A literature search assignment: Mobile phones, location, contact tracing and epidemics

Niklas Åsberg, Tim Björk and Daniel Alm Lundin
Department of Mathematics and Computer Science
Karlstad University
Karlstad, Sweden
Email: {niklas.olle.asberg, timpish29, danielahlm96}@gmail.com

***Abstract*—*Objective*: To identify, characterize and classify scientific articles regarding privacy friendly methods of mobile infection tracing (protocols, privacy concepts) as well as threats to citizens from such smartphone health surveillance infrastructures.**

***Method*: A literature review based on a keyword search and snowballing process were a inclusion/exclusion and classification criteria is presented and derived articles are evaluated from said criteria.**

***Results*: Results show that fully developed, peer-reviewed privacy friendly contact tracing solutions are scarce in this field but a huge quantity of newly published none peer-reviewed papers are being published as the need of stopping the spread of COVID-19 in a privacy friendly way is a huge focus for governments and researchers all over the globe.**

***Conclusion*: Based on the results, it is concluded that privacy is a rather recent addition to contact tracing as most of the solutions proposed in later articles did not specifically focus on privacy.**

I. Introduction

The pandemic COVID-19 has caused severe global economic disruption as well as causing hundreds of thousands of deaths world wide. Governments and researchers all over the globe are rushing to find a solution to stop the spread of the virus. A prominent approach to stop the spread is the use of contact tracing based on e.g location data and other personal data. However, this approach imposes great privacy endangering threats to citizens in the form of mass surveillance. This literature search is performed to find and classify privacy friendly methods of mobile infection tracing (protocols, privacy concepts) and threats to citizens from such smartphone health surveillance infrastructures. The following sections presents the inclusion/exclusion criteria, classification criteria, keyword search process, snowballing procedure and finally the classification of the definitive articles, a discussion of the result from the classification and the conclusion.

II. Criteria

A. Classification Criteria

These are the classification criteria:

[b]

Category	Description
Solution Proposal	The study provides a protocol or tool related to contact tracing with mobile applications • Focused on privacy • Not focused on privacy The study provides content regarding threats to citizens from mobile surveillance infrastructure
Philosophical	The study is related to the area of interest and discusses about, without presenting or evaluating solutions nor using techniques or methods, the topic of interest.
Experiment	The study focuses on investigating an area of interest with regards to contact tracing with mobile applications.
Case Study	The study focuses on a lesser portion of a problem related to the topic, where the boundaries of the context are relatively unclear.
Threat model	The study presents threats to citizens from smartphone health surveillance infrastructures

B. Inclusion-exclusion criteria

Every article found went through the same process where first of all, the title/abstract had to match a couple of keywords, such as “contact tracing” and “privacy”. If it did not, the article could be ruled out. Next, if the article passed the first criteria, the introduction and/or abstract was thoroughly read

and the rest of the article was skimmed. Moreover, the date of publicity was checked to see if it was "up to date". If the article was found interesting, well written and relevant enough, the reference was written down for further use with a short describing text of the article.

III. Keyword search

The opening selection of primary studies were given as an example by the task description as well as gathered by querying Google Scholar and KAU library. These choices of databases were based on the fact that they present a good way of avoiding bias in favour of any specific publishers. A search string, consisting of an expression in conjunctive normal form, was then applied in the gathering of the opening section of primary studies. The literals of the expression were keywords covering the important aspects that relevant articles should cover and address. The keywords aimed to find relevant articles were: privacy, mobile, infection, contact tracing, and threat. The search string was then adapted to comply with each query language syntax of the databases. In order to minimize the result of the initial search, the expressions were formed to only yield articles that include all the keywords. The initial search yielded 649 articles and was conducted may 7 2020.

