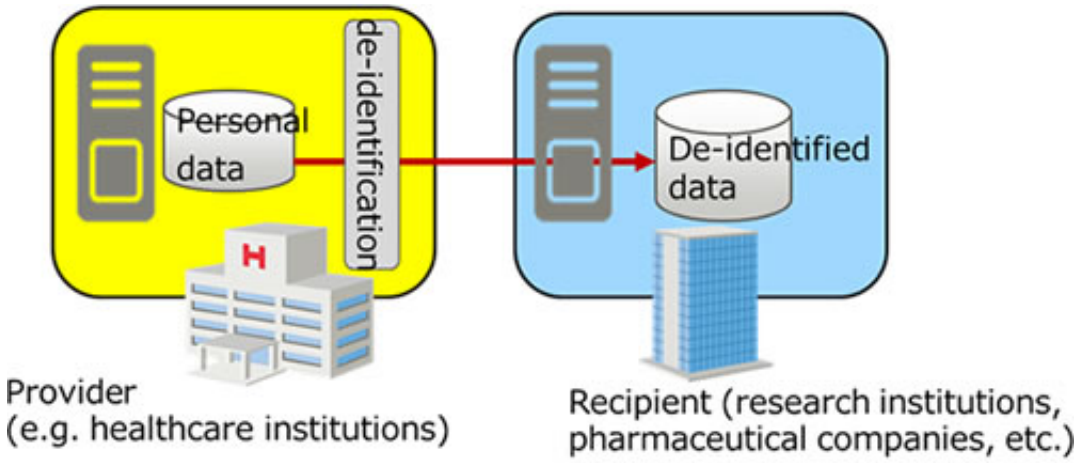
Novel technology to automatically assess personal data privacy risks

July 20 2016



Fujitsu Laboratories Ltd. today announced development of a unique new technology to automatically assess the privacy risk of personal data.

Under Japan's Amended Act on the Protection of Personal Information, which goes into effect in 2017, it will become permissible to provide third parties with personal data that has been processed to prevent the identification of a specific individual, or "de-identified," even without the individual's consent. Before providing de-identified data, the provider must first ensure it complies with guidelines and evaluate the risk that specific individuals could be recognized, which in cases outside Japan has led to experts spending many days in investigation. This is why

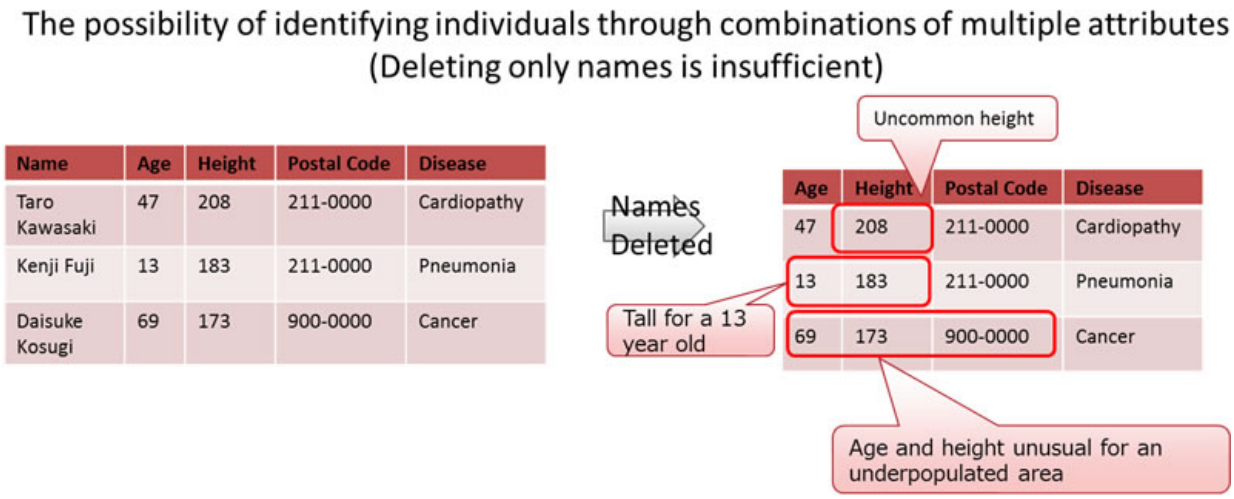
Fujitsu Laboratories developed novel technology to automatically evaluate the risk that an individual could be recognized from personal data. This technology enables data to be quickly and safely shared across multiple organizations, and can be expected to lead to improvements in the quality of products and services in a variety of fields, as well as to the resolution of social problems through co-creation between different industries. Details of this technology have been announced at the Information Processing Society of Japan's Computer Security Group (CSEC) meeting, held July 14-15 in Yamaguchi Prefecture.

Development Background

Under Japan's Amended Act on the Protection of Personal Information, which goes into effect in 2017, it will become permissible to provide third parties with personal data that has been de-identified, even without the individual's consent. This makes it possible to safely utilize data in different organizations, and is expected to lead to quality improvements in new products and services, with inter-organizational connections resolving societal problems and kick starting co-creation. There are a variety of methods for de-identification, which must be differentiated depending on the field and various guidelines. For example, it is conceivable that once guidelines, based on the Amended Act on the Protection of Personal Information, are established for the healthcare field, data, such as examination results held by healthcare institutions, will be de-identified and used by research institutions or pharmaceutical companies (Figure 1). For this reason, Fujitsu Laboratories developed de-identification technology focused on "k-anonymization," which is a technology to process information so that a minimum of k people possess the same attribute. Fujitsu Laboratories has been moving forward on research to apply the technology to healthcare and other fields.

