INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 07, July 2024, pp: 687-694

e-ISSN : 2583-1062

Impact Factor: 5.725

MAINTAINING PRIVACY IN MULTI-CLOUD SECURE DATA INTEGRATION FOR INFECTIOUS-DISEASE EXAMINATION

U. Vara Prasad[1], Ch. Sai Krupa[2], S. Ravendra[3], Ch. Vamshi[4], E. Prassana[5], Ch. Prabhakar[6], D. Udayasri[7], L. Varshni[8]

[1,2]Associate Professor Computer Science And Engineering Bomma Institute Of Technology And Science Khammam, Telangana, India.

[3,4,5,6,7,8]B. Tech degree in Computer and science Engineering from the University of JNTUH Bomma institute of technology and science Khammam, Telangana, India.

ABSTRACT

People's data are frequently seen to be dispersed around many companies, yet when combined, these data can produce valuable insights. Even with certain security measures, data fusion from several data hosting sites may compromise user privacy. This work examines a data-analytic platform that integrates and analyzes user location and health data, which are kept in two different clouds, to identify infectious disease geographic hotspots using the Kulldorff scan statistic. We investigate the privacy risks associated with this platform, which employs a key-oblivious inner product encryption (KOIPE) approach to guarantee that the honest-but-curious (HbC) entity is only exposed to coarse-grained statistical data. Utilizing a game-theoretic strategy, we safeguard user privacy against the intended inference attack by providing incentives for users to create clusters that are anonymous and ensure quantitative privacy. We show the effectiveness of our system in terms of design overhead and privacy level through comprehensive simulations based on real-life datasets. Index Terms: game theory, secure multi-party computing, Bayesian inference, public health, and Kulldorff scan statistic.

Keywords: Kulldorff scan statistic, game theory, secure multi-party computing, Bayesian inference and public health.

1. INTRODUCTION

An Infectious illness is a significant public health issue, being among the top causes of mortality in the US [2]. Infectious illnesses have a huge impact, but sadly, as our society becomes more urbanized and globalized, our society becomes more susceptible to their epidemics. The COVID-19 epidemic of this year serves as a grim reminder of just how dangerous infectious diseases can be. Thus, public health agencies' primary responsibilities are prompt and effective identification of infectious disease epidemics and early response to them, should they arise. The spatial clustering analysis is crucial among many other analytic interests [3]. In the early stages of a disease epidemic, epidemiologists might detect geographic disease clusters by using people's health (such as fever, coughing) and location (such as zip code). Next, open-access resources

such as prophylactic antibiotic use might be assigned to stop its spread. The use of disease monitoring systems has increased recently. The COVID-19 web dashboard was created to monitor COVID instances worldwide [4]. A further initiative known as BACTracker, or Biological-Agent Correlation Tracker.

Nevertheless, there are certain drawbacks to these mobile participatory systems, including their imprecise timeliness, restricted representativeness, and unstable participation rate. As an illustration, certain systems gather data on a weekly basis; the bulk of participants are female; and participation rates are correlated with sickness, with new members being more likely to be unwell than returning ones. For a number of reasons, including sociodemographic variations and privacy concerns (such as unsolicited, invasive marketing), patients have also been seen to be typically quite unwilling to disclose their location and health information [6]. As much personal data as possible is ideal for high-fidelity clustering analysis, yet current mobile systems fall short in this area.

This study specifically makes the following significant contributions. We provide a new system for geographic clustering analysis of infectious illnesses that gathers users' multi-institutional data across several cloud platforms. Using the key-oblivious inner product encryption (KOIPE) technique, we create an SMC-based method that restricts access to statistical data of groups rather than individuals by untrusted actors. We show that Bayesian inference is effective.