The initial set of articles derived from the aforementioned method were screened for relevance based on various exclusion and inclusion criteria. The prominent inclusion criteria in the keyword search were that the title must mention an approach to achieve privacy friendly contact tracing methods in a way that convinced the author performing the screening that the scope of the article is in the area of interest. In total, 14 articles were found from the keyword search that were further investigated with additional exclusion/inclusion criteria. The 14 articles are:

1) Ekong, I., Chukwu, E. and Chukwu, M., April 2020. COVID-19 Mobile Positioning Data Contact Tracing and Patient Privacy Regulations: Exploratory Search of Global Response Strategies and the Use of Digital Tools in Nigeria. JMIR mHealth and uHealth, 8(4), p.e19139.
2) Altuwaiyan, T., Hadian, M. and Liang, X., 2018, May. EPIC: Efficient Privacy-Preserving Contact Tracing for Infection Detection. In 2018 IEEE International Conference on Communications (ICC) (pp. 1-6). IEEE.
3) Abeler, J., Bäcker, M., Buermeyer, U. and Zillessen, H., 2020. COVID-19 Contact tracing and data protection can go together. JMIR mHealth and uHealth, 8(4), p.e19359. April 2020
4) Roomp, K. and Oliver, N., 2020. ACDC-Tracing: Towards Anonymous Citizen-Driven Contact Tracing. arXiv preprint arXiv:2004.07463.
5) Kuhn, C., Beck, M. and Strufe, T., 2020. Covid Notions: Towards Formal Definitions–and Documented Understanding–of Privacy Goals and Claimed Protection in Proximity-Tracing Services. arXiv preprint arXiv:2004.07723.
6) Chan, J., Gollakota, S., Horvitz, E., Jaeger, J., Kakade, S., Kohno, T., Langford, J., Larson, J., Singanamalla, S., Sunshine, J. and Tessaro, S., 2020. Pact: Privacy sensitive protocols and mechanisms for mobile contact tracing. arXiv preprint arXiv:2004.03544.
7) Shukla, M., Lodha, S., Shroff, G. and Raskar, R., 2020. Privacy Guidelines for Contact Tracing Applications. arXiv preprint arXiv:2004.13328.
8) Liu, J.K., Au, M.H., Yuen, T.H., Zuo, C., Wang, J., Sakzad, A., Luo, X. and Li, L., Privacy-Preserving COVID-19 Contact Tracing App: A Zero-Knowledge Proof Approach.
9) Trieu, N., Shehata, K., Saxena, P., Shokri, R. and Song, D., 2020. Epione: Lightweight contact tracing with strong privacy. arXiv preprint arXiv:2004.13293.
10) Bell, J., Butler, D., Hicks, C. and Crowcroft, J., 2020. Tracesecure: Towards privacy preserving contact tracing. arXiv preprint arXiv:2004.04059
11) Beskorovajnov, W., Dörre, F., Hartung, G., Koch, A., Müller-Quade, J. and Strufe, T., 2020. Contra corona: Contact tracing against the coronavirus by bridging the centralized–decentralized divide for stronger privacy. Cryptology ePrint Archive, Report 2020/5/05.
12) Tang, Q., 2020. Privacy-Preserving Contact Tracing: current solutions and open questions. arXiv preprint arXiv:2004.06818.
13) Hekmati, A., Ramachandran, G. and Krishnamachari, B., 2020. CONTAIN: privacy-oriented contact tracing protocols for epidemics. arXiv preprint arXiv:2004.05251.
14) Yasaka, T.M., Lehrich, B.M. and Sahyouni, R., 2020. Peer-to-Peer Contact Tracing: Development of a Privacy-Preserving Smartphone App. JMIR mHealth and uHealth, 8(4), p.e18936.

