Inter-Cloud Data Security for Secure Cloud-Based Business Collaborations

● Hiroshi Tsuda ● Akihiko Matsuo ● Kenichi Abiru
● Takayuki Hasebe

With the advent of cloud computing, the boundary separating internal and external data has become increasingly blurred due to the utilization of external services. As a result, existing methods of preventing data leakage, such as only using a gateway to block the outflow of confidential data, have become insufficient. Therefore, there is increased demand for new security technology to allow confidential data to be safely used even in the cloud. We have developed new cloud information gateway and access gateway technologies that can mask confidential information contained within data before it is processed in the cloud. They can also transfer applications from the cloud to inside the company for internal processing. In this way, they make it possible to utilize cloud services without transmitting actual data. These technologies enable users to safely utilize confidential data in the cloud, encouraging new uses of cloud computing, such as cross-industry collaborations and specialized uses in specific industries. This paper outlines information gateway and access gateway and describes their applications.

1. Introduction

The forms of use of cloud computing (hereafter "cloud") are rapidly becoming diversified, and this diversification has given rise to a variety of security requirements. Users may increasingly often judge the confidentiality of information, security systems of cloud and convenience offered by use of cloud and entrust clouds with appropriate information in appropriate forms. Using cloud and external services blurs the boundary separating internal and external data. In this situation, existing methods of preventing data leakage, such as using a gateway at the boundary with the outside of the company to simply block the outflow of confidential data, are insufficient. Accordingly, new security technologies for safe use of confidential information are required on the premises, in public and hybrid clouds, and in third parties' SaaS.

This paper describes new inter-cloud data security technologies for tackling these challenges.

2. Partner cloud and its challenges

Our independent survey conducted on future forms of cloud use in North America has shown that users have high expectations for business collaboration. For example, such collaborations include joint development of new products, cooperation, software development, operations outsourcing and information exchange. We refer to this form of using one or more cloud services for cooperation between different organizations as "partner cloud."

It has also been found that there is a strong need for partner cloud in industries such as the retail, consumer packaged goods (CPG), medical and healthcare industries. In the retail and CPG industries, for example, there is demand for using cloud to jointly develop new products with

other companies and customers, in addition to supply chains as has been the case up to now. In the healthcare industry, users wish to migrate to cloud for health information exchange (HIE), in which home doctors, hospitals, pharmacies and testing institutes work together.

In this partner cloud, the security of information exchanged between clouds is a major concern of users. One big challenge is how to protect legally regulated information including medical records and information such as intellectual property generated by joint development in the partner cloud. In such a situation, simply blocking or encrypting confidential information makes it hard to use the services. It is also inadequate as a security means for collaboration between two or more organizations when they wish to use each other's information safely in a cloud environment. In view of using information with SaaS applications in other companies' clouds, we need to be able to control information between clouds by freely combining three conditions, namely user environment, service and information.

3. Information gateway

To address the issues described above, we have developed a new cloud information gateway technology that can flexibly control the information exchanged between parties inside the company and the cloud and between clouds according to the content of the information and partner service. **Figure 1** shows an example of applying an information gateway to partner cloud.

In this information gateway, security is not ensured by blocking communication data as is done with the conventional firewalls. The following three technologies are used to provide processing while making visible the content of the data exchanged between clouds and execution environment functions. They allow users to safely use information in various applications such as information linking on the premises and with other organizations via the cloud.

1) Information masking

Confidential information is automatically masked when it is sent to partner services in the cloud and automatically restored when the results of the processing are received from the cloud.