Reducing needless data dissemination as much as you can, ideally to only statistical information, is another way to protect privacy. A cryptography tool called Remind offers safe computing for statistical analysis. Among the secure procedures it implements are statistical tests, covariance, and quantiles, to mention a few. Recent research has demonstrated, however, that privacy is not maintained on a platform that exclusively makes statistical data available

through safe calculations [8], [9]. The membership inference attack is the process of exposing user data in an aggregated statistical dataset using a basic inference method such as random forest. In order to protect user privacy, implemented differential privacy as a countermeasure to an aggregated dataset. SMC has been frequently used recently to secure data analytics from numerous sources. To prevent record linking, suggested an SMC technique to eliminate duplicate entries across several databases. Nonetheless, have noted that SMC outputs can still leak some important input data. same time to 1) control access to outsourced data by performing re-encryption techniques and 2) manage the dynamic ownership when real data owner is offline or revoke his/her ownership. exchanges between users and KS, which provided the attackers with chances to extract valuable information from the exchange. Miao et al. presented a secure deduplication method for multi-server-aided. In order to accomplish detailed data deduplication.

2. RELATED WORKS

Infectious disease monitoring and analysis:

Recently, there has been a lot of interest in the spatial study of infectious illnesses. Creating surveillance systems for data collecting is one area of academic focus. The COVID-19 web dashboard is one example of a model platform [4]. Nevertheless, the participation rate of these systems is not consistent, and the representativeness of the data obtained is restricted. An additional, but useful, method for tracking infectious diseases has emerged with the rise in popularity of social media and other cloud services. For example, Twitter data is used for Ebola and flu tracking. The efficiency of tracking 42 infectious illnesses, including dengue and malaria, in China using Weibo data was demonstrated. A privacy-preserving approach for combining social and health clouds to forecast users' health conditions based on social connections.

Privacy and security in infectious disease analysis:

In order to protect privacy throughout the collection, distribution, and use of data for health applications, randomization (such as differential privacy) and anonymization are often used methods. Nevertheless, these methods frequently either cause significant distortion that causes mistakes in the data analysis or experience re-identification assaults that undermine the effectiveness of protective measures. Another way to protect privacy is to limit the amount of needless data that is disclosed, ideally to only statistical data. A cryptographic tool called Remind offers safe computing for statistical analysis. Among the secure procedures it does are statistical tests, covariance, and quantiles, to mention a few. Recent research, however, has demonstrated that privacy is not maintained on a platform that solely makes statistical data available through safe calculations. The membership inference attack refers to the disclosure of user information in an aggregated statistical dataset via a straightforward inference technique such as random forest. In order to protect user privacy, added differential privacy to an aggregated dataset as a countermeasure. SMC is currently common to protect data analytics from many sources. As an illustration, suggested an SMC approach to eliminate redundant entries from many databases in order to prevent record linking. But as recently noted, some sensitive input data could still escape from SMC outputs. The objective of this study is to fill a gap in the literature by utilizing a non-perturbative method to safeguard user privacy against inference attacks on statistical data, particularly that obtained from the SMC output. In addition to protecting privacy, this method won't prevent the infectious disease monitoring system from gathering high-quality data. To put it another way, it ought to encourage people to submit their data for examination.

3. SYSTEM MODEL

The technical model for data fusion and the security model are briefly discussed in this subsection.

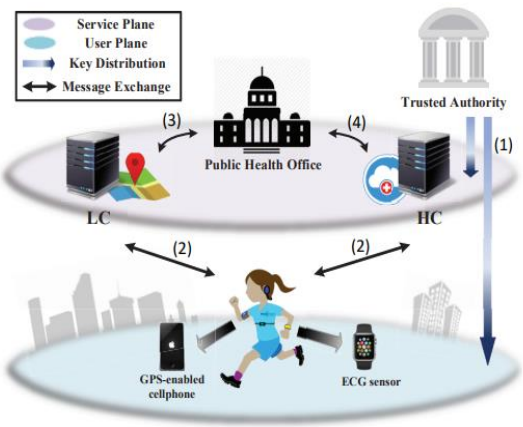


Fig.1. System model for the data fusion framework exploiting two cloud services.

Fig. 1 depicts the five entities that compose up this system. The public health office (PHO), cloud server for location-based services (LBS), health service cloud server (HC), trusted authority (TA), and users are all together referred to as this group.