The 14 articles that were found within the prominent inclusion criteria were then screened if they were published via a peer-reviewed channel or not, resulting in the exclusion of the articles 6, 7, 8, 9, 10, 11, 12 and 13. A second round of key wording was then performed on the abstracts of the remaining relevant articles . The abstracts of 1, 2, 3, 5, 14 were thought to be adequate enough in length and insightfulness to not extend the keyword search into the introduction and conclusion section. However, the examination of the fourth candidate was extended into the introduction section as it did not precisely mention privacy in the abstract. After extending the keyword search into the introduction the conclusion was made that it did in fact fulfill the inclusion criteria. Compared to the results of the start set in the article [1], the results from the aforementioned keyword search is more diverse as the resulting 6 articles are from different publishers and have no author in common, thus removing biases. However, the search did not resulting in a great diversity of publication years as the majority of articles were published in April or May 2020.

Resulting articles from the keyword search:

1) Ekong, I., Chukwu, E. and Chukwu, M., April 2020. COVID-19 Mobile Positioning Data Contact Tracing and Patient Privacy Regulations: Exploratory Search of Global Response Strategies and the Use of Digital Tools

in Nigeria. JMIR mHealth and uHealth, 8(4), p.e19139.
2) Altuwaiyan, T., Hadian, M. and Liang, X., 2018, May. EPIC: Efficient Privacy-Preserving Contact Tracing for Infection Detection. In 2018 IEEE International Conference on Communications (ICC) (pp. 1-6). IEEE.
3) Abeler, J., Bäcker, M., Buermeyer, U. and Zillessen, H., 2020. COVID-19 Contact tracing and data protection can go together. JMIR mHealth and uHealth, 8(4), p.e19359. April 2020
4) Roomp, K. and Oliver, N., 2020. ACDC-Tracing: Towards Anonymous Citizen-Driven Contact Tracing. arXiv preprint arXiv:2004.07463.
5) Kuhn, C., Beck, M. and Strufe, T., 2020. Covid Notions: Towards Formal Definitions–and Documented Understanding–of Privacy Goals and Claimed Protection in Proximity-Tracing Services. arXiv preprint arXiv:2004.07723.
6) Yasaka, T.M., Lehrich, B.M. and Sahyouni, R., 2020. Peer-to-Peer Contact Tracing: Development of a Privacy-Preserving Smartphone App. JMIR mHealth and uHealth, 8(4), p.e18936.

IV. Snowballing

The snowballing technique was used to find five articles both forward and backward from the initial set of articles given in the assignment description. The goal was to iterate twice, both forward and backward, in order to find a total of five articles in each direction.

A. Forward Snowballing

The initial set of articles were used and searched for in Google Scholar, in combination with the "Cited by" option. Depending on which article was searched for, the results varied a lot. Some of them had not been cited at all, whereas some of them were cited over 50 times. However, the amount of references found with the "Cited by" function did not, in some cases, matter because of the relevance of the references. A few of the references which had been cited many times, did not present a single relevant article.

To sort out which of the references should be used, the aforementioned inclusion-exclusion criteria were used. In some cases, the articles were locked, which meant they had to be searched for in some other scientific database through "kau.se" to get full access. The timestamp for the forward snowballing was fine for every reference.

When all references were gone through, a comparison had to take place because of the task specification. The specification said that only five references were to be produced in each snowballing direction, which meant that the 10 found articles had to be cut down to 5. To do so, the describing texts were compared which ruled out a few articles. There was a discussion about the remaining articles where pros and cons were presented. Then, a vote had to take place where all 3 parties voted which ones were best suited and most relevant for this particular topic.

One of the criteria for the article search was that the article had to have been peer reviewed, which you can not see in google scholar. Therefore, the articles that was found relevant had to be searched for in the KAU library where there is a setting to search for only peer reviewed articles. For the forward snowballing that became a problem because second to none of the articles were peer reviewed. One of the reasons for that could be because of the fact that the subject of contact tracing and privacy preservation in mobile health apps is a contemporary problem. Consequently, hundreds(if not more) of new articles about that exact problem are published every week, which means very few of them will be peer reviewed. This meant that, to produce any references, the criteria of peer review had to be disregarded.

