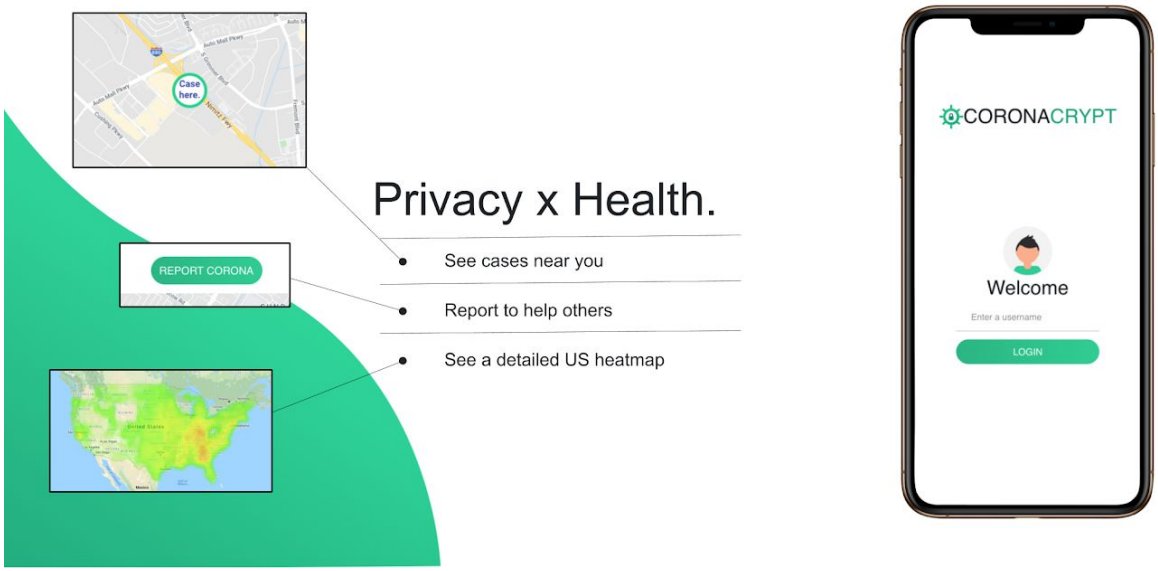
CoronaCrypt

Preserving privacy, protecting health.

By Joseph Zhang, Arnav Garg, Luke Zhao, Ronak Badhe, Alden Gu, and Tevin Wang

Test our software (Highly Recommended)**:**
https://tevinwang.github.io/coronacrypt/ (Location permissions required)

Source code:
Client: https://github.com/tevinwang/coronacrypt
Server: https://github.com/r2dev2bb8/coronacryptServer



Introduction

As more countries attempt to "flatten the curve" and limit the spread of this highly contagious virus, many cities are enforcing social distancing and shelter-in-place policies. **Individual location data sharing provides valuable surveillance insights** leading to more accurate data and increased model accuracy. With this increased knowledge, medical professionals and administrative leaders can better **determine hot spots and overall trends** and improve their response. For instance, they are informed of **citizen compliance** with social distancing and where infected patients have travelled, gaining knowledge about the viral spread. This allows for **more effective testing and smarter healthcare resource management**.

However, **contact tracing implies severe privacy risks** because sensitive location data and health information can be exposed. Recently, the governments of South Korea and China, have employed **intrusive tactics** to access smartphone location data and publish this information publicly. This lends itself to the possiblity of **unethical ad targeting and base discrimination against at-risk individuals.** Currently, there are digital approaches to conduct contact tracing with location histories, but many operate on a **skewed balance between privacy and effectiveness**. Even when systems do attempt to preserve privacy, methods are often insufficient. They often involve simply anonymizing user identifiers or encrypting in the backend, which, as described above, is catastrophic with an untrusted authority.

Our software is significant because **it refutes the previous notion that privacy reduces effectiveness in a contact tracing application**. We propose a technology-based solution for contact tracing and spreading information on the current magnitude of infections while protecting the privacy rights of citizens and diagnosed carriers.