Users can access services from LC and HC using off-the-shelf gadgets like wearables and smartphones with GPS capabilities. Users can obtain location and health services via LC and HC, which are run by separate businesses, respectively. PHO is a public health organization that monitors and analyzes diseases. PHO can obtain data from LC and HC with user agreement. TA creates and disperses secret materials to other organizations in addition to bootstrapping the system. Additionally, it resolves conflicts and, if necessary, revokes misbehaving entities. A few real-world systems and business scenarios serve as the foundation for the assumption that there are three separate and functional entities. In this regard, during the COVID-19 pandemic, Apple and Google created their own application programming interfaces (APIs) in their respective app stores to gather user location data (i.e., physical contact) and health information. This allowed the authorities to send out contract warnings and facilitate the tracking of the virus's spread by the likes of the CDC. The privacy protection mechanisms presented in this research can be included into the current platforms.

Security Model

TA cannot be hacked by an enemy since it is completely trusted by other system components. Sincere but inquisitive are LC and HC (HbC). In other words, they sincerely adhere to protocol while being interested in the location and health data of consumers. Delivering advertisements, refusing insurance to sick users, and many other things are among their inducements for bad conduct. PHO is also presumed to be HbC in that it really does statistical analyses of infectious diseases, but it also requests personal information from users in order to perform tasks such as grouping individuals who are infected, which goes against their will and compromises their privacy. It is believed that PHO, LC, and HC operate independently of one another. This system's users cannot be trusted. By sending LC and HC bogus or duplicated data, they may initiate Sybil attacks to skew PHO's analysis. There is an incentive to either spread fear in an unaffected area or lessen PHO knowledge in an infected area. This pertains to companies attempting to outperform one another commercially or bioterrorists.

4. PRELIMINARIES

Kulldorff Scan Statistic [22]:

Initially introduced in 1997 [22], the Kulldorff scan statistic has grown to be a potent instrument for conducting temporal and geographic clustering analyses for infectious illnesses. Small areas (like a school or a mall) with noticeably higher disease densities can be found by spatial analysis using the Kulldorff scan statistic. We go over how the Kulldorff scan statistic functions in the sections that follow.

A surveillance region K is divided into subareas {s1, s2…, sk} of any arbitrary level of resolution, and $\kappa = \cup_{i=1}^{k} s_i$. The disease headcount and population in each subarea, denoted as {c1, c2…, ck} and {p1, p2…, pk}, respectively, are collected. Let the total disease case count$C_{tot} = \sum_{i=1}^{k} C_i$ and census population $P_{tot} = \sum_{i=1}^{k} P_i$ of whole region k. The Kulldorff scan statistic is then applied should look through every potential grouping of nearby subareas to find any unusual ones that have an excess of illness. Let's say that the set of such clusters is denoted by the notation {s1, s2..., sk}, with a population of Pj and a disease case count of Cj. Next, the corresponding cluster density Dj is computed using the Kulldorff spatial scan statistic.

$$C_j \log\frac{C_j}{P_j} + (C_{tot} - C_j)\log\frac{C_{tot} - C_j}{P_{tot} - P_j} - C_{tot}\log\frac{C_{tot}}{P_{tot}}, \qquad (1)$$

If $\frac{C_j}{P_j} > \frac{C_{tot}}{P_{tot}}$ and 0, otherwise.

In so doing, the maximum density mrd =$max_{sj} D_j$ and the corresponding cluster mdr = arg$max_{sj} D_j$ in the region k can be identified. To evaluate if this cluster is statistically significant, the Kulldorff spatial scan statistic assumes that Ci follow inhomogeneous poisson processes and a randomization testing approach is conducted to examine mdr. The statistical significance (i.e., p-value) is then calculated so that the cluster is considered as the outlier or being statistically significant when p≤ 0.05.

Privacy Metric

It adopts the widely accepted privacy metric [8] and define user privacy level as the opposite of the capacity of an attacker to accurately deduce personal information about a user. Area Under Curve (AUC) is specifically utilized to

quantify the adversary's overall inference performance, and the privacy loss (PL) score is computed to indicate the loss of privacy experienced by a specific user.

