Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
On privacy enhancement using \( u \)-indistinguishability to COVID-19 contact tracing approach in Korea

Hyoung Joong Kim\(^1\), Suah Kim\(^1\), Sungho Lee\(^2\)

\(^1\)GRADUATE SCHOOL OF INFORMATION SECURITY, KOREA UNIVERSITY, SEOUL, SOUTH KOREA; \(^2\)SCHOOL OF MEDICINE, KOREA UNIVERSITY, SEOUL, SOUTH KOREA

1. Introduction

It is widely agreed that South Korea successfully managed its COVID-19 outbreak [1]. As of April 25, 2020, there had been 10,718 confirmed cases of COVID-19 in South Korea, with a total of 240 deaths [2]. The number of daily confirmed patients had been suppressed by about 10 people, mostly arriving from abroad since March 18, 2020 [3]. In early May, a young man who is believed to be one of the individuals behind the new COVID-19 cluster visited several bars and nightclubs tested positive. He is one of the latest super spreaders. As a result, quarantine authorities are still on high alert with resulting spread of confirmed cases among some nightclub patrons, but the public has full confidence in the ability of Korea Centers of Disease Control and Prevention (KCDC) to quickly stabilize the situation. This country has flattened its coronavirus curve by adopting an aggressive trace, test, and treat strategy. It is notable that South Korea calmed the coronavirus outbreak without lockdowns, roadblocks, and restrictions on movements. In addition, South Korea held a nationwide election on April 15, 2020 in the midst of a pandemic without a single one of the more than 29 million voters getting infected with the novel virus afterward [3,4]. During the COVID-19 pandemic situation, both the Korea Baseball Organization and K-League started regular season, and their games have been broadcast live abroad.

After the success of South Korea and Singapore in controlling COVID-19, many countries are now considering introducing contact tracing systems. Contact tracing is the process of identifying carriers of disease, and with whom they have come into contact, and let them quarantine as needed [5]. The contact tracing system of KCDC is a centralized system (Fig. 36.1 shows the data used by KCDC). This KCDC system is more comprehensive because it uses not only smartphone logs but also other data like credit card transactions and CCTV footage. Needless to say, privacy concerns and controversies
have been raised [6]. South Korea’s Minister of Foreign Affairs, Kang Kyung-wha, said that “privacy is a very important human right, but it is not an absolute right” and “we have a very robust legal system in place where it is clearly identified as to where these rights might be restricted” [7]. In any case, patient information sharing must be prudent [8,9].

Pan-European Privacy-Preserving Proximity Tracing (PEPP-PT) and Decentralized Privacy-Preserving Proximity Tracing (DP-3T) make use of Bluetooth Low Energy (BLE) only to discover and locally log smartphone users in close proximity of an infected patient. The PEPP-PT protocol conducts two functionalities: logging and reporting. The protocol with BLE module anonymously detects close encounters within six feet, and keeps the log in a user’s phone. The user can reports that he/she has tested positive and upload its log of contacts to the competent health authorities. The protocol adopts a centralized reporting server to process contact logs and notify users of potential contact with an infected patient [10]. On the other hand, in the DP-3T protocol, the health authority never has access to contact logs, and only serve to test patients and authorize report submissions [11]. Since the infection reports are processed on the client side.

The problem with the KCDC tracing system is that it is centralized and too detailed. This paper will address countermeasures to the detailed information, leaving aside the centralization challenge for another discussion. The more detailed the information that is disclosed, the more useful the data are, but the greater the degree of privacy breach. Several solutions have been proposed, but they are not yet fully formed and are not suitable for use in contact tracing. These include the privacy enhancing technologies of differential privacy, federated analysis, homomorphic encryption, zero-knowledge proof, and multiparty computation [12,13]. Furthermore, the quasi-identity (QID) attributes blurring method called $k$-anonymity [14], sensitive data mixing method called $l$-diversity [15], and $t$-closeness [16] have also been proposed. These last three techniques specifically play an important role in privacy enhancement for medical data. However, the privacy breach that has resulted from KCDC contact tracing data is a unique situation, which requires a new technique different from these three standard approaches. This paper will examine the privacy breach problems of the KCDC COVID-19 response and propose a $u$-indistinguishability concept and demonstrate its effectiveness to solve this problem.

