Personal Privacy Protection in Big Data Environment Under the New Coronavirus Situation

Xing Chen [a] *
The National Computer Network Emergency Response Technical Team/Coordination Center of China Zhejiang Branch, Hangzhou, China.
* [a] ying_luo@163.com

***Abstract*—In early 2020, a massive outbreak of a new coronavirus pneumonia swept across the globe. Governments quickly moved to detect and track and isolate patients and doctors and nurses to effectively prevent the spread of virus and protect the safety of healthy people. Tracking personal trajectories, contact surfaces and other information by means of technology can help us quickly detect the close contacts of the virus. But these scattered data are aggregating together, which is the data of citizens' personal privacy. Once the data are used by malicious elements, it will cause incalculable economic losses and even public crises. Therefore, we must pay close attention to the protection of personal privacy. This paper first introduces the measures adopted by various countries in the world to fight against the epidemic situation, and discusses the problems and risks faced by them. Then, the mature privacy protection technology is analyzed. Finally, the research direction of the privacy protection stage during the epidemic period is proposed.**

Keywords—Big data, Personal privacy protection, Covid-19.

I. INTRODUCTION

In early 2020, a massive outbreak of a new coronavirus pneumonia swept the world. The world is now in an unknown situation. By the end of June 17th, there were about 8 million cases of Covid-19 reported worldwide, with 435 thousand deaths. In the Americas, Africa and South Asia, the number of cases is still rising rapidly[1]. The lives and safety of the world's people have been seriously threatened, and the governments of various countries are moving quickly. The treatment of patients, the search for specific drugs and the development of vaccines are also carried out by means of detection, tracking and isolation of patients and medical patients to effectively prevent the spread of virus and protect the safety of healthy people.

A. COVID-19 Dedicated Application Tracking

Covid-19 application specific tracking only refers to users installing special applications for covid-19 tracking. Users actively or passively report relevant information. Contacts will track application software to combine self-reported health status and location history to prevent users from touching the virus.

Utah has released an application called Healthy Together to track contacts in the state by tracking the location of mobile phones. The APP allows public health departments to track the location and the way of contact of people who are positive for coronavirus.

In addition, in April 10th, Apple and Google announced plans to jointly develop a contact tracking application using Bluetooth technology. This application allows users to report their positive diagnoses and receive warnings that they are in close contact with an infected person. [4]

The Singapore government has also launched an application called Trace Together, which allows mobile users to voluntarily share their location data, including GPS, mobile base station location data and Bluetooth signal data. [5]

B. Track Phones' Location

Currently, this method is the most common way. In South Korea, Singapore, Taiwan and other countries and regions, this way is widely used for personnel management. In Taiwan, the government used "electronic fence" technology to monitor the whereabouts of home segregated people by means of mobile location. More accurately know the movement of personnel. Once the home quarantine personnel leave the designated area, the system will send "alarm messages" to the parties, epidemic prevention departments and other related units synchronously, so as to grasp the whereabouts of the relevant personnel.

In South Korea, government departments are also using the location data of infected people to track their whereabouts. The government has drawn up data maps of mobile phones provided by telecom and credit card companies, and used contact tracing to send emergency alerts to explain the sex, age and other details of the patients. [2]

C. Health Code

In the mainland of China, "grid management" is a common mode of community management. Whether it is home segregation or fixed point isolation, it relies mainly on the supervision of community workers. But after the spread of the epidemic, the "Health Code" technology based on online self reporting will soon be popularized throughout the country. Health codes are classified into green, yellow and red according to their risk level. Green risk is the lowest and red is the highest. From the residential district to the company building where work is located, many places need to produce green codes that represent health to allow in and out.[3]

The application of health code can be completed on the mobile phone. After finding the health code program that you need, you can enter your personal name, certificate number, detailed address, the travel experience and health condition of the last 14 days according to the prompt, then the system will automatically show the result of health code. If any one of the contents changes, the result of health code may also change.

In addition, there are other ways to divide the population into online forms. For example, China Mobile, China Unicom and telecom operators launched the epidemic travel enquiry

platform to show whether users had left the local area or went to high-risk areas during the past 14 days. This has been used in some public places in Beijing or hospitals, railway stations, and so on.

II. Personal Privacy Risks

The application of various scientific and technological means will greatly help the government to control the spread of the epidemic and ensure the safety of the people. However, we should also see that the government has collected personal privacy information, especially location information, at different levels. The leakage of privacy information is not only the disclosure of personal information, but also the serious consequences brought to the public after the information leakage. It will even affect social stability and the normal operation of the economy.

When all the data of the individual come together, we can predict the behavior and hobbies of the person by using big data mining technology. For example, we can accurately guide our residence, work place, and even hobbies by visiting places.

In South Korea, a man's covid-19 test was positive. He provided government officials with data on his whereabouts. The purpose is to let officials warn the public not to enter certain areas and allow people who may have contact with him to be tested. But according to Broadcasting British Corporation, the man's data are sufficiently specific to determine his age, place of work and place of residence. And the fact that he was infected in the course of sexual harassment. [6]

From the above case, we can see that privacy protection must be paid enough attention to fight against the epidemic clock. Especially in the age of big data, there are still some information about individuals intentionally or unintentionally leaking on the network. This is also one of the great challenges in the era of big data.

III. Privacy Protection Technology

A. Protection During Data Transmission

After collecting user information through various channels, the government departments need to transmit to the centralized storage service nodes through the network. During this transmission process, we need to do a good job of protection. The common transmission protection is encrypted transmission channels. For example, the application of SSL certificate and the establishment of VPN channel. This encryption technology can be divided into two kinds: symmetric key encryption technology and public key encryption technology according to the different encryption algorithms.