AUC Score: Assume that the adversary's inference is $x* \in \{0, 1\}$ and that the user's private information is $x \in \{0, 1\}$. We possess the subsequent metrics:

- True Positive (TP) when $x* = 1$ and $x = 1$;
- True Negative (TN) when $x* = 0$ and $x = 0$;
- False Positive (FP) when $x* = 1$ and $x = 0$;
- False Negative (FN) when $x* = 0$ and $x = 1$;

It allowed us to calculate the True Positive Rate (TPR) as TP/(TP+FN) and the False Positive Rate (FPR) as FP/(FP+TN). Next, the Receiver Operating Characteristic (ROC) curve might be constructed based on various discrimination thresholds. The area under the ROC curve (AUC) represents the adversary's overall success in inferring the user's information x.

PL Score: Assume that everything goes according to plan, meaning that the adversary must randomly guess (AUC = 0.5) $x*$ in order to obtain any prior knowledge about users' private information. We characterize the PL score as the adversary's relative belief gain over its initial random guess, and we find that AUC may rise through the inference attack:

$$PL = \begin{cases} \frac{AUC-0.5}{0.5} & \text{if } AUC > 0.5 \\ 0 & \text{otherwise} \end{cases}$$

5. SECURE MULTI-CLOUD DATA FUSION FOR INFECTIOUS DISEASE ANALYS

Protocol Overview:

Fig. 1 depicts the general information flow. TA creates and provides security materials to PHO, LC, HC and users during system initialization.

After that, the system can go offline (shown as flow (1) in Fig.1). As will be mentioned, PHO first chooses an incentive to encourage users to provide their data whenever an infectious disease epidemic occurs. Referred to as flow in Fig. 1, LC and HC are responsible for gathering user location and health data, respectively.

An anonymity scheme based on identity-based encryption is intended to prevent Sybil attacks, while an authentication scheme based on anonymous group signatures is meant to provide controlled link ability of user data on HC and LC. At last, PHO collaborates with LC and HC to collect users combined location and health data, which are designated as flow in Fig. 1 for spatial analysis.

Only the SMC-based design for statistical data computation, will be presented for the sake of conciseness. This design serves as the basis for our investigation into the danger to and protection of privacy in Section 6. Referring readers to our early work [1], we provide further information about our design for flows (1).

SMC-Based Secure Data Aggregation

The population Pj and the number of infected users Cj in each geographic grid j are the only census data needed by the Kulldorff scan statistic, therefore the design aim is to restrict PHO's access to these statistical (i.e., aggregated) data alone. Furthermore, because spatial analysis works with unreasonably big datasets, significant computational performance is required. In particular, these are the processes that comprise our design.

PHO begins by associating users' encrypted identities, uid at LC and uid at HC1, to match their health and location information, respectively, at HC and LC.

Subsequently, PHO and HC work together to securely compute the total illness numbers. In Fig. 2, we provide a toy example for demonstration. PHO, on the one hand, creates a query matrix Q that includes users' locations inside a single grid. In such grid, the user is represented by a value of 1; otherwise, it is 0. However, in vector H, HC keeps users' infected status. Then, using the inner product of Q and H, PHO can effectively determine the disease count vector CNT in each geographic grid. According to our design, HC should not see PHO's query matrix Q (for location privacy) and PHO should not see HC's health vector H (for health privacy). Essentially, our plan consists of a secure multiparty computation (SMC) design in which PHO and HC work together to jointly compute the inner product of Q and H while keeping other details secret.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 07, July 2024, pp: 687-694

Impact Factor: 5.725

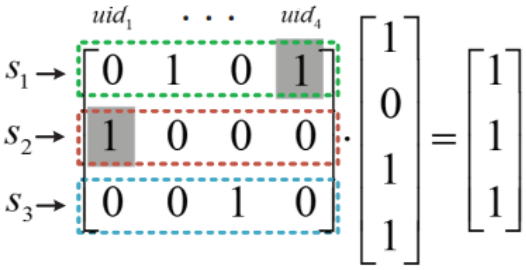


Fig.2. A model of the batch queries the four users and three grids in PHO's query matrix are represented by HC, which also provides the state of each user's infection. The inner product indicates how many users are infected in each grid.