![FIGURE 36.1 Information flow of COVID-19 contact tracing system in South Korea.](image)
2. Related technologies

Samarati [17] and Sweeney [14] introduced the \( k \)-anonymity concept. This measure makes each record indistinguishable with at least \( k - 1 \) other attributes. In other words, \( k \)-anonymity allows each equivalence class to have at least \( k \) attributes. Table 36.1 shows an example with nine patients. Attackers will try to reidentify the patients by associating Table 36.1 with other databases like the voter registration list as Sweeney did [14]. To make reidentification difficult, suppression and generalization techniques are used. The suppression technique replaces the lower value with \("*"\) so that \( k \) attributes can be expressed with the same value. In Table 36.1, from the top, the three patients are in their 20s. The suppression technique replaces the first digit of the age of these three patients with \("*"\) (See Table 36.2 of the top three patients). The generalization technique replaces the individual values of attributes with a broader category. In Table 36.1, from the bottom, their age is over 40. The generalization technique replaces their age more broadly with \(">40"\) (See Table 36.2 of the bottom three patients). The suppression is a special case of the generalization with a narrower range. For example, the age range \("2*"\) of the suppression is equivalent to \("20–29"\) of the generalization. Each QID tuple in

| Age | ZIP   | Disease       |
|-----|-------|---------------|
| 1   | 29    | 12345         | Cancer       |
| 2   | 22    | 12345         | Cancer       |
| 3   | 25    | 12346         | Cancer       |
| 4   | 32    | 23456         | Heart disease|
| 5   | 36    | 23459         | Stomach ulcer|
| 6   | 34    | 23491         | Heart disease|
| 7   | 42    | 54321         | Epilepsy     |
| 8   | 51    | 54322         | Flu          |
| 9   | 56    | 54320         | Angina pectoris|

**Table 36.1** An original patients table.

| Age | ZIP   | Disease       |
|-----|-------|---------------|
| 1   | 2*    | 1234*         | Cancer       |
| 2   | 2*    | 1234*         | Cancer       |
| 3   | 2*    | 1234*         | Cancer       |
| 4   | 3*    | 234**         | Heart disease|
| 5   | 3*    | 234**         | Stomach ulcer|
| 6   | 3*    | 234**         | Heart disease|
| 7   | >40   | 5432*         | Epilepsy     |
| 8   | >40   | 5432*         | Flu          |
| 9   | >40   | 5432*         | Angina pectoris|

**Table 36.2** A 3-anonymity version of Table 36.1.
Table 36.2 occurs in at least three attributes with 3-anonymity. Thus, 3-anonymity is achieved in this manner. A more extensive research results on $k$-anonymity can be found in Ref. [18].

Two attack models against $k$-anonymity are proposed: homogeneity attack and background knowledge attack [15]. The homogeneity attack leverages the case where all the values for sensitive information within a set of $k$ attributes are similar or identical. Given the same disease name (i.e., cancer of top three patients in Table 36.2), the attacker may guess that Alice, 29, living in an area with ZIP code 12345, had cancer. The background knowledge attack increases the accuracy of inference using background information about the patient. For example, Bob has a friend who is Japanese (of age 36 with ZIP code 23459), who has been admitted to a hospital. However, Bob does not know what illness his friend was hospitalized for. Assume that it is a well-known fact that Japanese have an extremely low chance of heart disease. Then, Bob can guess that his friend is suffering from a stomach ulcer. To beat these attack scenarios, $l$-diversity concept has been proposed [15].

The $l$-diversity principle lets every equivalence class contain at least $l$ sensitive values. The top three patients in Table 36.2 suffer from cancer, a single disease case. However, the top three patients in Table 36.4 suffer from three different diseases; skin cancer, 

### Table 36.3 An original patients table.

| Age | ZIP    | Disease          |
|-----|--------|------------------|
| 1   | 29     | 12345 Skin cancer|
| 2   | 22     | 12345 Psoriasis   |
| 3   | 25     | 12346 Dermatitis  |
| 4   | 32     | 12346 Arrhythmia  |
| 5   | 36     | 12344 Angina pectoris |
| 6   | 34     | 12333 Arteriosclerosis |
| 7   | 24     | 12354 Gastritis   |
| 8   | 32     | 12,349 Stomach ulcer |
| 9   | 49     | 12347 Stomach cancer |

