IoT Security for Utilization of Big Data: Mutual Authentication Technology and Anonymization Technology for Positional Data

● Takashi Shinzaki ● Ikuya Morikawa ● Yuji Yamaoka ● Yumi Sakemi

As the Internet of Things (IoT), is becoming increasingly widespread and measures are being taken to utilize the collected IoT data in big data analysis and applications. There is a wide range of target IoT items such as sensors, automobiles, consumer electronics, and wearable devices, and this makes it necessary to apply security technologies according to the device and usage, while considering the types of data exchanged. This paper presents the requirements for IoT security. It also describes the following specific approaches of Fujitsu Laboratories: technology that realizes mutual authentication and encrypted communication at the same time by applying in an extended way an ID-based key sharing scheme to Transport Layer Security (TLS); and technology that allows more small-section data to be safely disclosed through the use of multiple layers of fixed mesh on a map, as anonymization technology for the more efficient utilization of positional data.

1. Introduction

As the Internet of Things (IoT) becomes increasingly pervasive, measures are being taken to utilize the collected IoT data in big data analysis and applications. The things (devices) that are connected to networks in the IoT include factory equipment, sensors, automobiles, terminal equipment, consumer electronics, and wearable devices. This interconnection of devices and the use of the data they provide as databases make it possible to create value of a new kind.

These IoT devices handled by users use an architecture premised on connection to the cloud, and multiple clouds are expected to form amoeba-like connections, producing a hyperconnected cloud.

As IoT devices are of many different kinds and likewise the data they exchange is of many different types, high security is required. On the other hand, the information processing capability of these devices varies greatly, making the application of the security technologies conventionally used in personal computers and smartphones difficult, and thus it is necessary to apply different security technologies according to the device and usage.

Further, the IoT in turn uses the results obtained from the analysis of the data collected from various devices. As the data may include personal data, technology for the protection of privacy is also important.

This paper first introduces the technologies required for IoT security. It goes on to introduce mutual authentication technology for IoT devices via ID-based encryption for big data utilization, and anonymization technology for the use of positional data, as specific cases.

2. Technologies required for IoT security

In Fujitsu Technology and Service Vision[1)], Fujitsu provides an overview of the ways in which information and communications technology (ICT) contributes to business and social innovation. Among these, "authentication and authorization," "data privacy protection," and "security intelligence" are considered to be important pillars for building a secure ICT environment. These, applied to IoT security, could be said to be "device authentication and access control," "data protection and privacy protection," and "device vulnerability countermeasures and cyber attack countermeasures."

1) Device authentication and access control

This consists of technology for mutual authentication and access control between connected devices. Such technology can protect devices from external

attacks by performing mutual authentication and secure data communication between devices. Also, to enable the efficient use of the data sent from IoT devices, authentication technology is required that links the device and the person, as well as technology that connects the sensor ID to the personal ID, in order to establish whose data is being measured and who can use that data.

2) Data protection and privacy protection

To protect the data that is exchanged among IoT devices from external attacks, data protection technology is required. Privacy protection technology such as anonymization is also essential for the use of the results of analysis of the exchanged data, as such data is assumed to include personal data.

3) Device vulnerability countermeasures and cyber attack countermeasures

Along with the application of device vulnerability countermeasures, technology to quickly detect and handle any signs of potential security incidents is also required. IoT devices being connected to networks, it is assumed that they will be targets of attacks that exploit vulnerabilities, just like personal computers and smartphones at present. Thus, in addition to vulnerability countermeasures that boost resistance to such attacks, early detection of attacks as well as early verification of damage and countermeasures in case of intrusion are required.

Even in the event of a cyber attack on devices in a factory, a rapid and accurate grasp of the malware infection and the damage status can limit any damage, so that only a part of the factory's operations are disrupted.