It alters the original design of the safe k closest neighbor (kNN) method, which served as our inspiration, due to two reasons: (i) it only gives relative distance rather than an exact number; and (ii) both entities know the random matrix, which makes our approach insecure. To solve these issues, we suggest the Key-Oblivious Inner Product Encryption (KOIPE) technique. Fig. 3 provides an overview of our concept, and the following provides a thorough description.

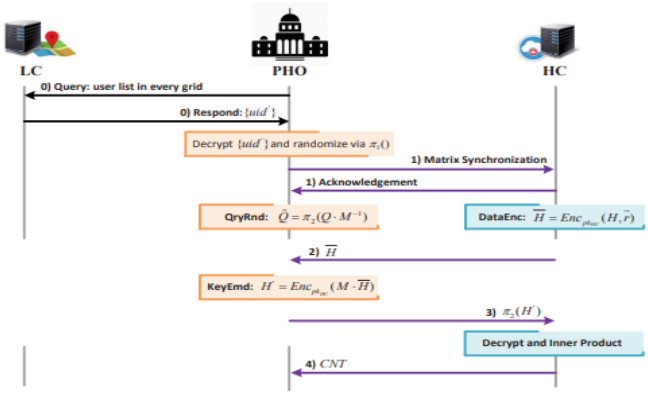


Fig.3. Diagram for information exchange and secure data aggregation

MtxSync: For the purpose of removing mismatched records, PHO and HC "synchronize" their matrices first. PHO utilizes a safe permutation process π1 to generate a random user list uid, which is subsequently forwarded to HC for rearranging and trimming H.

6. INFERENCE ATTACK ON STATISTICAL DATA

In light of its side information and inference capabilities, researchers assess PHO's ability to compromise an individual's health data in this section. Imagine a situation where PHO continuously tracks public health at the municipal level. In order to support its statistical analysis, PHO, in other words, continually gathers data from HC and LC. PHO may be able to determine a user's health status more accurately by combining data from several time instants, but this might decrease the user's degree of privacy. We next use a real-world dataset to illustrate the efficacy of this kind of inference attack.

Bayesian Inference Attack:

In this case, we choose to use the Bayesian inference model to instantiate PHO's inference function. When PHO examines the gathered dataset D, which is referred to as evidence, they may determine the posterior knowledge given the prior knowledge P. This procedure is repeated t times to record PHO's current perception of the health status of the users. PHO's assessment of whether user i has the illness or not is specifically described by time t. In this case, the health condition H is a collection of random variables that adhere to nonidentical Bernoulli distributions and are presumed to be independent of one another.

Evaluation

Researchers implement a real-world dataset from Gowalla [30], a location-based social networking site where users check in to report their position, to capture the mobility features. It includes, in particular, 6,442,892 user check-ins from February 2009 to October 2010. A user ID, location ID, latitude, longitude, and time stamp are included in every entry. We concentrate on the users who checked in on March 14, 2010 in Austin, Texas, USA for our analysis. A rectangular grid with the center at position (30.0°N, 98.0°W) and the corners at coordinates (29.5°N, 98.5°W), (29.0°N, 97.5°W), (30.5°N, 97.5°W), and (30.5°N, 98.5°W) was used to approximate the geographic region of Austin. Within the allotted period, 1,801 distinct users in Austin reported 10,638 check-ins.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 07, July 2024, pp: 687-694

e-ISSN : 2583-1062

Impact Factor: 5.725

It carries out a number of inference attacks on every user to ascertain PHO's inference capacity. Initially, the area of concern is divided into grids, such as 10 x 10, 15 x 15, and 20 x 20. Next, PHO's previous knowledge of each user's sickness status is set to 0.5 and the population's infection rate is assumed to be 10%, staying constant during the considered time. Next, based on the rate of infection, we randomly allocate each user a disease status (i.e., 0 or 1). Let's now assume that PHO gathers the statistical data throughout the course of a day at a temporal granularity of hours (e.g., 1 hour, 2 hours, and 3 hours).