1) Symmetric encryption key technology

This technology uses the same key when encrypting and decrypting, or uses two keys that can be easily extrapolated. In fact, this group of keys becomes a common secret between two or more members in order to maintain exclusive communication links. Common symmetric encryption include AES; Salsa20; DES; IDEA; RC5 and so on.

Symmetric encryption key technology has the advantage of short encryption key, low computation cost, fast speed, and difficulty in decoding. It is suitable for small or massive data encryption. It is the mainstream algorithm. Its disadvantage is that both sides can not effectively guarantee the security of the key by using the same key. The difficulty of key management grows with the increase of key data volume. Moreover, when it is used in distributed network system, the key management is difficult and the cost is high.

2) Public key encryption technology

It also known as asymmetric encryption key technology. It is a kind of cryptography algorithm. It needs two secret keys. One is the public key, the other is the private key; the public key is used for encryption, and the private key is used for decryption. The ciphertext with plaintext encryption can only use the corresponding. Private key decryption and get the original plaintext. Because encryption and decryption require two different keys, it is called asymmetric encryption; unlike encryption and decryption, the same key is used. Symmetric encryption is public and can be released arbitrarily. The private key can not be disclosed. It must be kept strictly secret by users. It will not be disclosed to any communication side. Common algorithms include RSA, ECC and Diffe_Hellman, DSA and so on.

The main advantage of public key encryption technology is higher security, public key is open, secret key is kept by itself, no need to give private key to others, and its disadvantage is that encryption and decryption takes a long time and is slow, it is only suitable for encrypting a small amount of data.

As a key privacy protection technology, data encryption technology can effectively prevent data leakage in the process of transmission, and ensure user information security.

B. Protection of Data Stored Procedures

After the user data is transferred to the node server safely, we need to store the relevant data in the database. The database is the main body of the information system, but it is also the most vulnerable part. It will also be threatened by internal staff [7], including unauthorized access, tampering with data, and breaking the availability of data. For this reason, we should strengthen relevant protection, mainly from the following aspects:

1) Strengthen rights management. Control the internal authority, classify and manage related permissions, ensure the necessity and minimization of authority. Add two verification mechanisms or IP and Mac address binding measures to high privileged users to ensure account safety. At the same time, access control strategy is needed to increase security audits.

2) Encrypts and stores privacy information. Encrypting data can enhance the security of data, but it will also increase the overhead of the system, such as increasing computing resources and reducing access speed. Therefore, we need to choose encryption according to the actual selection, such as encrypting sensitive data, encrypting and decrypting the data needed.

Data Storage Encryption ensures that user sensitive information exists in the form of ciphertext. It is stolen and cannot be read and used by hackers immediately. At the same time, the application of security audit and access control strategy effectively standardizes the operation of internal staff and reduces the risk of information leakage.

C. Protection of Data Release Process

The government analyzes data collected to make decisions conducive to society. Therefore, it is necessary to publish relevant information. For example, during the epidemic period,

in order to enable people to better protect themselves, Singapore, South Korea and other governments will publish detailed information about their lives, work and entertainment, including trajectories, residence time and so on.

If publishers do not consider privacy protection and publish data, it will cause serious consequences to citizens and bring them trouble. Therefore, the challenge of data publishing is that the published data can not only guarantee personal privacy information, but also maximize the effectiveness of publishing data. For this reason, we need anonymity to protect personal privacy when distributing messages.

Anonymization is the goal of protecting privacy by hiding user identities and sensitive data. Currently, anonymous methods mainly include generalization, suppression, clustering, decomposition and replacement. Generalization and suppression are some details of hiding quasi identifiers (which can identify user attribute sets), so that a specific value can be replaced by a universal value. Clustering is to divide data sets into clusters according to a given rule, to ensure that objects in clusters are similar and objects of different clusters are different. Decomposition and replacement are decoupling between quasi identifiers and sensitive attributes by grouping and mixing sensitive attributes.[8]

Although anonymity can protect users' privacy, it also has fatal defects. When attackers have a large amount of basic information, it can easily infer sensitive information of a certain record with comprehensive analysis of the published information. Therefore, we must study his privacy protection technology.

IV. Conclusion

In general, our privacy protection is in the stage of perfection. From the point of view of laws and regulations, we must establish a complete set of legal system to ensure that the government and enterprises can use user information reasonably and correctly. At the same time, we should strengthen education and publicity, and enhance people's awareness of personal information. This paper makes a simple analysis of privacy protection technology. It is expected that there will be a relatively complete technology plan to protect users' privacy information and meet the needs of users and markets in the future.

References

[1] Information on: www.who.int/zh/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---17-june-2020

[2] Information on: www.sohu.com/a/384365092_650579

[3] Information on: zh.wikipedia.org/zh-hans/%E5%81%A5%E5%BA%B7%E7%A0%81

[4] Information on: support.google.com/android/announcements/9929436?hl=en

[5] Information on: www.tracetogether.gov.sg

[6] Information on: www.cnbeta.com/articles/tech/965497.htm

[7] Bertino E, Sandhu R, Database security-concepts, approaches, and challenges[J]. IEEE Trans on Dependable and Secure Computing. Vol.2 (2005) No.1. pp. 2-19

[8] Xiangwen Liu, Liangmin Wang, Data publishing anonymous technology progress. Journal of Jiangsu University(Natural Science Edition). Vol. 37 (2016) No.5. pp. 562-571.

[9] Sharma Kartik; Aggarwal Ashutosh; Singhania Tanay; Gupta Deepak; Khanna Ashish (2019). Hiding Data in Images Using Cryptography and Deep Neural Network. Journal of Artificial Intelligence and Systems, 1, 143–162.