Round two of the forward snowballing was similar to the first round but the snowballing was based on the references produced from the first round, although since the work was divided over the 3 parties, the voting phase was skipped.

The five definitive references were:

1) Liu, J.K., Au, M.H., Yuen, T.H., Zuo, C., Wang, J., Sakzad, A., Luo, X. and Li, L., Privacy-Preserving COVID-19 Contact Tracing App: A Zero-Knowledge Proof Approach
2) Hekmati, A., Ramachandran, G. and Krishnamachari, B., 2020. CONTAIN: privacy-oriented contact tracing protocols for epidemics. arXiv preprint arXiv:2004.05251.
3) Demirag, D. and Ayday, E., 2020. Tracking and controlling the spread of a virus in a privacy-preserving way. arXiv preprint arXiv:2003.13073
4) Berke, A., Bakker, M., Vepakomma, P., Raskar, R., Larson, K. and Pentland, A., 2020. Assessing disease exposure risk with location histories and protecting privacy: A cryptographic approach in response to a global pandemic. arXiv preprint arXiv:2003.14412
5) Cho, H., Ippolito, D. and Yu, Y.W., 2020. Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs. arXiv preprint arXiv:2003.11511.

B. Backward Snowballing

The first step of backward snowballing was to screen as mentioned by first looking at the title of the references in the lists of references in the initial articles. This step yielded 12 candidate articles.

The articles were then further screened by looking at the abstract with consideration to the inclusion-exclusion criteria. At this step it appeared that the overall level of privacy focus in these types of articles got lower the further back in time they were written. Consequently, most of the articles found in the first step failed to pass our initial criteria, which in turn led to a modification of the thought behind the inclusion-exclusion criteria. Instead of concurrently requiring a privacy oriented solution and the threats concerning these, it was sufficient for an article to mention one of these subjects in a way that could be useful on its own or along with other included articles.

Along with this step, the articles were also checked via the KAU library to make sure that they had been peer-reviewed.

Conducting all of these steps resulted in keeping four articles from the first iteration.

For the second round, the process was iterated again on the four articles from round one. The second iteration yielded an additional set of eight articles. These articles were then screened much in the same way as earlier. Some of these articles had to be ruled out for reasons such as for example having the same authors as articles from the first round. The result was three candidate articles. The last step was to compare the three articles to those from the first iteration. At least one article was going to be kept without replacing another, but for the rest a decision had to be made whether the second iteration articles should replace one from the first round. The end result was that one article from the second round was kept and none of the four from the first iteration removed.

The final five articles were:

1) K. Zhang, X. Liang, J. Ni, K. Yang, and X. Shen, "Exploiting social network to enhance human-to-human infection analysis without privacy leakage," IEEE Transactions on Dependable and Secure Computing, 2016.
2) Sacks, J.A., Zehe, E., Redick, C., Bah, A., Cowger, K., Camara, M., Diallo, A., Gigo, A.N.I., Dhillon, R.S. and Liu, A., 2015. Introduction of mobile health tools to support Ebola surveillance and contact tracing in Guinea. Global Health: Science and Practice, 3(4), pp.646-659.
3) De Montjoye, Yves-Alexandre and Hidalgo, César A and Verleysen, Michel and Blondel, Vincent D, Unique in the crowd: The privacy bounds of human mobility, Scientific reports, Vol. 3, 2013, Nature Publishing Group.
4) Iyengar A, Kundu A, Pallis G. Healthcare informatics and privacy. IEEE Internet Computing. 2018;22(2):29-31.
5) Zhou, J., Lin, X., Dong, X. and Cao, Z., 2014. PSMPA: Patient self-controllable and multi-level privacy-preserving cooperative authentication in distributed-healthcare cloud computing system. IEEE Transactions on Parallel and Distributed Systems, 26(6), pp.1693-1703.