Evaluation of Performance

The following subsection provides a numerical evaluation of the calculation overhead that was incurred for HC and PHO during the SMC-based data fusion stage. Additionally, we investigate how encouraging user engagement through the game theoretic method might strengthen users' privacy defenses against Bayesian inference attacks.

Simulation Setup

In order to mimic a user's operating environment, LC/HC and PHO, we employ a workstation with a 32GB RAM and a 3.4GHz Intel(R) Core (TM) i7 CPU. Similar to RSA, we use a 2048-bit modulus as the secret key length in the Paillier cryptosystem. The implementation is predicated on John Bethencourt's open-source platform, which was constructed using the C language's GNU Multiple Precision Arithmetic Library (GMP).

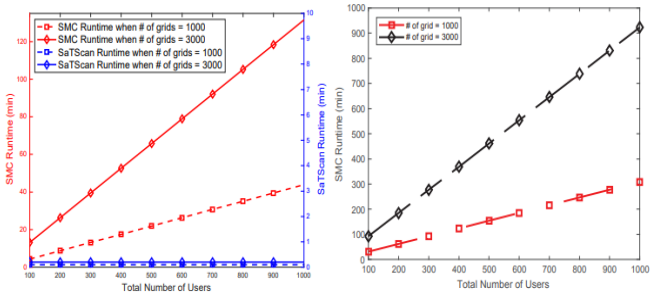
For Kulldorff scan statistics, we use a single dataset, the cancer incidence in New York State. It was created by compiling 67,217 tumour occurrences between 2005 and 2009, which covered 13,848 geographical groupings and an average population of 19.34 million according to the 2010 population census. In order to assess the effectiveness of our planned scheme, we take this dataset and filter just the incidence of lung tumours. We then use the random-drop method to modify the number of users, the number of diseases, and the number of clusters.

Additionally, we investigate ways to manipulate user behavior in a large sample group under various incentives to "hide" users' personal information. Assume that a single geographic grid has N = 200 users. The system parameters γ and λ are designated as control variables, signifying the user's partiality towards privacy and reward, and the PHO's inclination towards utility over money, respectively.

Analysis of Computational Overhead

It presented the theoretical element-wise analysis of the computing overhead for each protocol step in our early work [1], therefore we will not discuss it here and instead concentrate on numerically analyzing the SMC-based data fusion design's performance.

Here especially present an end-to-end computing overhead that entails executing the SMC protocol to acquire the statistical data and locating the spatial clusters using the Kulldorff scan statistic. It wants to compare the extra overhead caused by the non-functional security approach by doing this. Firstly, the employ the SMC protocol that is based on KOIPE and is based on the Paillier platform developed by John Bethencourt. However, to execute the Kulldorff scan statistic, can employ the simulator made available by the open-source SaTScan, as described in [42]. The situation is that a region of disease over-density is being tested to see if it is statistically significant among others using the inhomogeneous Poisson process. We create R = 999 replications for each statistical analysis, and we set up to 0.05. To eliminate randomness, ten separate runs of the SMC and SaTScan simulations are performed.



(a) PHO (b) HC

Fig.4. Computation overhead for PHO and HC.

Three major findings can be drawn from this simulation, as seen in Fig. 4. More users and tighter grids make the difference between the SMC and SaTScan runtimes more noticeable. The granularity of spatial grids is the sole factor that affects the SaTScan duration; user count has no bearing on it. PHO has a substantially shorter runtime than HC. For our initial finding, the overhead associated with the SMC approach is reasonable (on a minute scale) given that a disease monitoring system at the city level may only gather data on an hourly or daily basis. Regarding the second

observation, this is because the Kulldorff scan statistic is designed to evaluate spatial clusters, while Eq. 1 normalizes the illness and user numbers. The third discovery can be explained by the fact that HC computes Paillier's encryption and decryption functions, but PHO just performs arithmetic modular exponentiation and multiplication.