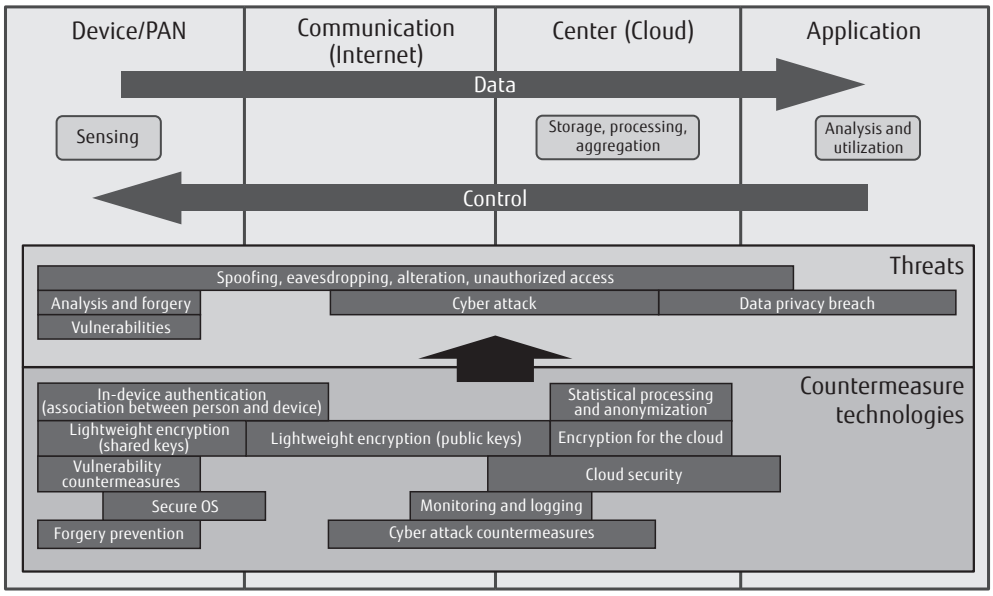
IoT security requires that all the above be realized in a very constrained environment. In other words, security must be considered taking into account limiting factors such as low device performance (in terms of processing capability, amount of memory, and power supply), restricted network connections, and limited user operation.

Figure 1 gives an overview of the threats to IoT security and the technologies to counter them.

3. Mutual authentication technology for IoT devices

3.1 Background

As IoT systems become more widespread, new businesses that use the analysis results obtained from the information collected by devices have started attracting attention. For example, equipping industrial machinery such as tractors with functions to communicate their current status and analyzing the degree



PAN: Personal area network

Figure 1
IoT security threats and countermeasure technologies.

of wear of parts makes it possible to systematically replace parts at the appropriate time. Another example is energy management systems such as building energy management system (BEMS), which can collected information on the state of devices, their power consumption, and so on, through the network connection of air conditioning and lighting equipment. Further, BEMS allow power saving and limiting of power consumption during peak periods by sending control information to devices based on the result of analysis of power usage data collected from devices.

In this way, IoT systems form a two-way communication model whereby not only device data is collected by servers, but control data is also transmitted from servers to devices. In such applications, mutual authentication that checks the validity of both device and server, rather than unilaterally checking whether the collected information has been transmitted from legitimate devices, or simply authenticating the server in order to prevent illegitimate control of devices, is essential. Further, as the number of connected devices is expected to reach about 50 billion in 2020, the applied mutual authentication technology will need to be extremely efficient. Fujitsu Laboratories conducts research and development of technologies that can efficiently execute the mutual authentication between devices required for IoT systems.

In today's Internet, an authentication and encrypted communication technology called Transport Layer Security (TLS) is widely used. As functions that realize mutual authentication, TLS offers TLS-PSK, which uses pre-shared keys (PSK), and the TLS-DHE-RSA authentication method, which uses RSA and Diffie-Hellman (DH) key exchange, which are both public key cryptographic protocols.

