, which shall remain hidden. The participants only learn their final result y_i, but not the input data of others. Any function f that is solvable in polynomial time can be represented as an MPC protocol [3, Chapter 22.2]. For our application we only consider two parties. We assume a semi-honest security model for the participants in our protocol. A semi-honest model can be reinforced to provide security in a malicious setting by accepting a performance penalty [4].

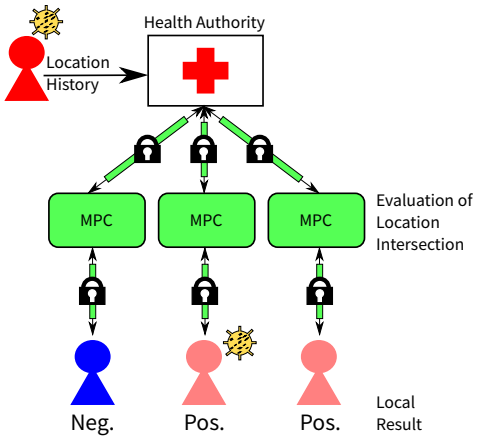

Fig. 1. An infected person shares their location data with the health authority. Three untested individuals starts separate MPC sessions with the HA. Note that not every positively evaluated contact is necessarily infected.

One way to realize a MPC protocol are *Yao's garbled circuits* [3]. Running an MPC protocol requires the one side to create a *circuit* from the function to be calculated and send it to the other party. The other side evaluates the circuit. Evaluation requires oblivious communication between both parties.

B. Oblivious Random Access Memory

Data-dependent accesses to memory are difficult to realize in MPC because the used index is leaked to attackers. To solve this issue, *oblivious random access memory* (ORAM) was invented [5]. It enables reading and writing data stored at a secret index i. The index is hidden by a set of random accesses to the ORAM. The oblivious database can either be located on a single server participating in the MPC execution or shared between all parties in the form of secret shares. State of the art ORAMs such as Floram [5] require only $\mathcal{O}(\sqrt{n})$ in computation and communication per access.

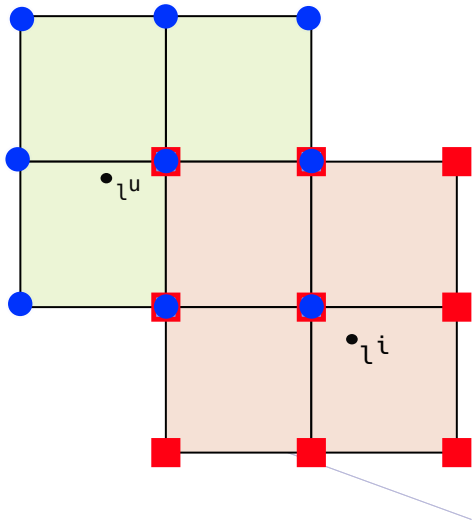

Fig. 2. Locations have multiple dimensions. A data point l is rounded to the closest position on the grid. Using this as center, the set L of adjacent grid locations is computed covering the region close to l. If a set L^u belonging to a users position intersects with the set L^i of an infected individual, the user can have contracted the disease.

C. Contact Tracing with Secure Multiparty Computation

Our proposed system (see Figure 1) takes advantage on the existence of health authorities (HA) that are collecting location histories of infected users, as done in many countries hit by the epidemic [6]. We also assume that a vast majority of individuals use location-based services that store their history locally. The HA can use the data points of infected patients (and the associated timestamps) to initiate MPC sessions with everyone who wants to trace themselves. The HA creates a circuit which it will send to all interested individuals. During the evaluation, each individual has to perform oblivious communication with the HA. Together, the parties determine where trajectories of infected and non-infected people intersect. Contact tracing is done in private so that only the traced individual can learn their status. No information about past locations of infected people or users of the system is revealed to either side.