### Table 36.4 A 3-diversity version of Table 36.3.

| Age | ZIP    | Disease          |
|-----|--------|------------------|
| 1   | 2*     | 1234* Skin cancer |
| 2   | 2*     | 1234* Psoriasis   |
| 3   | 2*     | 1234* Dermatitis  |
| 4   | 3*     | 123* Arrhythmia   |
| 5   | 3*     | 123** Angina pectoris |
| 6   | 3*     | 123** Arteriosclerosis |
| 7   | 20—50  | 123** Gastritis   |
| 8   | 20—50  | 123** Stomach ulcer |
| 9   | 20—50  | 123** Stomach cancer |
psoriasis, and dermatitis, respectively. Similarly, the middle three and bottom three patients have different diseases. It shows that 3-diversity is achieved in Table 36.4. However, the homogeneity attack model is also possible against \( l \)-diversity anonymization. In Table 36.4, given the similar kind of diseases, the attacker may guess that Mary, 22, living in an area with ZIP code 12345, is suffering from gastrointestinal disease. To beat this attack scenario, the \( t \)-closeness concept has been proposed [16].

The \( t \)-closeness model treats the values of an attribute distinctly by taking into account the distribution of data values for that attribute. An equivalence class is said to have \( t \)-closeness if the distance between the distribution of a sensitive attribute in this class and the distribution of the attribute in the whole database is no more than a threshold \( t \). The \( t \)-closeness model is similar to the \( l \)-diversity model in terms of the diversity of the sensitive attributes. However, the former is stricter than the latter because that considers the distribution of the attributes. Table 36.5 shows that the top three patients’ illnesses are diverse, and their distribution has entries of dermatology, respiratory, and gastroenterology. So do the middle three and the bottom three patients. It shows that 3-closeness is achieved in Table 36.5.

A new type of supplementing technique for the \( k \)-anonymity principle blends a number of dummy identities with fake location records [19] in location-based service applications. This approach is appealing because it effectively achieves \( k \)-anonymity without sacrificing data quality [20]. Applications of the blending methods like \( k \)-anonymity, \( l \)-diversity, and \( t \)-closeness can be found in Refs. [21–23].

3. Contact tracing in South Korea

South Korea’s COVID-19 policy is considered effective, but unique. The uniqueness claim stems from privacy infringement concerns on the contact tracing and movement data sharing. Epidemiological investigators get real-time data feed from the KCDC on patients. After the investigation, KCDC releases the patients’ movement information including their whereabouts, transportation used, people they contacted, times spent at
specific locations, CCTV footage, credit card transactions, and immigration reports, if any. KCDC also releases patients’ personal information including gender, nationality, age, routes of infection, date of confirmation, name of hospitals admitted, number of contacts, and movement information. Local governments add residential area information which is equivalent to the ZIP code in the United States. The KCDC released the movement information release guidelines on March 14, 2020 and revised it on April 2, 2020 [24].

To deter a reidentification attack, a quasi-identifier masking technique called \( k \)-anonymity [14,25,26] has been proposed. Sensitive attribute blending technique called \( l \)-diversity [15,22] and sensitive attribute distribution assimilating technique called \( t \)-closeness [16,22] have also been proposed. However, none of these techniques have been applied to the KCDC patient tracking approach. KCDC QID includes gender, age (or year of birth), nationality, routes of infection, date of confirmation, admitted hospital name (See Table 36.6), and residential area name added by local government (See Table 36.7). KCDC sensitive attributes are not seen in Table 36.6, but is obviously COVID-19 itself.

Movement information contains more detailed data including on times, places, and activities (See Table 36.7). Since April 12, 2020, the KCDC guidelines have changed to release the movement information only for two weeks [24]. The movement information does not belong to the category of sensitive attributes as aforementioned. The movement data part (second column in Table 36.7) is called an auxiliary attribute in this paper. Auxiliary attributes include five pieces of information: WHO did WHAT, WHEN, WHERE, and BY. The WHO attribute is not explicitly in the auxiliary section, but is in the QID part.