TLS-PSK is defined as a mutual authentication scheme designed for use in environments where device performance is constrained. In this scheme, the two entities that are to perform mutual authentication prove their legitimacy to each other by sharing secret information (pre-shared keys) beforehand. Since only symmetric key encryption scheme is used in the authentication process, this scheme is suitable for constrained computational performance environments such as sensor networks. However, in this scheme, the security of the entire system is at risk if the pre-shared keys are compromised by theft and analysis of devices.

On the other hand, the TLS-DHE-RSA scheme is an authentication method that uses public key cryptography and is widely used for Web services on the Internet. Usually, use of public-key cryptography requires certificates that are used to prove the owner ship of public key. During mutual authentication, the two parties first send each other their respective certificate, and public key cryptographic processing called certificate verification and key exchange is performed. In public-key cryptography, a different secret key is used by each entity, and unlike TLS-PSK, the security of the entire system is maintained even if secret keys are leaked. However, public-key cryptography has the demerit of requiring all the devices and servers on a system to have a certificate, which imposes a high processing load. Therefore, for the IoT, which encompasses an enormous number of devices, the huge number of man-hours required for providing certificates of all the devices and managing them, and the explosive increase in the amount of cryptographic processing and communication for certificate verification are significant issues. In response, Fujitsu Laboratories developed new technology that realizes mutual authentication without having to use certificates.[2)]

3.2 Developed technology

The technology developed by Fujitsu Laboratories uses a public-key cryptography technology called ID-based cryptography. This is a technology that allows the use of information (ID) associated with a user, device, or server, such as a device ID, e-mail address, or a fully qualified domain name (FQDN), as a public key. RSA and elliptic curve cryptography, which are public key cryptography used by conventional TLS, use random numbers as keys, which are generated independently from IDs. As a result, it is necessary to obtain the certificate of the other communication party beforehand, and to verify the validity of that key (random number). In contrast, in the case of ID-based cryptography, the ID of the peer is the key itself, so that encryption is possible without having to obtain certificates in advance and perform key verification with these certificates. This concept of cryptographic technology was proposed by Shamir in 1984,[3)] launching active research that resulted in the proposal of a practical scheme was proposed by Sakai et al. in 2000, followed by the proposal of another scheme by Boneh et al.,[4),5)] and so on.

Among these various ID-based cryptography, the technology developed by Fujitsu Laboratories uses ID-based key exchange that realizes secure key exchange through the use of IDs without resorting to certificates. In order to exchange keys, partial keys created using the device ID of the peer are first exchanged. Next, a shared key is generated from the partial key obtained from the peer, using the secret key corresponding to one's own ID. Since the shared key can be generated only by the legitimate device that has a secret key corresponding to the ID used to generate the partial key, subsequent communication facilitated by the shared key enables mutual authentication and encrypted communication simultaneously.

Fujitsu Laboratories have developed technology that applies an ID-based key exchange scheme that extends TLS. The developed scheme and the conventional TLS authentication procedure are compared in **Figure 2**. In the TLS extension, shaping and optimization are performed for the generation of shared keys by ID-based key exchange technology so as to comply with the existing TLS protocol. Conventional TLS in particular was subject to the constraint that only elements such as the IP address or domain name could be used as the device ID without certificate exchange. However, implementing an extension to efficiently transfer the information of the sender at the start of communication makes it possible to use any information as the device ID. Through this extension, use of mutual authentication technology via device ID with the same degree of convenience as that offered by TLS thus far is possible.