Social Impact

a. CoronaCrypt can **inform users about a potential risk of infection** from someone they have been in close proximity to. Users can then effectively self-isolate and **reduce the spread of COVID-19.** Furthermore, a study from Nature Medicine estimates that 44% of all COVID-19 infections are from presymptomatic patients, and people are most infectious before they experience any symptoms. Therefore self-isolation, even before symptoms emerge, is critical, and CoronaCrypt allows people to self-isolate under **any risk of proximal infection**.

b. In addition, location data from users allows for **more accurate models and disease distribution maps**. Better visualization of the disease and its spread would **allow health workers to selectively test with more efficiency**. This would increase the effectiveness and availability of testing by directing tests to specific areas that need it the most. With more accurate visualizations of COVID-19 infections, **experts are able to make more accurate predictions and models of the disease's prevalence and spread**. Government officials are also

able to **better make decisions regarding the allocation of lifesaving resources and the reopening of businesses** in accordance with more accurate disease models across the country.

c. Our software also attempts to eliminate the uncertainty and anxiety that comes along with this disease.

d. The argument that there may be a possibility of "false" self diagnoses fails to account for the impact that CoronaCrypt has **in reducing viral spread, addressed global socioeconomic damage which is vitally important**.

e. With other contact-tracing apps, one issue is efficacy: If not many people are willing to use the app in the first place, the practicality is limited. However, **we plan to transform CoronaCrypt into a plugin in the future, making it significantly more accessible to users**. Our GPS contact-tracing method can use location data already collected by other apps, like Google Maps, to inform users of infection risk, eliminating the need to download other apps.

Literature Review

The keyword search of contact tracing applications revealed **2150 publications and other scholarly articles, with only 71 regarding privacy measures** (see Table I.). The current mechanisms being researched concern using smartphone Bluetooth signals and blockchain technology.

Database	ProQuest	EBSCOhost	JSTOR	Science Direct	Total
Hits	855	312	722	261	2150
Relevant Hits	54	11	4	2	71

Table I. Results of the Keyword Search of Current Contact Tracing Applications.

Some applications, such as the one made by Apple and Google, use Bluetooth to receive and record contacts with other users' phones. However, **Bluetooth tracking is only limited to simultaneous contact** and cannot accurately record transmission through commonly-touched surfaces, something our system is capable of assessing. **Another issue with Bluetooth is its low scalability because users must download a special application**, whereas GPS contact tracing can be achieved through a plugin as most users already have applications recording users' location history.

On the other hand, contact tracing with blockchain technology leverages the fact that a decentralized paradigm c**an be completely trusted to keep authorities from maliciously**

accessing sensitive, personal data. This pandemic has shown that **it is difficult to make a practical, perfectly decentralized contact tracing application**. For instance, CoEpi, an open source app, places more trust in the individual user rather than the central authority. However, their design is susceptible to privacy attacks from other users. CoronaCrypt addresses the shortfalls of these other technologies and pursues further sophisticated privacy policies, such as unlikability and confidentiality of a user's personal information.

Privacy-Preserving Scheme

We present a solution for **contact tracing and distributing health information while protecting the privacy rights** of both infected users and other citizens. Confirmed as important by the publications found in the literature review, the following privacy and design objectives for our contact tracing application center around.

- **Confidentiality** of the users' location and health information against external parties and other users.
- **Unlikability** between location data and users.
- **Efficiency**, **scalability**, and **usability**.

Methods

1. **Data Collection**:

The application collects **timestamped GPS coordinates to 4 decimal places every *t* minutes**. This results in points around 20 feet from each other to be the same. These values are stored locally in an array-like data structure in addition to being encrypted and sent to the database over a secure channel daily. We anonymize the identity and locations of the user by using a **one-way hash function** (NIST standard SHA-256 hash algorithm). The server **automatically deletes any points more than *n* days ago**, where *n* represents the incubation period (14 days for 99% of cases). This way we can accurately come up with a comprehensive overview of someone's whereabouts for the last *n* days. The overview of this end-to-end encryption procedure is shown in Fig. 1.