Privacy and Incentive under Game-theoretic Approach

The choices made for γ and λ, being system parameters, have a significant effect on how the game-theoretical model is solved. We show how the ideal payment/incentive P varies with respect to γ and λ in Fig. 5. It is evident that PHO tends to provide a greater P for any γ as λ declines. This is because users tend to favor privacy protection, particularly when $\gamma < 1$. Consequently, in order to optimize its usefulness, PHO has to provide a bigger incentive P. Another noteworthy finding is that PHO's incentive grows exponentially for every γ as λ rises, and then stays constant beyond a certain threshold. This cutoff suggests the minimal P at which every user is encouraged to take part.

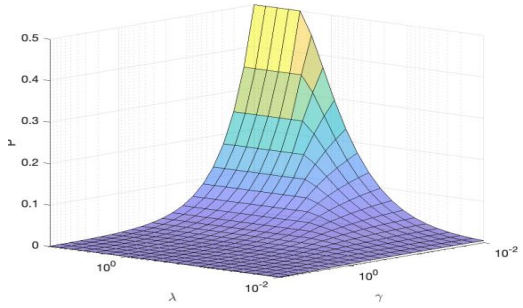


Fig.5. Selection and Impact of system parameters.

Next, let's look at how the PHO's incentive influences users' decisions, which in turn influences the PHO's own payout, based on the aforementioned numerical guideline for parameter selection. We used $\gamma = 0.005$ and $\lambda = 80$ for this evaluation, which suggests that the user values privacy over incentive and that PHO values usefulness over payment. The findings for the users' approach and PHO's reward in relation to the incentive offered by PHO are displayed in Fig. 6. It notes that as the PHO's payment increases, the threshold $t*$ rises gradually at first and sharply towards the end. This suggests that most users opt not to participate for a minimal incentive (privacy will be compromised otherwise), but that they all decide to join at the same time once the payment can make up for their loss of privacy and a sizable anonymity set is created. Additionally, $t* = 1$ is always true for $p > 1$. Users' best course of action is evidently to participate in consensus since this gives them the biggest aggregate group with the least amount of privacy loss. yet, PHO's reward is maximum when every user participates; yet, when payment grows, it decreases, which is understandable given that utility stays constant despite an increase in overall payment. Therefore, the best course of action for PHO is to urge all customers to participate by offering a minimum payment (in this case, p = 1).

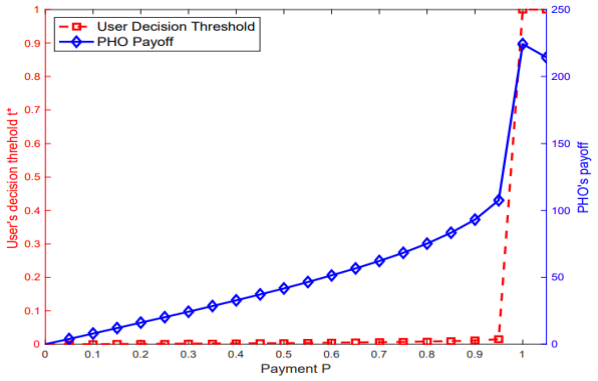


Fig.6. Users' and PHO's strategy under the game model.

Fig. 7 depicts how the minimal requirement for the aggregation group size might affect user privacy, guided by the best strategy derived from the game theoretic method. The identical configuration as previously described is used to perform the simulation. 10% is the infection rate, 10 x 10 is the grid size, and once per hour is the inference frequency. As more users contribute to the aggregate statistics, there is a discernible decrease in privacy loss. For example, reducing the average privacy loss by as much as 28% (from 0.24 to 0.18 in Fig. 7) is possible for a small aggregate size of 20 users. Furthermore, it is demonstrated that, in the absence of privacy protection, this approach benefits the outlier users whose PL was 1. However, as the game-theoretic model is a strategy-making model for agents of competing interests, it often falls short of providing proven privacy, meaning that most users' privacy cannot be totally secured (i.e., privacy loss = 0). To address it, one may look for more robust privacy ideas like differential privacy.