Issues

Providers of personal data must be prepared for the risks associated with de-identification processing, such as checking whether or not they have met the guidelines for each industry, or if privacy could be violated from the de-identified data. It is not easy, however, for data providers to evaluate the risk that an individual could be identified from de-identified data and to take countermeasures, so evaluation and confirmation were previously left to experts, and the time required became an issue. There are reports, for example, of cases where healthcare institutions outside Japan de-identified data they held for use in medical research, and the process took more than half a year. For these reasons, Fujitsu Laboratories decided that, in order to quickly evaluate the risk that an individual might become known from de-identified data and take countermeasures, it is important to analyze the attributes that make it easiest for an individual to be identified, and then apply appropriate de-identification methods. Because it is possible, however, to identify an individual from the combination of multiple attributes (such as gender, telephone number, address) (Figure 2), the calculation of the combinations of attributes that make it easiest to identify individuals became so large that searching in a realistic amount of time became difficult.



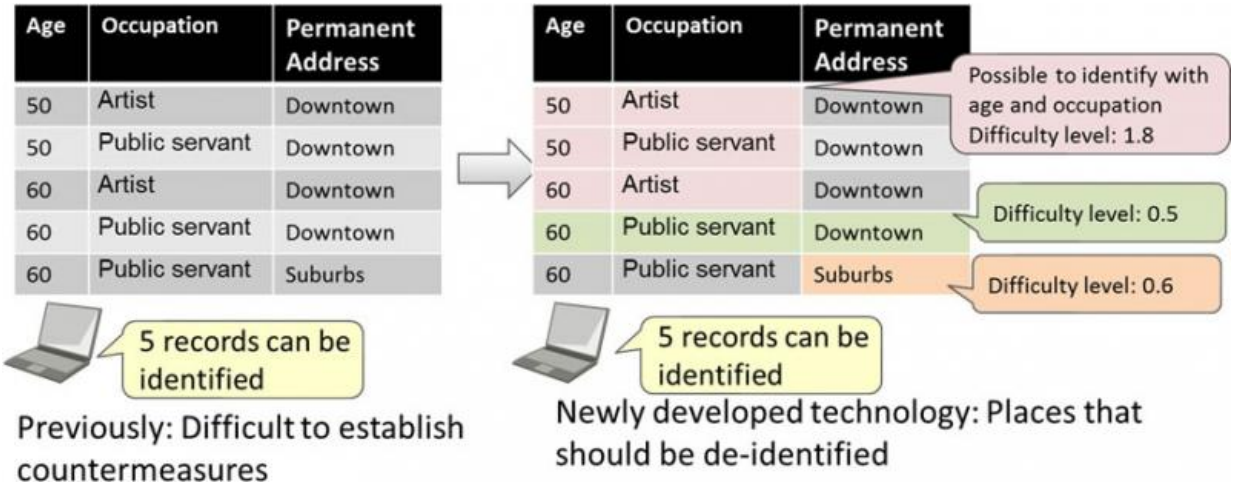
About the Newly Developed Technology

Based on data distribution, Fujitsu Laboratories has now developed a new technology to automatically search for combinations of attributes that make it easiest to identify individuals, as well as quantifying that ease of identification, in a realistic timeframe. This will enable data providers to rapidly analyze risks and take countermeasures. Features of the newly developed technology are as follows:

1. Technology to efficiently search for attribute combinations that enable identification

Fujitsu Laboratories developed a technology to efficiently analyze privacy risks by extracting the attributes that should be assessed, prioritizing from the combinations of attributes that make it easy to identify individuals. The system uses the fact that lines in the database (records) that can be identified by a combination of fewer attributes can

also be identified by combinations of more attributes to eliminate unnecessary analysis. For example, a record that could be identified with just age and occupation could naturally also be identified from age, occupation, and permanent address, so analysis of the latter combination can be omitted. This enables efficient analysis without the need to calculate huge numbers of combinations of attributes.



2. Technology to quantify the ease of personal data identification

Fujitsu Laboratories developed technology that searches for combinations of attributes in the data that make it easiest to identify individuals, and that can quantify that difficulty level to compare the ease of identification by individual. This makes it possible to quickly see which attributes should be prioritized for de-identification (Figure 3, Right). In addition to this technology, Fujitsu Laboratories developed technology to calculate potential damages if data is leaked, as well as to determine compliance with the various de-identification guidelines.

With these technologies, users can evaluate broad privacy risks, and easily carry out appropriate de-identification processing based on those risks.

Effects

Using this technology, it is now possible to automatically evaluate the risks of data for 10,000 people with 14 attributes in the realistic time frame of less than three minutes. This technology evaluates risk based on data distribution, so it does not require anything like a dictionary defining the weight of attribute values. Use of this technology can contribute to faster and safer provision of de-identified personal data to third parties, not only in the healthcare field, but also in finance, local government, and others. This makes possible safer data sharing among different industries, which can be expected to lead to improved quality of services and products, and to the resolution of problems in society through co-creation with different industries.

Fujitsu Laboratories is planning to verify the effects in a real environment and bring it into practical implementation around fiscal 2017.

Provided by Fujitsu