3.3 Evaluation

To evaluate the effect of the developed technology, the developed technology was implemented to OpenSSL, which is widely used as open source software (OSS) for TLS, and performance evaluation was carried out. In the evaluation, assuming the collection of data from devices, the client was implemented in a small single-board computer, and the server in a general-purpose personal computer. As a result, the communication traffic volume was reduced to approximately one-sixth compared with conventional

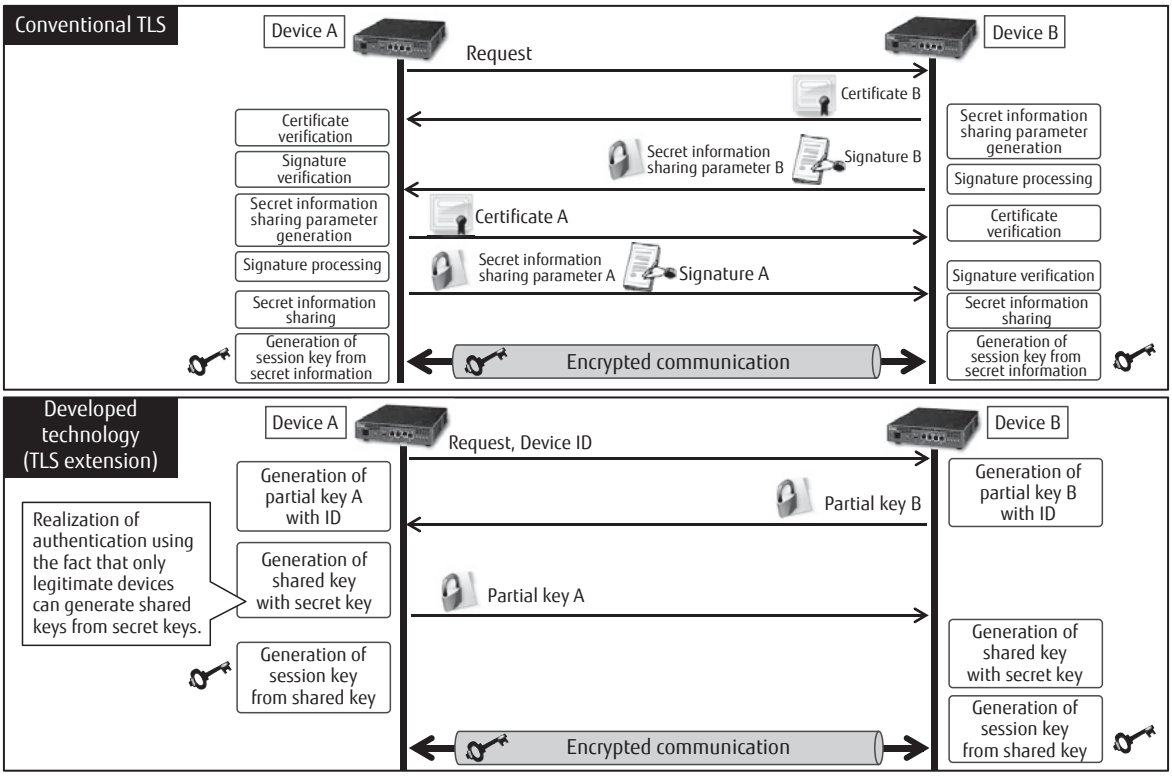


Figure 2
Comparison of developed technologies and conventional TLS.

TLS-DHE-RSA, and the processing time was shortened to approximately one-fifth (**Figure 3**). Thus the developed technology was found to make non-onerous mutual authentication between devices possible, while achieving a level of security equivalent to that of Web services and the like even for IoT.

The authentication technology developed in the present study will be used as a fundamental part of IoT. Going forward, it is important that Fujitsu not go it alone but collaborates with other companies and research institutions in securing interoperability in the IoT era, in which tens of billions of devices will be connected via networks. To this end, in collaboration with the Green University of Tokyo Project (GUTP), a Green ICT initiative of the University of Tokyo, we are promoting the introduction of this authentication technology to the IEEE 1888 communication standard for BEMS.[6] IEEE 1888 is a communication protocol developed as the result of standardization activities by the University of Tokyo, various companies and universities in China, and others, and was formally established as an international standard in 2011. This protocol, which adopts a communication method based on HTTP and XML, achieves the integrated management of various communication standards such as ZigBee and the Building Automation and Control Networking protocol (BACnet), which is a communication standard for building automation and control networks. IEEE 1888 not only collects data from devices, but also supports control communication to devices. For this reason, the developed technology has become a model for the targeted two-way communication. Next, through demonstration experiments with GUTP, we will promote interconnectivity verification, the identification of issues, and their solution, toward adoption of this technology in this project and IEEE standardization.