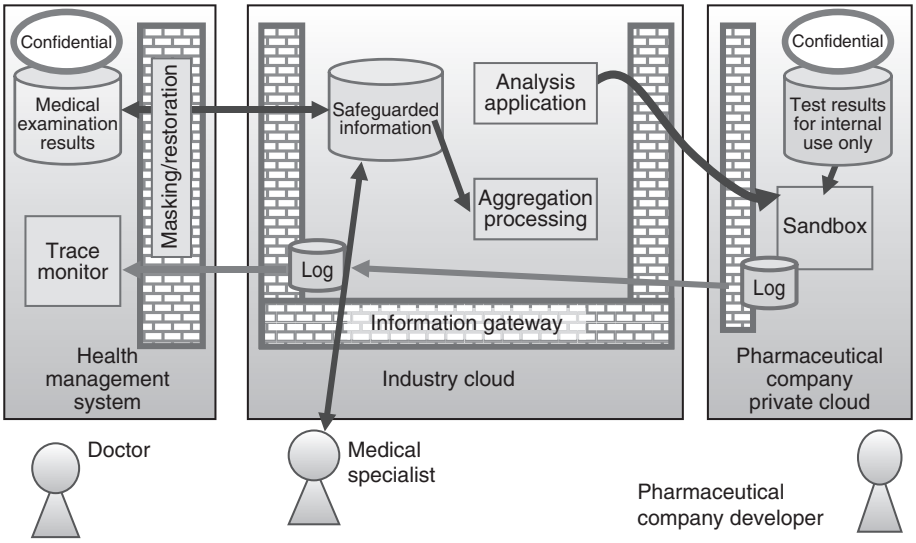

Figure 1
Application of information gateway to partner cloud.

2) Information traceability

The state of usage of information is traced and visualized based on the usage log and content of information across clouds.

3) Secure sandbox

Applications in the cloud can be placed in internal controlled execution environments for safe execution to process confidential information.

4. Information masking

Simply encrypting confidential information to store it in the cloud means that it cannot be processed unless the keys are provided in the cloud. Accordingly, we have developed an information masking technology. It takes away the confidential portions included in confidential information or replaces the information with information safeguarded by using special encryption techniques. Thus, the information can be processed in the cloud.[1)]

Take the utilization of medical examination results in an industry cloud as an example. Replacing patients' names and addresses with assumed ones when sending the information to the industry cloud outside, and restoring the information when it is received as the results of analysis by medical specialists can be executed in a way that is not apparent to the users and creators of cloud applications.

Information in spreadsheets owned by users such as the number of infected patients in respective areas may, if used as it is, result in leakage of sensitive information showing the concentration of patients in certain areas to other users in the cloud. To solve this problem, we have developed a privacy-preserving scrambled aggregation technology as shown in **Figure 2**. With this technology, special random numbers are added to hide the original values on the