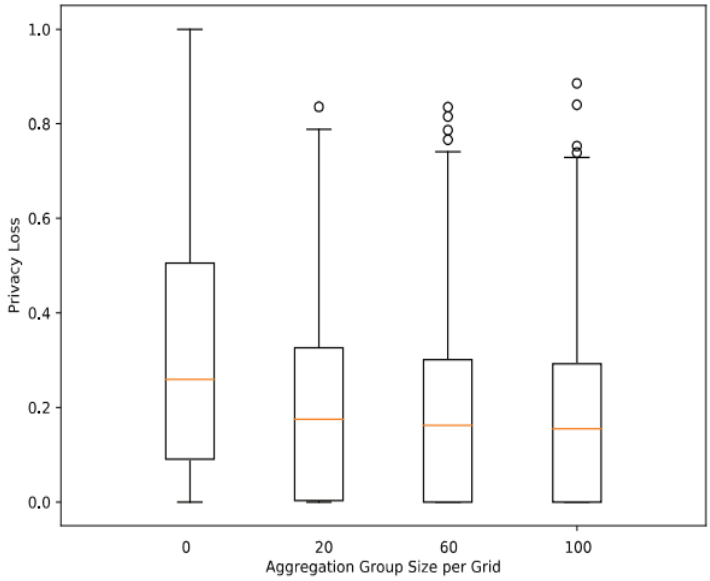


Fig.7. The impact of the minimum aggregation group size on user privacy loss.

7. CONCLUSION

The danger to privacy posed by a multi-cloud secure data fusion approach for infectious disease research has been investigated in this study. We have demonstrated the influence of a straightforward yet efficient Bayesian inference approach on the privacy loss experienced by users as a result of the disclosure of just statistical data from a built secure multi-party computing protocol. A game-theoretic strategy is put out to protect privacy against Bayesian inference assaults and to encourage logical users to donate their data for public health. Quantitative simulations using real-world datasets have been carried out and have shown the design overhead and privacy gain.

8. REFERENCES

[1] J. Liu, Y. Hu, H. Yue, Y. Gong, and Y. Fang, "A cloud-based secure and privacy-preserving clustering analysis of infectious disease," in 2018 IEEE Symposium on Privacy-Aware Computing (PAC). IEEE, 2018, pp. 107–116.

[2] H. Nichols, "The top 10 leading causes of death in the United States," 2017. [Online]. Available: https://www.medicalnewstoday.com/articles/282929.php.

[3] J. Cordes and M. C. Castro, "Spatial analysis of covid-19 clusters and contextual factors in New York city," Spatial and Spatio temporal Epidemiology, vol. 34, p. 100355, 2020.

[4] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track covid-19 in real time," The Lancet infectious diseases, vol. 20, no. 5, pp. 533–534, 2020.

[5] A. Szpiro, B. Johnson, and D. Buckeridge, "Health surveillance and diagnosis for mitigating a bioterror attack," Lincoln Laboratory Journal, vol. 17, no. 1, 2007.

[6] K. El Emam, J. Hu, J. Mercer, L. Peyton, M. Kantarcioglu, B. Malin, D. Buckeridge, S. Samet, and C. Earle, "A secure protocol for protecting the identity of providers when disclosing data for disease surveillance," Journal of the American Medical Informatics Association, vol. 18, no. 3, pp. 212–217, 2011.

[7] R. Dautov, S. Distefano, and R. Buyya, "Hierarchical data fusion for smart healthcare," Journal of Big Data, vol. 6, no. 1, p. 19, 2019.

[8] A. Pyrgelis, C. Troncoso, and E. De Cristofaro, "Knock, who's there? membership inference on aggregate location data," are Xiv preprint arXiv:1708.06145, 2017.

[9] F. Xu, Z. Tu, Y. Li, P. Zhang, X. Fu, and D. Jin, "Trajectory recovery from ash: User privacy is not preserved in aggregated mobility data," in Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2017, pp. 1241–1250.

[10] F. Eigner, A. Kate, M. Maffei, F. Pampaloni, and I. Pryvalov, "Differentially private data aggregation with optimal utility," in Proceedings of the 30th Annual Computer Security Applications Conference. ACM, 2014, pp. 316–325.