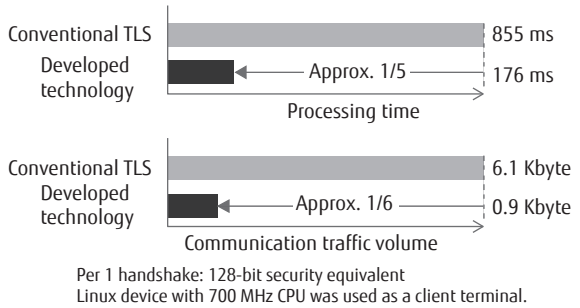


Figure 3
Performance evaluation results.

4. Anonymization technology for the utilization of positional data

This section explains our activities in the area of anonymization technology for the utilization of positional data.

4.1 Background

In recent years, car navigation systems and wearable devices with GPS receiver functions have begun providing positional data via the Internet to collecting entities that process that data and make it available to third parties. For example, a service involving the collection by companies (collectors) of positional data of a large number of automobiles (providers) traveling in the areas stricken during the Great East Japan Earthquake and the disclosure of that data in plotted form on maps to allow third parties (users) to obtain actual traffic information, attracted attention.[7]

With the rise of IoT and modifications of the legal framework, including Act on the Protection of Personal Information, for the proper utilization of personal data, the provision of positional data to third parties is expected to become increasingly widespread.

However, positional data from providers may include data of a private or confidential nature. For example, log of visited shops and facilities may make it possible to estimate personal preferences and circumstances that the individuals in question might prefer to keep unknown by others. Thus, in order to protect the privacy and confidentiality of providers from users, providers often need to anonymize positional data before disclosing it. Against this time background, Fujitsu Laboratories has been conducting research and development of anonymization technology for various types of data, including positional data.

4.2 Developed technology

The positional data introduced in this paper is a collection of records that include the provider's identifier (ID) and latitude and longitude information. Further, it is assumed that each provider continually provides records at short time intervals. For example, in the case of the previously described traffic information service, the provider ID is the ID of the automobile

such as car navigation system ID, etc., and the latitude and longitude are the positional data of the automobile. A batch of records is collected for each ID at short intervals, such as one-second intervals, during travel.

Figure 4 a) shows an example of the positional data of a single day plotted on a map. Let's assume that provider A travels from E's home to D's home by car, but does not want E at E's home to know that she dropped by D's home. However, if E is an authorized user, and E learns that the only vehicle that arrived at E's home on the way is A, then E can infer from this data that A dropped by D's home. This is the leakage of personal information.

In order to prevent the leakage of personal information and confidential information from positional data, masking the identity of the provider, in other words anonymization of the positional data, is effective. Users often need not know information specific to individuals or companies, and the analysis of overall trends covering a large number of citizens or commercial vehicles, for example, is often sufficient. An example of that is the previously described traffic information service.

One simple means of anonymization is the deletion of IDs. Through this means, it is possible to reduce the risk of personal and confidential information leakage at locations with a large amount of data. However, at locations where the amount of data is not so large, the effect of anonymization will be limited. For example, if the data shown in Figure 4 a) is anonymized, E cannot be sure about which point is point A, but it is still possible for E to estimate this based on the distance between points and the trajectory.

Another approach is filtering based on map information. For example, in the previously described traffic information service, of the information for main roads, only the parts for which positional data was available are said to have been disclosed. As information for community roads facilitates leakage of personal and confidential, this method has the problem that it cannot be readily applied to roads other than main roads.