### Table 36.6 Patient information revealed by Korea Centers of Disease Control and Prevention (in part).

| Gender (nationality, Patient year of birth) | Routes of infection                        | Date of confirmation | Admitted hospital | # Of contacts (# of isolates) |
|---------------------------------------------|--------------------------------------------|----------------------|-------------------|-----------------------------|
| 47 Female (Korean, 1957)                     | Contacted patient #31 (under investigation)| 2/19                 | DMC               | Investigating               |
| 31 Female (Korean, 1959)                     | Investigating (investigating)              | 2/18                 | DMC               | 1160 (1160)                 |
| 28 Female (Chinese, 1989)                    | Contacted patient #3 (secondary infection) | 2/10                 | Discharged        | 1 (1)                       |
| 12 Male (Chinese, 1971)                      | Contacted patient in Japan (secondary infection) | 2/1                  | SNUH B            | 422 (0)                     |
| 6 Male (Korean, 1964)                        | Contacted patient #3 (secondary infection) | 1/30                 | SNUH              | 17 (1)                      |
| 3 Male (Korean, 1966)                        | Visited Wuhan (primary infection)          | 1/24                 | Discharged        | 16 (0)                      |
| 1 Female (Chinese, 1984)                     | Visited Wuhan (primary infection)          | 1/20                 | Discharged        | 45 (0)                      |

Captured and translated into English on April 27, 2020 at [http://ncov.mohw.go.kr/bdBoardList_Real.do](http://ncov.mohw.go.kr/bdBoardList_Real.do).
Table 36.7  Example of movement information revealed by three local governments.

| Patient | Personal information | Movement details |
|---------|----------------------|------------------|
| #63 of Gangnam, Female Seoul | 39 years old | April 24 to 25: Stay home (disinfection finished; three contacts under self-quarantine) April 26: 08:50 visited a café near subway station exit #1 (disinfection finished; under investigation) 09:10 tested at the public health center 10:00 hung around ATM near subway station (disinfection finished; under investigation) 10:30 stay home (disinfection finished) 11:30 visited a restaurant near substation exit #3 (disinfection finished; under investigation) 16:00 stayed home (disinfection finished) April 27, 2020 08:00 confirmed positive |
| #31 of Dongjak, Female Seoul | 20s living in Heukseokdong | April 13 08:00 confirmed positive |
| #16 of Junggu, Male Seoul (#10028 of KCDC) | Born in 1991 American Performer contacted #9864 (Canadian) | March 11 Arrived at Incheon International Airport and went to a hotel. March 12–29 Stayed at the hotel and went to and from a performance venue in Yongsan March 31 09:30 visited the breakfast lounge in the second floor of the hotel 16:00 visited rooftop of the hotel 18:00–20:00 performed at the venue in Yongsan 23:15 returned to the hotel April 1 Tested at the public health center and self-quarantined April 2 Confirmed positive and moved to a hospital |

Captured and translated into English on April 27, 2020 at https://www.gangnam.go.kr/index.htm, https://dongjak.go.kr/, and https://www.jongno.go.kr/Main.do?menuId=400516&amp;menuNo=400516, respectively, from the top to the bottom.

The WHAT attribute is also not there in the auxiliary section, but is implied. The WHEN and WHERE attributes are explicitly in the auxiliary section. The BY attribute represents a means of transportation (i.e., by taxi, by subway). If there is no BY attribute, it means that the patient has moved on foot.
Activity information (i.e., WHAT) guessed from the movement data is potentially very harmful to the patient [27]. For example, patient #3 in Table 36.6 suffered from a rumor of an extramarital affair. The speculation was derived based on the fact that he accompanied a Chinese woman (patient #28) to a hotel and a plastic surgery clinic twice [28]. The reason why the movement information is dangerous is because the activity information is inferred no matter whether it is true or not. If a man in his 50s and a woman in her 30s went to a hotel together, it is not unusual to assume that this is because of an affair in Korea. This serious misunderstanding began simply because the two people’s movements coincidentally overlapped. Unfortunately, the place and time were the same and led to an inaccurate conclusion.