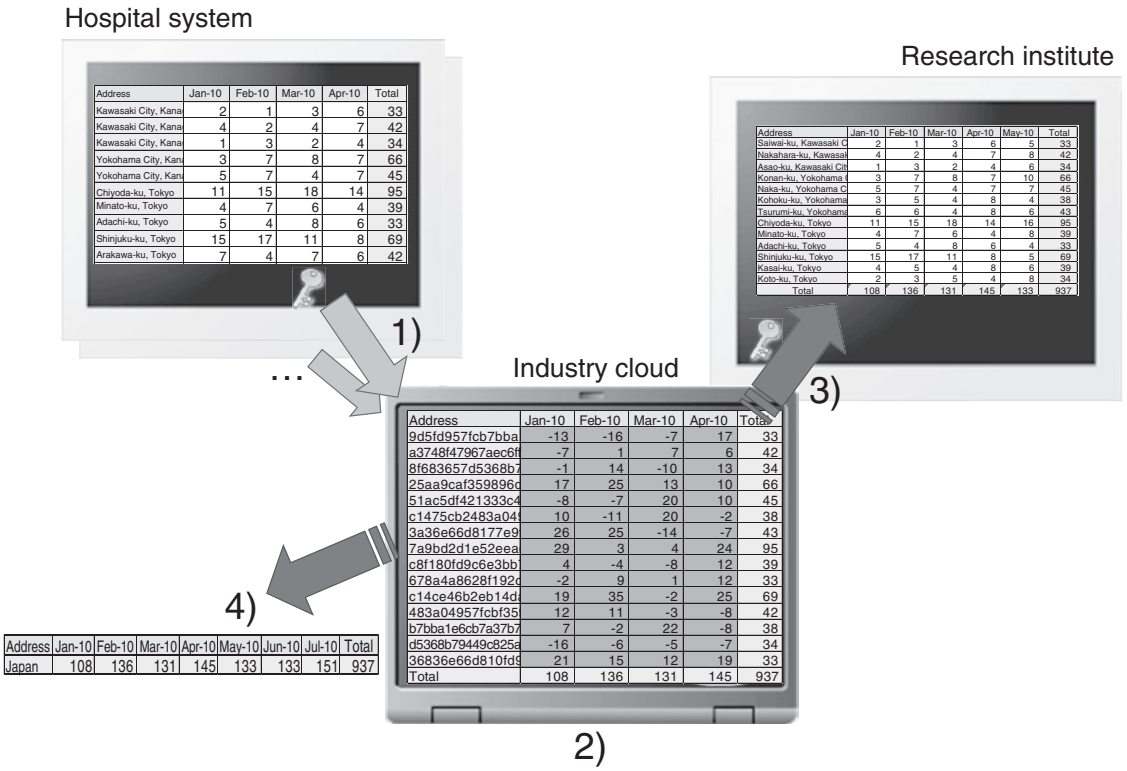

Figure 2
Example of privacy-preserving scrambled data aggregation.

premises when sending them to the cloud [1)] and multiple total values can still be correctly aggregated down and across as they are [2)]. The results of aggregation are available at various granularity levels (prefecture level, municipality level, etc.) according to the level of the decryption key of the individual user [3)]. Information can be processed without the need to store the original table information or key for restoration in the cloud, and one aggregation result can be used as the basis for handling multiple user levels. This provides the characteristic ease of database management. Even persons without the key have access to limited information: the overall total values only [4)].

For these information masking policies, the administrator can use XSLT (XML Stylesheet Language Transformations) on the information gateway provided on the premises or in the cloud to define detailed masking and restoration rules by the tag according to the service in the partner's cloud. Information that does not match these rules can be blocked, thereby ensuring that information exchanged between clouds is controlled and has an appropriate level of masking.

One of the related technologies is tokenization.[2)] For example, tokenization masks credit card numbers included in information to be handed to a subsystem to save on the costs of auditing under the PCI-DSS (Payment Card Industry Data Security Standard). In contrast, our technology is characteristically capable not only of flexibly describing rules including restoration but also of conducting specific analysis (aggregation) while information is kept masked. Another technology is homomorphic encryption, which allows combinations of additions and multiplications of encrypted numbers without using keys.[3)] However, it involves enormous computation costs and users cannot use partial results in the cloud such as total sum of the spreadsheets. Our approach, on the other hand, is a realistic one with restrictions applied to data structures.

5. Information traceability

In the survey conducted in North America mentioned above, technology for tracing accesses and changes to information entrusted to the