Further, in a different method that is frequently employed for positional data, the space that is covered is divided into a mesh of a predetermined size, and the number of records included into each section is recorded. In this paper, this is called the fixed mesh method. In the fixed mesh method, the information of sections with only a small number of IDs is often not disclosed for privacy and security protection. The Statistics Bureau of the Ministry of Internal Affairs and Communications of Japan uses the fixed meshed method for regional mesh statistics, and defines section size in five steps, from approximately 80 km^2 to approximately 250 km^2, assigning a code to each section. However, in order to loosely distinguish for example a specific road, use of a finer mesh of approximately 20 m^2 is required.

Figure 4 c) shows an example leaving sections with two or more IDs, as the result of the application of the fixed meshed method using the mesh shown in **Figure 4 b)** to the data in Figure 4 a). "*N*/*V*" in each section indicates the number of records *N* and number of IDs *V*. For example, 8/2 indicates that data consisting of a total of 8 records belonging to 2 providers exists in that section. In this example, of the 48 original records, 27 records cannot be disclosed and are deleted.

The fixed mesh method is an excellent method that requires only a small amount of calculation and reduces the risk of privacy and confidentiality breaches as long as the data of mesh sections with only a small number of IDs are not disclosed. However, in the case of a fine mesh, this method has the problem that the amount of data disclosed becomes extremely small if the boundaries of the mesh occur by mischance inside a road, such as shown in Figure 4 c), where almost no records as disclosed for the portion at the road of the center. As a solution to this problem, Fujitsu Laboratories has developed a new technology that can safely disclose a greater amount of data for small areas.

The developed technology increases the area for which data can be disclosed through the use of multiple layers of fixed meshes. **Figure 4 d)** shows the situation when, in addition to the first-layer mesh used in Figure 4 c), the area that can be disclosed is increased through use of a second-layer mesh. An additional 14 records can be disclosed as the result of use of the second-layer mesh. Usually, a larger number of mesh layers results in a larger number of records that can be disclosed. As the amount of calculation grows linearly according to the number of layers but the amount of calculation per layer is small, it is unlikely to be a problem.

Further, a method for effective mesh placement was also developed. Placement of a new mesh is most