Clearly, this demonstrates the need for a method that maintains the accuracy of data while preventing inference. Table 36.6 itself is neutral and informative. This table can be used for the demographic study. Super spreaders (i.e., patient #31) can be easily identified based on this table, source of transmission found, incidents of cluster infection detected, and various social network analyses made possible by connecting the dots. Table 36.6 itself does not breach privacy much, but the mobility information associated with patients does. The mobility analysis is similar to the traffic analysis in military intelligence [29]. Attackers can infer so much information from the mobility analysis as demonstrated in the following hypothetical instance. Even if Carol does not know the actual details of the conversation between Alice and Bob, she can deduce the intention of the two from the meeting. Carol can infer that frequent communications or meetings between them denote the planning stage of an action. Observers can guess all sorts of information about the intentions and actions of the target. Table 36.6 contains little mobility information. Table 36.7 is different; this detailed movement data can cause serious privacy infringement. Table 36.7 is the extended version of the movement data provided by the KCDC.

4. Problems of contact data disclosure

The importance of the patient movement data is clear. These data are very useful for the epidemiological investigation and disinfection. However, disclosing these data creates several problems; one of them is the possibility of activity inference. The problem began with a separate release of movement information for each patient. It was inferred that patients #3 and #28 had been having an affair with each other because each movement information overlapped. Because the identities of the two patients (#3 and #28) have not been disclosed and have been pseudonymized, the accuracy of the movement information is not compromised by mixing their mobility data into one cluster. If it is not enough to combine two patients’ mobility data, then bundle three patients’, and if that is not still enough, then mix the $u$ patients of them. The number $u$ represents the number of patients. This $u$ can be adjusted according to the level of security requirement. The solution is simple and easy: anonymity loves company. The answer lies in mixing and unlinking so that QIDs are mixed along with (WHEN, WHERE, BY)-tuples, then, as a result, their linkability is blurred.
Movement information of the patient #63 in Gangnam for four days is reported in Table 36.7. It is easy to identify the name of café or restaurant since Table 36.7 discloses detailed information on location (i.e., name of the subway station which is blinded in this paper and exit number). All the people she contacted are identified. The identity of the contact person is identified through the patient’s statements and CCTV footage or credit card transaction verification. All close contacts should be tested for COVID-19. Disinfection is carried out in the places visited by the patient.

Information for the patient #31 in Dongjak does not disclose much information since the woman followed the self-isolation guidelines well. On the other hand, the patient #16 has stayed a long time in Korea and met many people. In the table, his occupation is identified. Newspapers reported he is an American actor appearing in “The Phantom of Opera.” A total of 8578 audience members who had watched the musical from March 18 to 31 were monitored by the Metropolitan Government of Seoul. Out of the total of 128 cast members, 126 were confirmed negative. A Canadian ballerina (#9864) tested positive on March 31, and the American actor (#10028) confirmed positive on April 2. The performance venue, Blue Square Interpark Hall, was closed after disinfection according to the guidelines. Table 36.7 does not disclose the detailed information surrounding the patient like the venue which is inferred or provided by the competent authorities.

5. \textit{u}-indistinguishability

The basic concept of \textit{k}-anonymity, \textit{l}-diversity, and \textit{t}-closeness is the same: blurriness by mixing. The \textit{u}-indistinguishability is no exception. Note that \textit{u}-indistinguishability is conceptually very similar to \textit{l}-diversity and \textit{t}-closeness. The last two mix the sensitive attributes as much as possible for the prevention of privacy infringement of the \textit{k} subjects. Since there are \textit{k} subjects and \textit{k} sensitive attributes in each cluster, the total number of WHO-WHAT ATTRIBUTE combinations is $O(k^2)$. However, each QID is associated with each sensitive attribute one-by-one. In addition, because of the suppression and generalization techniques, each cluster contains few distinct QIDs. For example, the top three QIDs in Table 36.5 are reduced to one representative QID (i.e., an unknown person of age 20–40 living in the area with ZIP code 123**). As a result, the total number of sensitive attributes pertaining to a particular person is around $O(k)$.

