Inter-Cloud Computing

DOI 10.1007/s12599-011-0158-4

The Authors

Prof. Dr. Tomonori Aoyama (✉)
Research Institute for Digital Media and Contents
Keio University
West Annex, 2-17-22 Mita, Minato-ku
Tokyo 108-0073
Japan
aoyama@dmc.keio.ac.jp

Dr. Hiroshi Sakai
NTT Information Sharing Platform Laboratories
Musashino-shi 180-8585
Japan
sakai.hiroshi@lab.ntt.co.jp

Received: 2011-01-31
Accepted: 2011-02-13
Accepted after two revisions by Prof. Dr. Sinz.
Published online: 2011-05-10

This article is also available in German in print and via http://www.wirtschaftsinformatik.de: Aoyama T, Sakai H (2011) Inter-Cloud-Computing. WIRTSCHAFTSINFORMATIK. doi: 10.1007/s11576-011-0272-4.

Chairman of the Global Inter-Cloud Technology Forum Japan: http://www.gictf.jp/index_e.html.

1 Reasons for Inter-Cloud Computing

This catchword article presents the concept of "Inter-Cloud Computing" which introduces an additional management layer on top of conventional Cloud Computing Systems. Its goal is to reach a higher level of sustainability by autonomously shifting resources among the participating Cloud systems when unexpected load levels occur or disasters strike.

Cloud systems are widely seen as a promising paradigm for an IT infrastructure that is capable of creating an added value for business, society, and administration. If Cloud systems are to be applied to critical areas, such as enterprises' core businesses, governmental systems, medical applications, and other social infrastructure services, it is crucial to guarantee end-to-end service quality that covers the involved external networks, and to meet a set of requirements regarding reliability (including security), compliance, governance, and power efficiency.

However, if services are provided by a single Cloud system – which is the usual case today – unexpected levels of overload traffic from the Internet or natural disasters may affect the system's reliable functioning. The ability of a single Cloud to request reserved resources is usually fixed by limited bounds. Therefore, unexpected loads and failures can easily overburden a single Cloud system and lead to unreliable and interrupted services. In order to enable a Cloud system to continue the delivery of guaranteed service levels even in such cases, it is indispensable to introduce a super-ordinate layer of Cloud organization which manages multiple Cloud systems so that they complement each other. Thus single Cloud systems, which are interconnected via broadband networks, can mutually request required resources from their peers, or provide available capacity to them, respectively. This flavor of Cloud Computing is called Inter-Cloud Computing.

2 From Systems to Standards to Applications

The idea of Inter-Cloud Computing first arose at Cisco Systems which coined the term "Inter-Cloud" Computing as a vision on interoperability of different Cloud platforms without the need of explicit referencing by the user. On the 4th Int. Conference on Internet and Web Applications Bernstein et al. (2009) formalized the term by proposing a set of Inter-Cloud protocols. The scenario of this approach is to connect the resources provided by the different Cloud service providers.

The Inter-Cloud Technology Forum (GICTF) was established in Japan in July 2009 to study subjects for Inter-Cloud Computing schemes and to promote standardization for corresponding technologies. The following discussion on objectives and functional requirements for Inter-Cloud-Computing solutions reflect current results of the GICTF, which aim to provide – on a global scale – higher-reliability and higher-quality Cloud services in case of service failures of Cloud systems caused by natural disasters.

In addition, there are a number of other initiatives which address technical problems related to Inter-Cloud Computing. One of the most important projects is RESERVOIR (Resources and Service Virtualization without Barriers) which aims to develop system and service technologies that will serve as the infrastructure (as a service IaaS), especially in utilizing virtualization and grid technologies across administrative domains (Celesti et al. 2010). The effort concentrates on tools for deployment, migration, and management of services across network, storage, and administrative boundaries. To overcome the problem of non-standard management interfaces, the RESERVOIR project defines an abstract layer to support the development of a set of high level management components. The project tackles among others the problem of the federation (sometimes synonymously used for Inter-Cloud or Cross-Cloud) of so-called Cloud islands. With horizontal federation, different providers could exploit economies of scale, making an efficient use of their infrastructure and increasing their capabilities as well as their service offers. A three-phase Cross-Cloud federation model was proposed: (1) discovery of services/Clouds, (2) matching discovered services to necessary requirements (e.g., SLA, QoS), and (3) authentication, i.e., establishing a trust context between selected Clouds.