2. **Database Operations**:

While the secure channel is established, the **received anonymized coordinates are compared with infected users' coordinates in similar timeframes**. This is done using a **cryptographically secure private set intersection**. If there is an overlap, the server calculates a risk value (done in step 3) and notifies the user. Since points around 20 feet from each other are encrypted to the same value, the set intersection will have an overlap defined to a distance less than 20 feet instead of no distance which dramatically speeds up the operation,

After a user is diagnosed as a carrier, **they have the option to anonymously share** their location history for the benefit of administrative agencies, medical professionals, and system users. Our system performs two operations with the user's anonymized location history.

a. The user's former GPS coordinates are **obfuscated and replaced with a larger geographic area and added to a colored, graduated heat map**, indicating the risk of people in those areas.
b. Each coordinate in the **infected user's location history is hashed and posted** on the server for all the system users to compare. We encrypt all coordinates to maintain **confidentiality** and **unlikability**.

3. **Risk Assessment**:

The app calculates a **quantitative value** for the risk of the user. This value is a combination of different metrics, including **duration of contact and the risk value of other users**. If the risk value passes a threshold, the app will notify the user and encourage them to **self-quarantine or request a test**. Note that we intend to calibrate this value in the future with professional medical assistance.

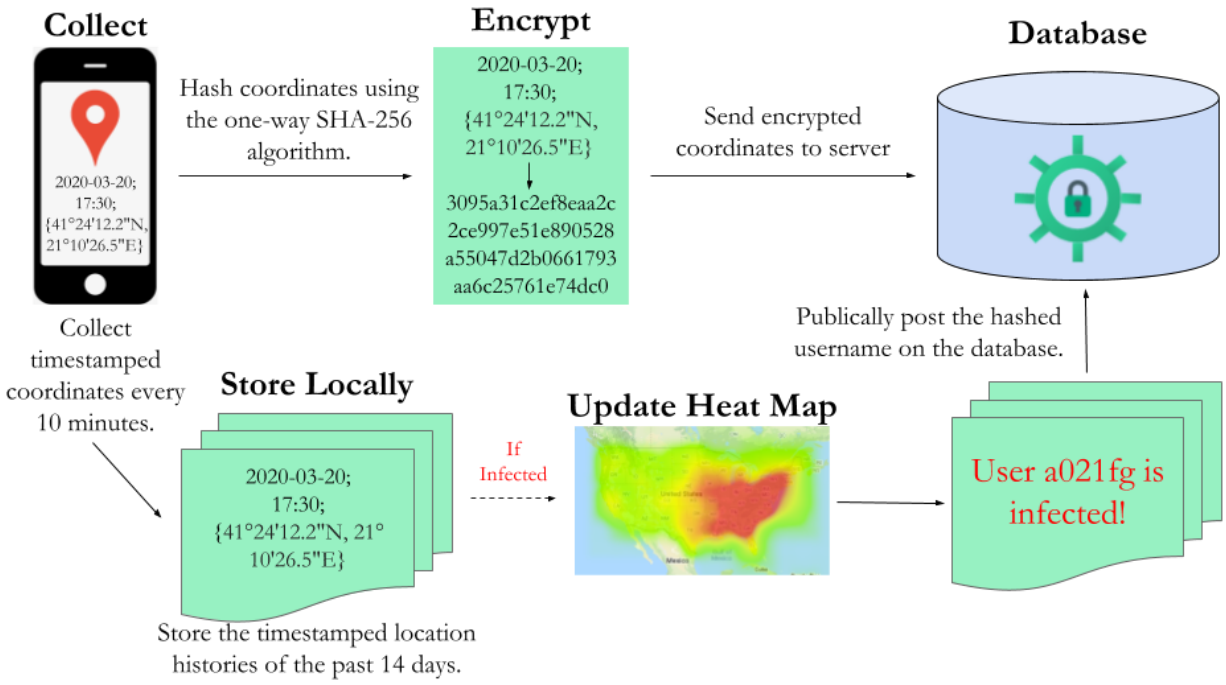


Fig. 1 Client Procedures; collected coordinates are encrypted before sending to the server.