On the other hand, the \textit{u}-indistinguishability blends WHO parts into a cluster, and WHEN and WHERE parts into another cluster, respectively, so that it is difficult to infer WHO did WHAT. Each cluster has at least \textit{u} members of QID and more than or equal to \textit{u} pieces of movement information. The first column of Table 36.8 has two clusters. One cluster has three persons’ information: two males of age 67 and 42, respectively, and a female of age 35. Another cluster also has three persons’ information; two males of age 75 and 60, respectively, and a female of age 40. In this manner, QIDs and its associated auxiliary attributes are clustered for mixing the movement information (see Table 36.8). Table 36.8 shows an example of 3-indistinguishability cases. Each row (or cluster) has at
least three patients and their seven pieces of movement information in the first row and four pieces in the second row. Note that auxiliary attribute mixing eliminates the overlapped movement information, if any, of accompanying people. The patient #3 and #28 in Table 36.6 had visited the same place (a plastic surgery clinic and a hotel) at the same time. Thus, if they are in the cluster, the duplicated data pieces can be removed.

QIDs in the example can be blurred by applying $k$-anonymity. Patients along with the movement information should be carefully selected and mixed based on the philosophy of $l$-diversity and/or $t$-closeness, if possible. The key point of this paper is in the $u$-indistinguishability; at least $u$ patients and their mobility information are mixed into each cluster. It is up to the implementer to apply this technique in more detail. If the number of patients is large, $u$ can be increased. If the number of patients is small, noise can be intentionally added to the movement information. In the cases of $k$-anonymity, $l$-diversity, and $t$-closeness, there is one-by-one mapping between the subjects in the first column and their associated attributes in the second column like Table 36.7. However, the first row of Table 36.8 shows that three subjects in the first column constitute a cluster, and seven attributes in the second column makes another cluster. Similarly, the second row shows a cluster of three subjects and another cluster of four attributes.

Under the premise that KCDC’s disclosure of contact tracing data is standard practice, this paper proposes a method named the $u$-indistinguishability to protect privacy. As is shown in Table 36.8, it is not clear WHO did WHAT. Attackers can infer WHO did WHAT, but their accuracy will drop sharply, especially when $u$ is large. The advantage of $u$-indistinguishability is that it makes it approximately $O(u^2)$ for $u$ patients and their more or less $u$ pieces of mobility information. However, the total number of mobility attributes pertaining to a specific person is around $O(k)$.

Attackers can conduct mobility analysis to infer WHO did WHAT. They will exploit all kinds of information from the mobility data columns. For example, Korean women do not go to a barber shop to get their hair done. Thus, it is not strange to assume that the

| Quasi-identities (age, gender, ZIP) | Movements (WHEN, WHERE, and BY attributes) |
|------------------------------------|---------------------------------------------|
| (67, M, 01234)                     | (2/20/2020 14:30, plastic surgery clinic A) |
| (35, F, 01234)                     | (2/20/2020 18:15, restaurant B, taxi from Banpo to Gangnam) |
| (42, M, 12345)                     | (2/20/2020 19:20, department store C)     |
|                                   | (2/20/2020 19:30, bookstore D)            |
|                                   | (2/20/2020 20:00, hotel E)               |
|                                   | (2/20/2020 20:15, karaoke F)             |
| (75, M, 67890)                     | (2/21/2020 13:00, flower shop G)         |
| (60, M, 09876)                     | (2/21/2020 15:40, barber shop A)         |
| (40, F, 76787)                     | (2/21/2020 15:40, café A)                |
|                                   | (2/21/2020 16:00, golf club A)           |
|                                   | (2/21/2020 17:00, cinema B)              |
barber shop is not included in the itinerary of a woman of age 40 in Table 36.8. Of course, it is possible that she took her son to the barber shop, but the probability is very low. Since it takes more than 20 min to get a haircut, it can be inferred that a man in Table 36.8 cannot also visit a cafe or golf club at the same time. We can make this reasoning using time information from the mobility data. As a result, careful information blending is required. Based on these assumptions, we can guess that a man visited the barber shop first and the cinema alone or visited the barber shop only. We can also guess that the man visited the barber shop first and the cinema with another man, or another woman, or both of them. It is clear that theoretically the man’s itinerary includes the barber shop, café, golf club, and/or cinema. However, since there are \( u \) patients, so many scenarios are possibly to be made by combining all of them.