D. Contact Tracing Algorithm

MPC allows inputs and outputs to remain hidden from other parties. Input locations $l := (x, y, t)$ consist of geographical coordinates and a temporal component. Each user u has n to-be-checked locations in their location history. The HA holds a number of m location data points from infected individuals. For a location l each component is rounded to a fixed granularity (e.g. 1 meter or 1 minute). Then, a set of locations L are calculated for which the Euclidean distance to l is smaller than a fixed threshold. Due to the reduced granularity the number of elements in the set is small. Both the HA and the user compute L for all their respective data points. The HA stores the result in an ORAM. For each position in $P = L_1^u \cup \cdots \cup L_n^u$ the user initiates a secure binary search on the ORAM. If an element from P is also found in the ORAM, then the user has been in contact with an infected individual. The number of contacts can be used to derive a risk score. Only the user herself will learn the final result. This algorithm described above is guaranteed to not leak more information than the ideal functionality to either side.

E. Performance Considerations

The binary search requires $\mathcal{O}(n \cdot log_2(m))$ steps per user. For testing our system we used the binary search implementation in the ACK library [7]. Depending on the amount of data to be stored, the library is capable of choosing between four different types of ORAM. For the tested value, it defaulted to Floram with CPRG [5]. Accounting for the runtime of Floram, the contact tracing algorithm takes $\mathcal{O}(n \cdot log_2(m) \cdot \sqrt{m})$. Using Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz and 16 GB RAM and setting $m = 3000$ and $n = 30$, the execution required 26.40 ± 6.63 seconds (standard deviation).

F. Decentralization

Our proposed system uses a central party (the HA) for contact tracing. Each person wishing to check their own history for contact points with infected people has to go through this central instance. Theoretically, a malicious HA could induce a movement of panic by tailoring additional locations as contact points. Withholding real contact points is also a possibility for the HA to touch up statistics. However, both of these attacks would not be helpful for reaching the overall target of tackling a pandemic situation, which is why we assume a semi-honest HA.

An alternative approach to contact tracing currently used by the Singapore Ministry of Health [1], [2] is to use Bluetooth and Wifi sensing. With the idea of Altuwaiyan et al. [8] it is possible to compare a set of observed devices to a set of devices belonging to infected individuals. In contrast to our proposal, this approach requires that every participant has their radio interface activated at all times.

References

[1] Hamilton, Isobel Asher, "11 countries are now using people's phones to track the coronavirus pandemic, and it heralds a massive increase in surveillance ," www.businessinsider.com/countries-tracking-citizens-phones-coronavirus-2020-3?r=DE&IR=T, accessed: 26.03.2020.

[2] Singer, Natasha and Sang-Hun, Choe, "As Coronavirus Surveillance Escalates, Personal Privacy Plummets," www.nytimes.com/2020/03/23/technology/coronavirus-surveillance-tracking-privacy.html, accessed: 26.03.2020.

[3] N. P. Smart, *Cryptography made simple*. Springer, 2016, vol. 481.

[4] S. Micali, O. Goldreich, and A. Wigderson, "How to play any mental game," in *Proceedings of the Nineteenth ACM Symp. on Theory of Computing, STOC*, 1987, pp. 218–229.

[5] J. Doerner and A. Shelat, "Scaling oram for secure computation," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 523–535.

[6] World Health Organization, "Operational Planning Guidelines to Support Country Preparedness and Response," www.who.int/docs/default-source/coronaviruse/covid-19-sprp-unct-guidelines.pdf, accessed: 26.03.2020.

[7] Doerner, Jack, "The Absentminded Crypto Kit," https://bitbucket.org/jackdoerner/absentminded-crypto-kit/src/master/, accessed: 05.04.2020.

[8] T. Altuwaiyan, M. Hadian, and X. Liang, "EPIC: efficient privacy-preserving contact tracing for infection detection," in *2018 IEEE International Conference on Communications, ICC 2018, Kansas City, MO, USA, May 20-24, 2018*. IEEE, 2018, pp. 1–6.