Interval Matching Algorithm Assessment

Before the application sends time-stamped coordinates, it hashes the coordinates using SHA-256. The time value is not hashed, so the database can check if there are proximity overlaps in adjacent time periods i.e. (virus was spread through a commonly-touched surface). Since the

locations were encrypted with a one-way hash, **it is impossible to reverse the hashed points and reveal the underlying location history, promoting perfect privacy and unlinkability.**

These hashed values are not meaningless, however, since a match with an identical value in a similar time period means that two users came into contact. We find these matches using the **cryptographically secure "Private Set Intersection" algorithm**. After much testing and deliberation, we created a function to **compare two encrypted coordinate time intervals and find the intersection without anything to the server except for the elements in the intersection**. If there is a match, the system will compute a risk value as indicated in Fig. 2.

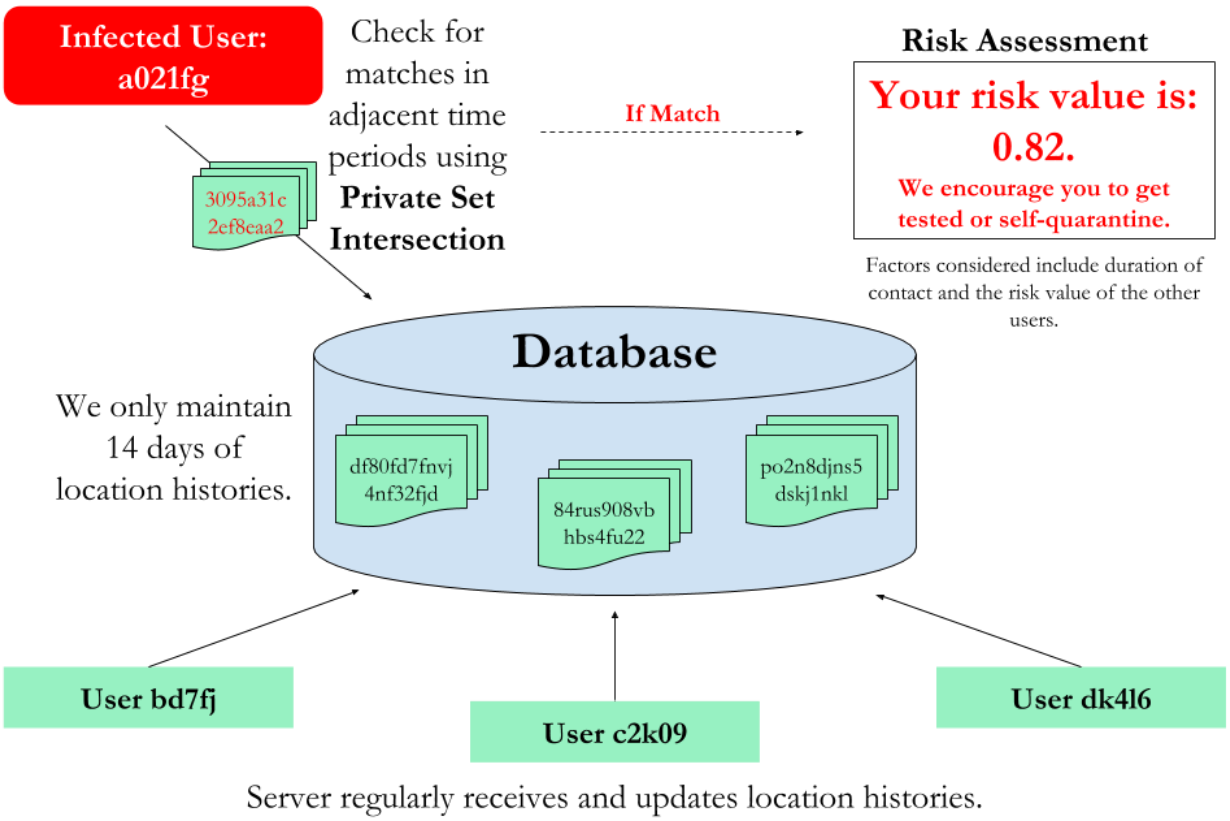


Fig. 2 Server Procedures; Can be operated by untrusted authorities coordinates are encrypted.

Heat Map Generation

After a diagnosed carrier grants our server access to the anonymized location history, we update the heat map with this information as indicated in Fig. 3.

Fig. 3 The generated heat map indicates areas with increasing case counts (data taken from Johns Hopkins University https://github.com/datasets/covid-19).

Using the **Google Maps API, we insert these coordinates which will dynamically scale the heat map colors to the greatest concentration of points at any particular pixel on the map**. A heat map can indicate how **infections are spread** across the country and more notably, allow users to **see if they came in contact with a possibly infected area.** As we constantly delete old infected users' location histories, the heat map will always be appropriate and accurately describe the **scope of the infection concern within a given region.**

Future Work

Our main goal for the future is to work with **medical professionals to better adjust our risk assessment system** to accurately dictate when users should get tested. Currently, our risk threshold value could be **calibrated later with more research and understanding.