In this paper, the \( u \)-indistinguishability is considered. The value \( u \) is a hyperparameter whose parameter value has to set before the learning process begins. In more detailed model, \((u, v)\)-indistinguishability can be used with two hyperparameters, \( u \) and \( v \). The value \( v \) denotes the number of moving activities. In general, the value \( v \) must be at least equal to the value \( u \), and, in general, larger than the value \( u \).

### 6. Conclusion

In this paper, the KCDC and local governments’ format used to share COVID-19 patient movement information is examined. The KCDC COVID-19 contact tracing system has been evaluated as having the potential to violate privacy with its new patient information disclosure format. The problem with this system is that it discloses too much information. Nevertheless, if Korea must continue to rely on this system, a sophisticated technique to reduce privacy breaches must be implemented. Under these circumstances, it is shown that the \( u \)-indistinguishability approach is sufficient to achieve the desired purpose by making it more difficult for inaccurate and unfair inferences to be made from the information provided during contact tracing. So far, South Korea does not apply the \( u \)-indistinguishability method.

### Acknowledgment

Thanks to Kari Karlsbjerg for proofreading the article and insightful suggestions. This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (grant number: KHIDIHI19C0785020020).

### References

[1] J. Cohen, K. Kupferschmidt, Countries test tactics in ‘war’ against COVID-19, Science 367 (6484) (2020) 1287−1288. https://science.sciencemag.org/content/367/6484/1287.
[2] Coronavirus disease-19. Republic of Korea. http://ncov.mohw.go.kr/en/bdBoardList.do.
[3] S. Neuman, South Korea reports no new domestic coronavirus cases, NPR (April 30, 2020). https://www.npr.org/sections/coronavirus-live-updates/2020/04/30/848181755/south-korea-reports-no-new-domestic-coronavirus-cases.

[4] S. Cha, South Korea: No New Domestic Coronavirus Cases, No Transmission from Election, April 30, 2020. https://www.reuters.com/article/us-health-coronavirus-southkorea/south-korea-reports-no-new-domestic-coronavirus-cases-first-time-since-feb-peak-idUSKBN22C05U?il=0.

[5] B. Powers, For contact tracing to work, Americans will have to trust Google and Apple, CoinDesk (April 14, 2020). https://www.coindesk.com/for-contact-tracing-to-work-americans-will-have-to-trust-google-and-apple.

[6] S. Park, G.J. Choi, H. Ko, Information technology—based tracing strategy in response to COVID-19 in South Korea: privacy controversies, J. Am. Med. Assoc. (2020), https://doi.org/10.1001/jama.2020.6602.

[7] T. Sparrow, South Korean FM on the Coronavirus: ‘We Have to Live With This risk’, DW, May 30, 2020. https://www.dw.com/en/south-korean-fm-on-the-coronavirus-we-have-to-live-with-this-risk/a-53427402.

[8] G. Cohen, M.M. Mello, Big data, big tech, and protecting patient privacy, J. Am. Med. Assoc. 322 (12) (2019) 1141–1142, https://doi.org/10.1001/jama.2019.11365.

[9] K. Huckvale, J. Torous, M.E. Karsen, Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation, JAMA Netw. Open. 2 (4) (2019) e192542, https://doi.org/10.1001/jamanetworkopen.2019.2542.

[10] J. Bay, J. Kek, A. Tan, C.S. Hau, L. Yongquan, J. Tan, T.A. Quy, BlueTrace: A Privacy-Preserving Protocol for Community-Driven Contact Tracing across Borders. White Paper, Government Technology Agency, Singapore, 2020. https://bluetrace.io/static/bluetrace_whitepaper-938063656596c104632def383eb33b3c.pdf.

[11] DP3T White Paper. https://github.com/DP-3T/documents/blob/master/DP3T%20White%20Paper.pdf.

[12] World Economic Forum, The next generation of data-sharing in financial services: using privacy enhancing techniques to unlock new value, White Paper (2019). https://www.weforum.org/whitepapers/the-next-generation-of-data-sharing-in-financial-services-using-privacy-enhancing-techniques-to-unlock-new-value.

[13] S. Fischer-Hbner, S. Berthold, Privacy enhancing technologies, in: Computer and Information Security Handbook, 2017, pp. 759–778. https://doi.org/10.1016/B978-0-12-803843-7.00053-3.