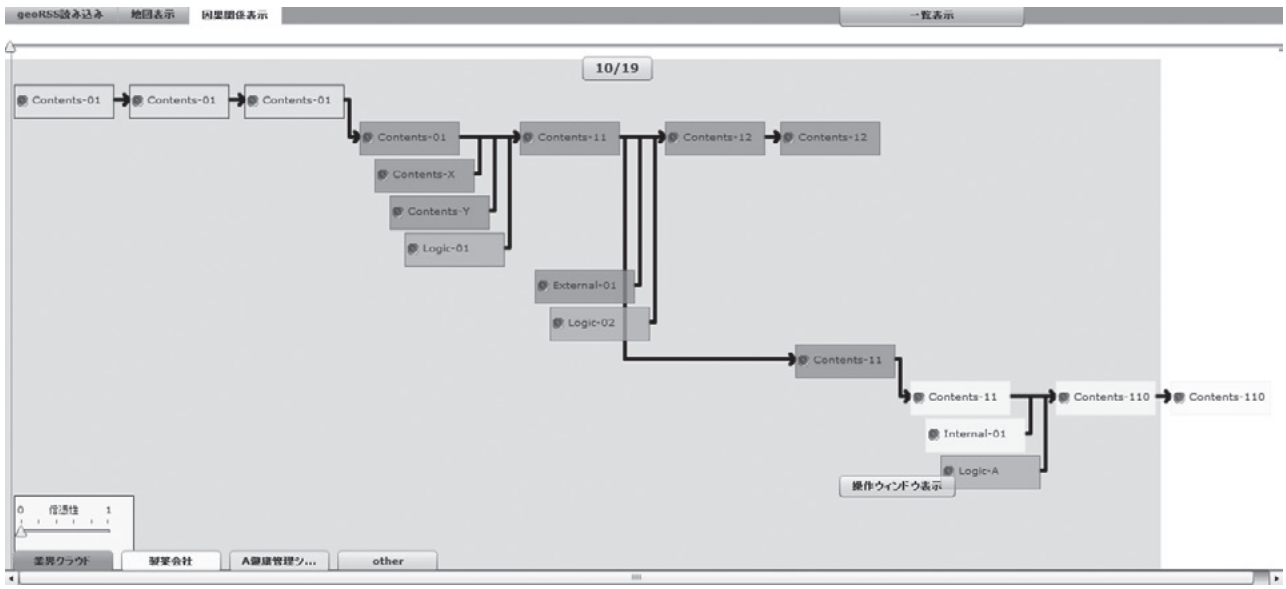

Figure 3
Screenshot of information traceability.

cloud was especially in high demand in the partner cloud. For a cloud operated by a third party accessed by other cooperating users, users need auditing to make sure that even masked information is used according to the contract.

To meet this need, we have developed information traceability technology that traces and visualizes the movement of information across clouds through an information gateway (**Figure 3**). The information gateway allows all cloud input and output logs to be recorded. For text documents in and out of clouds, the features of occurrence of keywords included in the text are recorded by using our proprietary fingerprint technique (content signature), which enables users to check the similarities of the text even if it is partially copied and edited. This fingerprint can be used to check data going out of the cloud to make sure that they do not contain any part of the confidential text and visualize the flow of the data.[1)]

One means of ensuring the traceability of information is to use electronic watermarks. Unlike that existing technology, our technology is based on the features of appearance of keywords in text. This eliminates the need for any special application and makes it possible to detect certain pieces of information with a small amount of text.

In addition to the inter-cloud input/output check by using information gateways, monitoring the state of use of data in the cloud by means of sandboxes, which is described below, can be combined. This allows users to check in joint development, for example, how data entrusted to the cloud have been used, including partial use of documents, to ensure they are used appropriately.

6. Secure sandbox

When services provided in public clouds such as SaaS applications are used on the premises, data are handed to the cloud for processing. For this reason, in view of the risk of information leakage, it has been difficult to handle data that contain confidential or private information intended only for internal use with external applications. As many good services will be provided in public clouds in the future, we expect to see increased demand for using them on the premises without anxiety. This will give rise to a demand for technology that allows services in the cloud to be used while information is kept on the premises.

To address this demand, we have developed a new technology for handling "internal use only" data with processing logic in a public cloud. It processes such information while it is kept on the premises by running the processing logic in the execution environment in an internal private cloud. With this technology, an application execution environment controlled from the information gateway is provided on the premises. To process information that cannot be let out of the premises, processing logic in the cloud is placed in this environment for execution based on the information placement policy. This execution environment, which is called a sandbox, is protected to prevent unauthorized operation. Even if any malicious code is included in the processing logic brought in from the cloud, there is no information leakage or other risks and secure execution is ensured. **Figure 4** shows how this secure sandbox can be combined with the information gateway.

7. Access gateway

As a form of using information in clouds, one possibility is to use it via partner services in other clouds. For example, SaaS applications such as print services and project management can be used to only show the necessary parts of documents by masking confidential information according to the user environment such as on the premises or at customers.