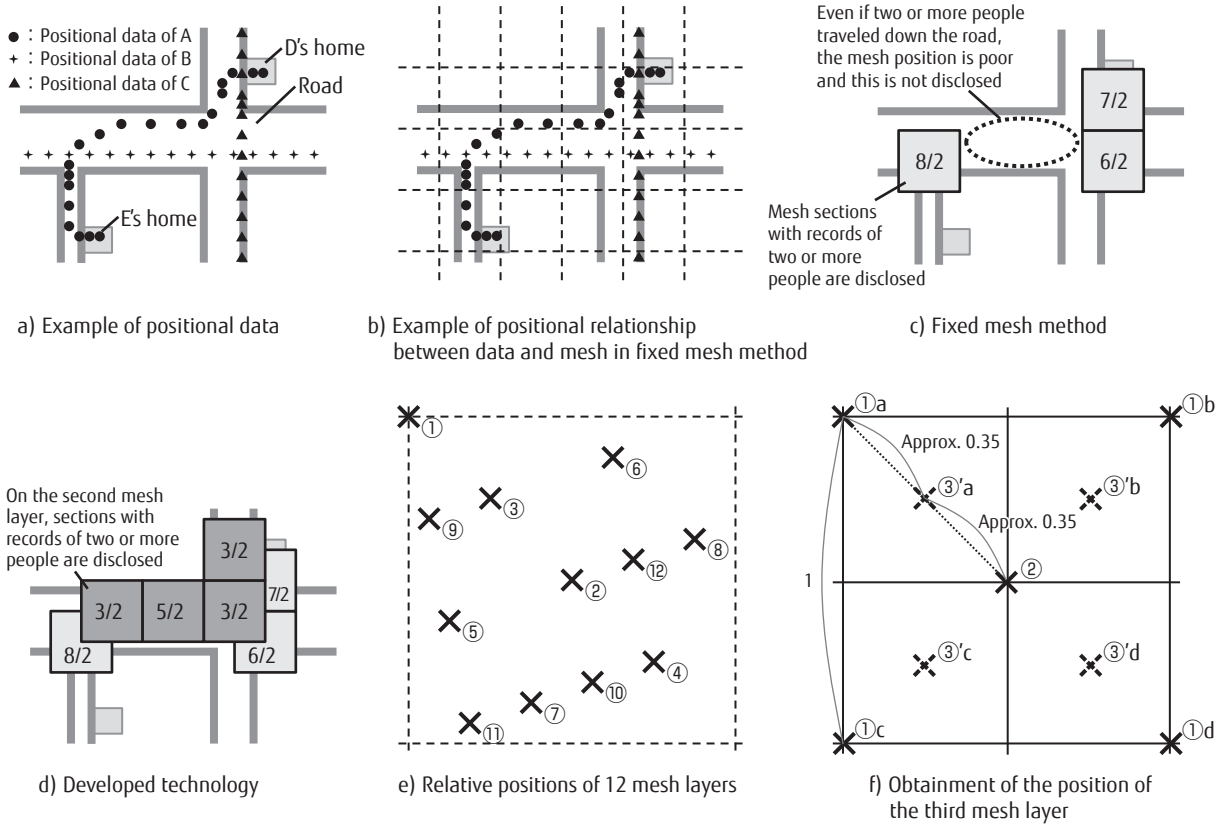
effective when the location is far away from any of the meshes that are already used. Therefore, among the points that are the most distant from all the sides of existing meshes, we decided to make the point that is the farthest from the vertex, which is the closest among all the vertices of the existing meshes, the vertex of the new mesh. **Figure 4 e)** shows the relative position of the mesh vertices of up to 12 layers. As an example, **Figure 4 f)** shows the obtainment of the position of the mesh vertices for the third layer. In Figure 4 f), the points most distant from all sides are ③'a to d, and the points closest to ③'a are ①a and ②. Assuming the length of all sides of the mesh is one , the distance between them is Approx. 0.35 each, and likewise for ③'b to d, there is no difference from ③'a to d, so that the vertex of the new mesh can be any of ③'a to d. This method can be applied also to three or more dimensions, making it applicable also the anonymization of data that includes the likes of elevation and time, in addition to latitude and longitude.

4.3 Evaluation

Application of this method to the actual data shown below showed the developed technology to increase the number of records that can be disclosed by approximately 7% compared with the fixed mesh method (**Figure 5**).

- Longitude and latitude unit: 0.1 second (corresponds to a length of approx. 3 m)
- 1 week's worth of second-by-second positional data of approx. 1,000 cars during travel
- Used area mesh code number: 5339-36 (Ariake area of Koto Ward, Tokyo, approx. three million records)

We selected 24 m^2 as the size of each mesh, and disclosed only the data of meshes with three or more IDs. Eight mesh layers were used for the developed technology. The processing speed was found to be in the order of 90 seconds for the developed technology, compared with 6 seconds for the fixed mesh method, using an ordinary personal computer—a performance



a) Example of positional data

b) Example of positional relationship between data and mesh in fixed mesh method

c) Fixed mesh method

d) Developed technology

e) Relative positions of 12 mesh layers

f) Obtainment of the position of the third mesh layer

Figure 4
Anonymization technology for the utilization of positional data.