The second project worth mentioning is conducted by the DTMF Industry consortium, which includes amongst others Cisco, EMC, and Microsoft. The goal of the consortium is to focus on standardizing interactions between Cloud environments to achieve interoperability and portability between different Cloud providers and their users (DMTF 2009, 2010). They are the only attempt to at least address the areas of managing security and business continuity risks across

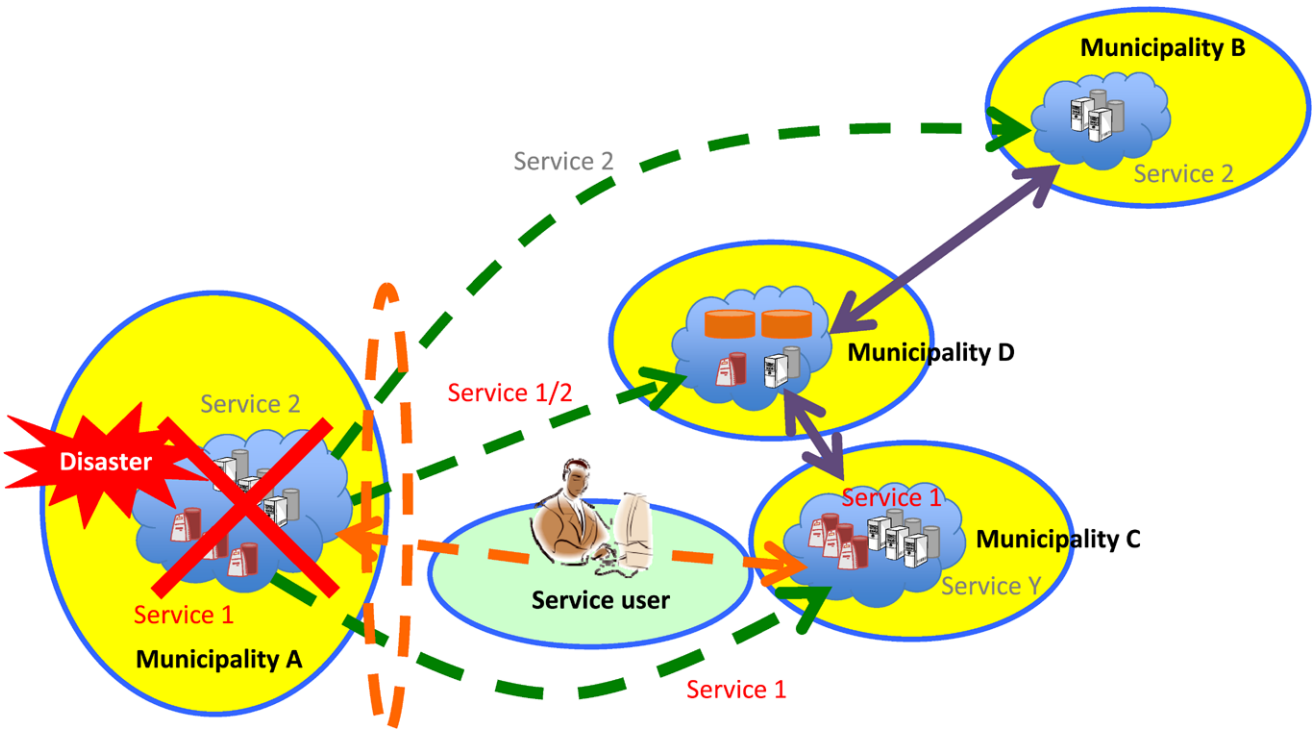

Fig. 1 Use case of guaranteed availability in the event of a disaster

several Cloud providers (QoS). They propose a specified SLA (Service Level Agreement) agreement in a distributed multi-provider environment. The initiative tackles three scenarios: (1) reducing cost increasing flexibility (prevent lock-in) when doing business with a new provider, (2) providing solutions for multiple providers to cooperate in order to meet customers' requirements, (3) defining tailored contractual agreements for different users with different needs. The consortium is still in an early stage, starting with defining use cases, service lifecycles, and a reference architecture. In the next phase interfaces and protocols will be developed.

This article illustrates the concept of Inter-Cloud Computing based on a use case about reaction to disasters as laid out by the Japanese administration. The Inter-Cloud Technology Forum (GICTF) was established in Japan to develop concepts for Inter-Cloud Computing solutions as a means to make administrative computer systems more robust against unexpected load levels and disasters, and thus sustainable. Building on results from GICTF, the article identifies the most important requirements that an Inter-Cloud system has to meet to accomplish the expectations the Japanese government has to Inter-Cloud Computing as a critical infrastructure in case of natural disasters. The use case is described in Sect. 3.

3 From Theory to Application: Natural Disasters – A Use Case for Inter-Cloud Computing

Figure 1 illustrates a use case which involves four municipalities of which each has at its disposal a single Cloud system. The Cloud system of Municipality A is damaged due to a natural disaster. It autonomously examines the impact of the disaster and determines that it is unable to continue to provide services and performs a disaster recovery procedure which involves using resources (such as applications, middleware, database servers, etc.) of the remote Municipalities B, C, and D, which are pre-arranged for such a service recovery scenario. Services that are normally provided by Municipality A are now temporarily provided by other municipalities, and users can continue to access the services without interruption. If the resources required for recovery cannot be provided for all services, those with a higher priority are preferred. Remaining services are recovered on a best-effort basis.