** We could use an **artificial intelligence approach for this issue as there are many factors** that go into calculating this value i.e. (duration of contact and the risk value of other users). Furthermore, we aim to better involve the users into our system, such as giving them the **privilege to manually redact any sensitive GPS coordinates**, including home areas or workplaces. Users should also be allowed to customize other values, like the time interval that GPS collects coordinates at as well as the granularity of the heat map (of course while maintaining a reasonable level of privacy). One potential step towards this goal would be to add the hashing, which would consist of small tweaks, to the server which would protect the information of the users. Finally, we aim to **transform this application into a plugin, so this technology can be more widely adopted** as many users' other applications already collect location history. **Maximizing the amount of users will translate to a better response and containment of COVID-19.**

Conclusion

CoronaCrypt is a privacy-preserving application with end-to-end encryption. **Unlike many other contact tracing applications that only encrypt in the backend, the user hashes all the collected coordinates** before sending them to the server to avoid an untrusted authority gaining access to personal location and health information. We used a novel method to perform set intersection operations on these hashed location histories, yielding an **efficient** and **scalable** way to compare adjacent time periods where users may have been infected. Furthermore, our model can compute a **quantitative risk value**, which reflects **the probability that users were infected** based on factors like the duration they were in contact with a carrier and the risk value of other users around them.

CoronaCrypt gives universal aid in our current time and crisis. By allowing people who have COVID-19 symptoms to **anonymously self-report their condition**, other people under risk of infection can self-isolate even before symptoms appear, minimizing the virus' spread. Self-reports can also be used to target testing towards certain areas, improving the accuracy of disease models and ultimately government decisions.

Major References

1. Dana M. Lewis et al. Coepi: Community epidemiology in action. https://github.com/Co-Epi, Accessed: 2020-05-05.
2. The New York Times. In coronavirus fight, china gives citizens a color code, with red flags. https://www.nytimes.com/2020/03/01/business/china-coronavirus-surveillance.html, Accessed: 2020-05-05.
3. "COVID-19 Contact Tracing Training Guidance and Resources." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 4 May 2020, www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/index.html. Accessed: 2020-05-05
4. "UPDATED: Who Is Most Susceptible to the New Coronavirus?" *Cancer Health*, 6 May 2020, www.cancerhealth.com/article/who-is-most-susceptible-new-coronavirus. Accessed: 2020-05-05