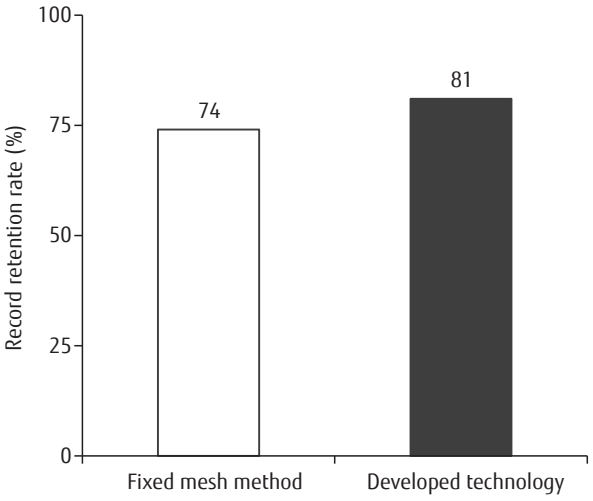


Figure 5
Effectiveness of the developed technologies.

level that can be said to be sufficiently practical.

As described above, Fujitsu Laboratories has developed a new technology that can safely disclose a greater amount of data for small areas. This development thus increases the possibility of offering traffic information services that are not limited only to main roads. Next, we hope to evaluate the effectiveness of the newly developed technology in actual use cases in terms of increasing the number of records that can be disclosed.

5. Conclusion

This paper has introduced the activities of Fujitsu Laboratories in the areas of IoT security and the utilization of big data.

To ensure IoT security, "device authentication and access control," "data protection and privacy protection," and "device vulnerability countermeasures and cyber attack countermeasures" are all important. The specific activities of Fujitsu Laboratories introduced in this paper are technology that extends the ID-based key sharing scheme to TLS to realize mutual authentication and encrypted communication at the same time, and technology for the safe disclosure of a greater number of records of small sections through the use of multiple layers of fixed meshes on a map as an anonymization technology for the utilization of positional data.

Going forward, we believe that IoT devices will assume an architecture predicated on connection to the cloud. We believe that a huge amount of IoT devices will be connected to the cloud, and moreover, that multiple clouds will interconnect in amoeba-like fashion, forming a hyperconnected cloud.

In anticipation of an IoT future, Fujitsu Laboratories is promoting research and development of IoT security technology assuming a hyperconnected cloud, and the application of developed technologies to cloud and IoT platforms. Further, Fujitsu Laboratories carries out research and development of schemes for the utilization of the collected IoT data on these platforms in a safe and secure manner.

References

1) Fujitsu: Fujitsu Technology and Service Vision. *http://www.fujitsu.com/global/vision/*
2) Y. Sakemi et al.: Application and implementation evaluation of TLS for key exchange with ID-based authentication function. Symposium on Cryptography and Information Security (SCIS) 2016, 2016.
3) A. Shamir: Identity-based cryptosystems and signature schemes. CRYPTO 1984, Springer, LNCS 196, pp. 47–53, 1984.
4) R. Sakai et al: Cryptosystems based on pairing. SCIS2000, 2000.
5) D. Boneh et al.: Identity-based Cryptography Standard (IBCS) #1. CRYPTO 2001, Springer, LNCS 2139, pp. 213–229, 2001.
6) H. Ochiai: IEEE 1888 Protocol Textbook, Impress Japan, 2012.
7) S. Sudo et al.: Use of probe information following occurrence of a wide-area disaster—through the case of the Great East Japan Earthquake—Journal of Information Systems Society of Japan, Vol. 8, No. 1, pp. 30–41 (2012).

Takashi Shinzaki
Fujitsu Laboratories Ltd.
Mr. Shinzaki is currently engaged in authentication technology related research and development.

Ikuya Morikawa
Fujitsu Laboratories Ltd.
Mr. Morikawa is currently engaged in authentication technology related research and development.

Yuji Yamaoka
Fujitsu Laboratories Ltd.
Mr. Yamaoka is currently engaged in Data privacy protection technology related research and development.

Yumi Sakemi
Fujitsu Laboratories Ltd.
Ms. Sakemi is currently engaged in authentication technology related research and development.