4 Requirements for a Critical Infrastructure

If a Cloud system experiences an unexpected overload or a natural disaster, it requires spare resources to cope with the situation. To guarantee the required service quality, such as service availability and performance, a flexible mechanism is needed to re-assign resources among the multiple Cloud systems. The primary objectives for such an Inter-Cloud Computing system are the following:

- Service Level Guarantees: To ensure a guaranteed performance level in the face of an abrupt increase in traffic reaching an unexpected level, a Cloud system must autonomously select an alternative provider offering an equivalent level. It further has to take care that the execution of differently prioritized tasks is arranged according to their priorities. Performance guarantees also include security and compliance requirements.
- Availability Guarantees: Guaranteed availability means that when a Cloud system is incapable of continuing its services due to a disaster, it must recover the services by interworking with Cloud systems located in areas unaffected by the disaster. If it is impossible to recover services in such a way that guaranteed quality levels can be

maintained for *all* services, they must be recovered according to priority levels. While guaranteed quality levels for high priority services must be kept, it must be attempted to maintain remaining services on a best-effort basis.
- Compatibility of Service Cooperation: In order to ensure the correct functioning of a service which is made up of several cooperating procedures, dependencies between these procedures must be maintained when the service is replaced. Replacement of a service due to an overload or a natural disaster must happen transparently to the service user.

5 Functional Requirements for Inter-Cloud Computing

To meet the above-stated objectives, this section describes the technical requirements for Inter-Cloud Computing as they have been identified by the GICTF, categorized into ten functions.

5.1 Provider Selection based on SLA-defined Quality Requirements

This function enables the selection of Cloud providers and Cloud systems that meet consumer quality standards stated in an SLA. The function matches the consumer's and provider's SLAs in order to guarantee quality requirements even in the events of service performance degradation or the occurrence of a disaster. In order to compare quality requirements, the SLA of a Cloud system must be defined and published to other Cloud systems using standard formats. This makes it possible to select appropriate providers for interworking by means of comparing (exact match or within a tolerable range) the items of the SLAs. Further, it must be possible to search for resources – including applications and middleware – held by other Cloud systems and, in turn, it must be possible for a Cloud system to be detectable by other Cloud systems.