V. Classification result

A. Solution proposal criteria

A couple of articles [2] [3] [4] present thorough solutions, such as applications for privacy-oriented protocols for contact tracing in epidemics whereas some, eg. [5] [6] [7] [8] [9], present more general solutions to avoid privacy violations. Some of these articles do not take privacy into account when presenting their solutions, such as [10] [11]. The fourth variation of articles are those that discuss, analyze and evaluate already created solutions.

B. Philosophical criteria

Some articles are well written and are very much related to the topic of interest but do not present solutions. Instead they discuss the different guidelines [12], considerations [13] or the subject in a general sense [14] [15]

C. Experiment

There are also articles which are more focused on investigating a certain method which concerns the area of interest, such as [16]

D. Threat model

All of the final articles presented and discussed general threats, such as misuse and leakage, associated with storing personal data e.g location data and medical data. The article [3] mentioned potential threats regarding false positives and false negatives in confirmed cases of the disease. [17] and [13] analyzed the associated threats with the applications and protocols used in China and South Korea. Both articles concluded that the solutions used in these countries protects the privacy of users from each other, but these solutions had serious privacy concerns with respect to the government access to the users data. The article [15] also discussed additional threats regarding the use of pseudonyms and how pseudonyms would enable adversaries to connect other data to the pseudonyms and thus identifying original users. [15] also mentioned the threat of leaking aggregate information as it could cause stigmatization of certain behavior, areas or companies/shops.

VI. Discussion

As shown in this paper, there are two different ways of article search; snowballing and database search(screening). Both of these methods are great ways of doing so, but there are situations where one is better than the other. One of snowballing's main advantages [1] is that you start from articles which are relevant, and from there you snowball in different directions which means the articles found will very often stay relevant. Furthermore, with the "cited by" function in google scholar the number of articles produced are often large if the papers are highly cited, resulting in examination of many papers. Although, with proper inclusion-exclusion criteria, the large number of articles becomes a big advantage.

When performing snowballing, one disadvantage is that it is possible to get caught in a smaller set of articles which refer to one another, depending on the initial articles from which the snowballing is performed. Instead, if the method of database search is used, the retrieved articles will not depend on another article, but the keywords used for the search. This will prevent getting caught in a small set and will grant a large variety of authors [1].

One of the criteria for the article search was that the article had to have been peer reviewed, which you can not see in google scholar. Therefore, the articles that were found relevant had to be searched for in the KAU library where there is a setting to search for only peer reviewed articles. For the forward snowballing that became a problem because second to none of the articles were peer reviewed. One of the reasons for that is because of the fact that the subject of contact tracing and

privacy preservation in mobile health apps is a contemporary problem. Consequently, hundreds(if not more) of new articles about that exact problem are published every week, which means very few of them will be peer reviewed. This meant that, to produce any references, the criteria of peer review had to be disregarded. The aforementioned argument could also be the reason that the keyword search only resulted in articles that were not published later than 2018/2019.

When trying to control for example an Ebola outbreak in Africa several years ago, privacy was not considered as important as it is today. Furthermore, the technique is more developed today, which has led to an increased amount of these applications and consequently more discussion about the privacy aspect of such applications. The result of the backwards snowballing is not what we initially anticipated or aimed for. Instead, the changed approach led to a more diverse set of articles that could either be relevant on their own or as complement to each other.

VII. Conclusion

From this study we queried mainly google scholar and the KAU library in order to find articles relevant to the subject. As the amount of articles we aimed to find was relatively low, we achieved the goal of finding in total 10 articles from using the snowballing technique and an additional six from performing a keyword search. The articles were categorized which showed that even with a low number of articles they included several different aspects of privacy in contact tracing.

The latest articles found from, especially forward snowballing, shows that the interest in privacy friendly contact tracing has increased significantly in the latest months compared to a couple of years ago. This could be expected considering the present-day world pandemic situation. This study shows that older solutions only focused on contact tracing still can be useful for designing new privacy friendly methods when complemented by for example more philosophical articles focused mainly on privacy.