As a technology for such cloud federation via partner services in other clouds where the information gateway cannot be deployed, we

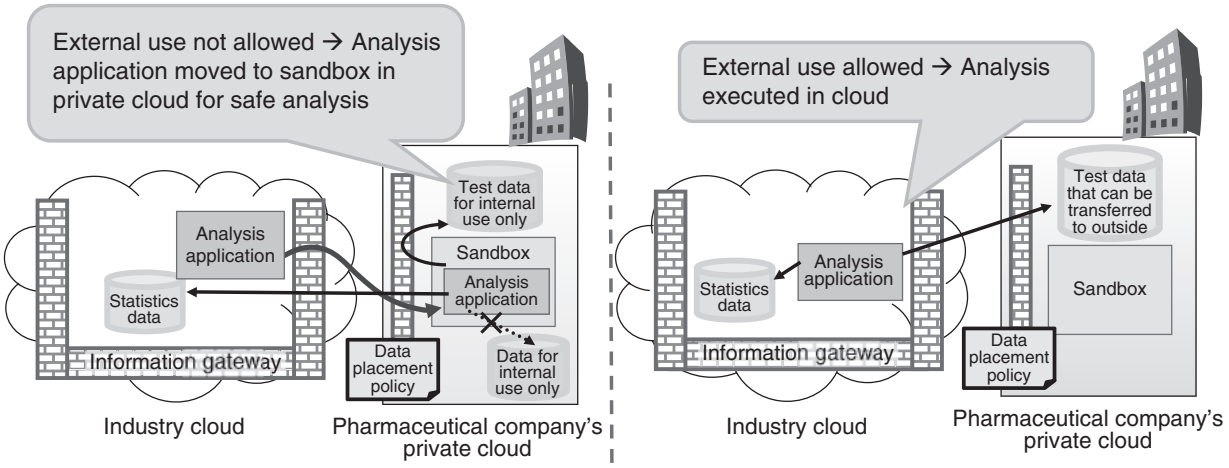

Figure 4
Combination of secure sandbox and information gateway.

have developed access gateway technology. It realizes inter-cloud data access by linking with the information gateway of the cloud that is entrusted with information.[4] **Figure 5** shows the developed combination of access and information gateways for location-based information control in a hybrid cloud environment.

Before confidential information entrusted to a cloud can be handed to other clouds, the validity of data access requests from other clouds must be confirmed. Currently, OAuth[5] is in increasingly wide use among Web services as a mechanism to check with users whether data can be handed over. For corporate use, however, those decisions on whether or not data can be provided should not be made by individual users. It requires access control based on the access rights for services used or data entrusted according to the predefined corporate user attributes (such as department and official position).

To meet this need, we have developed an access gateway provided on the network for combination with an information gateway. We have thus achieved control such as an ability to automatically change the confidentiality level according to the user environment or user attributes based on a judgment on whether or not to allow access to data to be handed to other clouds.

The access gateway allows a judgment to be made on a user's authority based on the user's environment (in the office, out of the office, etc.) and attributes (position, affiliation, etc.) at the time of logging in to the network service. In addition, we have extended OAuth so that in the OAuth phase it is possible to detect whether or not to hand data to partner services of other clouds by a substitute login in place of users to respective services. The results of the judgment on the users' authority can be announced at the same time. This allows users' use environment and attributes to be notified to the information gateway.

Having to configure different communication protocols and authentication methods for every sign-up with service providers in order to use various cloud services is cumbersome and time-consuming. However, collective control with a network service access gateway in this way helps to reduce the workload of enterprises.

8. Policy control

Making use of an information gateway to apply appropriate policies to information in the cloud eliminates erroneous drawing of information from the cloud and other risks and