[14] L. Sweeney, k-anonymity: a model for protecting privacy, Int. J. Uncert. Fuzz. Knowl. Syst. 10 (5) (2002) 557–570. https://epic.org/privacy/reidentification/Sweeney_Article.pdf.

[15] A. Machanavajjhala, D. Kifer, J. Gehrke, M. Venkitasubramaniam, L-diversity: privacy beyond k-anonymity, ACM Trans. Knowl. Discov. Data 1 (1) (2007). https://dl.acm.org/doi/10.1145/1217299.1217302.

[16] N. Li, T. Li, S. Venkitasubramaniam, t-closeness: privacy beyond k-anonymity and l-diversity, IEEE Int. Conf. Data Eng. 106–115 (2007). https://ieeexplore.ieee.org/document/4221659.

[17] P. Samarati, Protecting respondents’ identities in microdata release, IEEE Trans. Knowl. Data Eng. 13 (6) (2001) 1010–1027. https://doi.org/10.1109/69.971193.

[18] J. Salas, V. Torra, A General Algorithm for K-Anonymity on Dynamic Databases. Lecture Notes in Computer Science, vol. 11025, Springer, 2018. https://link.springer.com/chapter/10.1007/978-3-030-00305-0_28.

[19] H. Kido, Y. Yanagisawa, T. Satoh, An anonymous communication technique using dummies for location-based services, in: Proceedings of the International Conference on Pervasive Services, 2005, pp. 88–97. https://doi.org/10.1109/PERSER.2005.1506394.
[20] X. Liu, X. Li, Location Privacy Protection in Mobile Networks, Springer, 2013.

[21] B. Zhou, J. Pei, The $k$-anonymity and $l$-diversity approaches for privacy preservation in social networks against neighborhood attacks, Knowl. Inf. Syst. 28 (2011) 47–77. https://doi.org/10.1007/s10115-010-0311-2.

[22] Y. Sei, H. Okumura, T. Takenouchi, A. Ohsuga, Anonymization of sensitive quasi-identifiers for $l$-diversity and $t$-closeness, IEEE Trans. Dependable Secure Comput. 16 (4) (2019) 580–593. https://doi.org/10.1109/TDSC.2017.2698472. https://ieeexplore.ieee.org/document/7913692.

[23] Z. Tu, K. Zhao, F. Xu, Y. Li, L. Su, D. Jin, Protecting trajectory from semantic attack considering $k$-anonymity, $l$-diversity, and $t$-closeness, IEEE Trans. Netw. Serv. Manag. 16 (1) (2019) 264–278. https://doi.org/10.1109/TNSM.2018.2877790.

[24] H.J. Kim, South Korea’s COVID-19 patient movement information release guidelines: too much detail, Korea IT Times (May 4, 2020). http://www.koreaittimes.com/news/articleView.html?idxno=97267.

[25] J. Domingo-Ferre, A. Oganian, Local synthesis for disclosure limitation that satisfies probabilistic $k$-anonymity criterion, Trans. Data Priv. 10 (1) (2017) 61–81. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6760907/.

[26] H. Li, F. Guo, W. Zhang, J. Wang, J. Xing, ($a,k$)-Anonymous scheme for privacy-preserving data collection in iot-based healthcare services systems, J. Med. Syst. 42 (56) (2018). https://doi.org/10.1007/s10916-018-0896-7.

[27] N. Kim, More scary than coronavirus’: South Korea’s health alerts expose private lives, Guardian (March 6, 2020). https://www.theguardian.com/world/2020/mar/06/more-scary-than-coronavirus-south-koreas-health-alerts-expose-private-lives.

[28] M.J. Kim, S. Denyer, A ‘travel log’ of the times in South Korea: mapping the movements of coronavirus carriers, Wash. Post (March 13, 2020). https://www.washingtonpost.com/world/asia_pacific/coronavirus-south-korea-tracking-apps/2020/03/13/2bed568e-5fac-11ea-ac50-18701e14e06d_story.html.

[29] Wikipedia, Traffic Analysis, 2020. https://en.wikipedia.org/wiki/Traffic_analysis.