The current amount of interest in the subject is promising for the future considering the ongoing development. In the case of a similar situation to the COVID-19 outbreak, the progress made in the articles we found can prove to be useful to have at the beginning of the outbreak.

[4] T. Altuwaiyan, M. Hadian, and X. Liang, "Epic: Efficient privacy-preserving contact tracing for infection detection," in *2018 IEEE International Conference on Communications (ICC)*. IEEE, July 2018, pp. 1–6.

References

[1] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, 2014, pp. 1–10.

[2] A. Hekmati, G. Ramachandran, and B. Krishnamachari, "Contain: privacy-oriented contact tracing protocols for epidemics," *arXiv preprint arXiv:2004.05251*, April 2020.

[3] J. K. Liu, M. H. Au, T. H. Yuen, C. Zuo, J. Wang, A. Sakzad, X. Luo, and L. Li, "Privacy-preserving covid-19 contact tracing app: A zero-knowledge proof approach," May 2020.

[5] D. Demirag and E. Ayday, "Tracking and controlling the spread of a virus in a privacy-preserving way," *arXiv preprint arXiv:2003.13073*, Mars 2020.

[6] K. Zhang, X. Liang, J. Ni, K. Yang, and X. S. Shen, "Exploiting social network to enhance human-to-human infection analysis without privacy leakage," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 607–620, November 2016.

[7] A. Berke, M. Bakker, P. Vepakomma, R. Raskar, K. Larson, and A. Pentland, "Assessing disease exposure risk with location histories and protecting privacy: A cryptographic approach in response to a global pandemic," *arXiv preprint arXiv:2003.14412*, April 2020.

[8] I. Ekong, E. Chukwu, and M. Chukwu, "Covid-19 mobile positioning data contact tracing and patient privacy regulations: Exploratory search of global response strategies and the use of digital tools in nigeria," *JMIR mHealth and uHealth*, vol. 8, no. 4, p. e19139, april 2020.

[9] K. Roomp and N. Oliver, "Acdc-tracing: Towards anonymous citizen-driven contact tracing," *arXiv preprint arXiv:2004.07463*, April 2020.

[10] J. A. Sacks, E. Zehe, C. Redick, A. Bah, K. Cowger, M. Camara, A. Diallo, A. N. I. Gigo, R. S. Dhillon, and A. Liu, "Introduction of mobile health tools to support ebola surveillance and contact tracing in guinea," *Global Health: Science and Practice*, vol. 3, no. 4, pp. 646–659, December 2015.

[11] J. Zhou, X. Lin, X. Dong, and Z. Cao, "Psmpa: Patient self-controllable and multi-level privacy-preserving cooperative authentication in distributedm-healthcare cloud computing system," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 6, pp. 1693–1703, June 2014.

[12] M. Shukla, S. Lodha, G. Shroff, R. Raskar *et al.*, "Privacy guidelines for contact tracing applications," *arXiv preprint arXiv:2004.13328*, April 2020.

[13] H. Cho, D. Ippolito, and Y. W. Yu, "Contact tracing mobile apps for covid-19: Privacy considerations and related trade-offs," *arXiv preprint arXiv:2003.11511*, April 2020.

[14] A. Iyengar, A. Kundu, and G. Pallis, "Healthcare informatics and privacy," *IEEE Internet Computing*, vol. 22, no. 2, pp. 29–31, March 2018.

[15] C. Kuhn, M. Beck, and T. Strufe, "Covid notions: Towards formal definitions–and documented understanding–of privacy goals and claimed protection in proximity-tracing services," *arXiv preprint arXiv:2004.07723*, April 2020.

[16] Y.-A. De Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," *Scientific reports*, vol. 3, p. 1376, March 2013.

[17] J. Abeler, M. Bäcker, U. Buermeyer, and H. Zillessen, "Covid-19 contact tracing and data protection can go together," *JMIR mHealth and uHealth*, vol. 8, no. 4, p. e19359, April 2020.