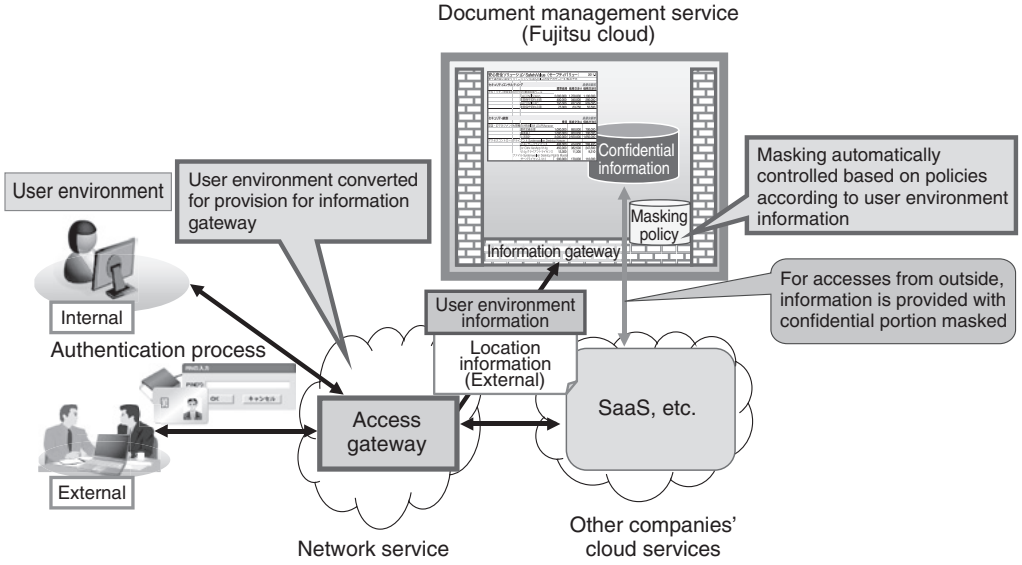

Figure 5
Combination of access gateway and information gateway.

allows information to be securely entrusted to clouds. From the viewpoint of utilizing this entrustment to the cloud, we need to consider the form in which information is provided, in addition to considering the conventional control of access to information.

Accordingly, as output control on information entrusted to clouds, we need control in terms of which piece of information is provided in what form to be used by whom (including user attributes and user environment), where (services, application, cloud services, and such like to which information is handed) for information gateways. Users need to make judgments on whether or not to output information, specify portions to be masked and change confidentiality levels according to these parameters. There are a variety of such parameters that may differ for every use and we need policies to allow them to be controlled without the need for separately configuring them. One possible way is to use a combination of access and information gateways to automatically switch confidentiality level for output.

In policy management, policies for information entrusted to the cloud are configured for individual pieces of information by users who own and have entrusted the information. The policies are configured with the information gateway and automatic control is provided according to the configured policies. In this way, we have realized granular control including output with the confidentiality level changed according to the user environment information, user attributes and partner services handed from the access gateway.

9. Conclusion

With the cloud information gateway technology and access gateway technology we have developed, three conditions including user environment, service and information can be freely combined for inter-cloud information control. This enables users to control information according to the degree of its confidentiality when it is to be exchanged with the cloud by preserving privacies. Taking processing applications in the cloud to be executed securely in the internal sandbox environment allows users to use applications without letting internal, confidental information off the premises. Furthermore, using an access gateway allows information

to be masked in various levels according to the access environment such as internal, external and domestic also when using other companies' partner services.

These technologies can be used as measures against information leakage in various situations of data linking such as between private clouds and public clouds, between public clouds, between in-house cloud services and third parties' services and between clients and the cloud. In the future, we intend to make use of these technologies to develop new cloud applications including cross-industry collaborations and specialized uses in specific industries.

References

1) K. Ito et al.: Data Masking and Traceability for Inter-Cloud Data Security. IEICE Technical Committee on Information Networks Invited Lecture, IN2010-129, 2011.
2) RSA Security Inc. Tokenization. *http:/www.rsa.com/*
3) C. Gentry: Fully Homomorphic Encryption Using Ideal Lattices. ACM Symposium on Theory of Computing (STOC2009), 2009.
4) T. Ogura et al.: Proposal of Secure Data/Service Collaboration Method among Public Clouds. IEICE Technical Committee on Information Networks, IN2011-57, 2011.
5) E. Hammer-Lahav, Ed.: The OAuth 1.0 Protocol. IETF RFC5849, April 2010.

Hiroshi Tsuda
Fujitsu Laboratories Ltd.

Dr. Tsuda is currently engaged in research and development of technology for secure knowledge processing.

Kenichi Abiru
Fujitsu Laboratories Ltd.

Mr. Abiru is currently engaged in research and development of technology for provisioning network services.

Akihiko Matsuo
Fujitsu Laboratories Ltd.

Mr. Matsuo is currently engaged in research and development of cloud federation technology and technology to improve software maintenance efficiency.

Takayuki Hasebe
Fujitsu Laboratories Ltd.

Mr. Hasebe is currently engaged in research on information security systems.