5.2 Monitoring

The task of this function is to collect and monitor the usage status and dead/alive status of each computing or network resource of a Cloud system, and to determine the need for load distribution or disaster recovery. To this end, it must be possible to – periodically or on a request basis – collect resource information (such as information about the performance and operation of each server, storage unit or network) for each service provided by a Cloud system. It must further be possible to exchange such resource monitoring information among other Cloud systems by means of commonly defined formats.

5.3 Provisioning

The provisioning function determines the resource requirements (volume, type, etc.) based on the SLAs and the concrete resources that are required to replace services while maintaining the guaranteed service level. It must be able to identify bottlenecks when the traffic loads from or to the network vary and to dynamically perform corresponding compensation planning, since the operational characteristics of an application differ from case to case. It must further be able to perform a planning differentiated according to given priorities.

5.4 Resource Discovery and Protection

This function's objective is to search and discover resources that can replace services to be protected from service degradation or disaster impacts. Its responsibility is also to request the discovered resources from the Cloud systems that own them, when necessary.

In detail, it should be possible to search and discover the available resources in the group of Cloud systems that have been selected as candidates for providing replacement services. The function should search resources in the own Cloud system first, and extend the search to the other systems, if necessary. It must be possible to replace resources based on the SLAs. For example, if latency is critical, the function must firstly find replacement servers that are near the user. In contrast, if bandwidth is critical, it must find networks with sufficient bandwidth. Further, there must be mechanisms to allow recovery according to a given priority. Since a large quantity of resources might have to be recovered in case of a large-scale disaster, not every service can potentially be kept up. For example, lifeline services should be recovered with a higher priority.

5.5 Resource Management

The task of this function is to manage configurations of resources that are needed to protect services from degradation or disaster. Within this function, it must be possible to describe supplementary information in a standardized manner (such as resource type and status), in order to be able to manage resources over multiple Cloud systems in an integrated manner. Here, the management of various resource configurations for each service, such as servers, storage units, and networks, should be possible in a unified manner. Further, there should be options to update configuration information of resources over multiple Cloud systems in synchronization with events (e.g., protection or release of resources from other Cloud systems) and to manage difference information, so that the difference in resource configurations can be easily understood when configurations change.

5.6 Service Setup

This function performs basic routines prior to service provisioning in order to connect Cloud systems via networks, remotely activate application or middleware, and transfer or copy data to enable the use of secured resources in the own Cloud system or other Cloud systems. Further, the service setup function has to ensure that it is possible to access backup data that might be used in the event that a Cloud system is damaged as the result of a disaster.

In detail, the function must include the possibility to remotely activate resources (such as virtual machines, applications, and middleware) that have been reserved by the function "Resource Discovery and Protection". It should be possible to activate these resources under consideration of the configuration values with regard to the environment of the Cloud systems that provides the secured resources.

5.7 Authentication Federation

Federation of authentication is necessary to combine consumer identification information (IDs) so that consumers can experience a seamless service usage spanning multiple Cloud systems. The function has to support a variety of consumer information formats used by different ID management schemes. To this end, functionality must be included to make interoperable these ID management systems (which might use different data models or schemes). Further requirements for this function include:

- An authentication scheme must be provided that is capable of performing trust-based inter-authentication among multiple Cloud systems based on consumer-provided credentials.
- A system must be installed that manages trust relationships between servers in a single Cloud system or among multiple Cloud systems.
- Search functionality has to enable look-up of ID information of any service consumer in a Cloud system.
- It must be possible to search, discover, and exchange information generated for ID federation between federated Cloud systems. To this end, the generation and execution of rules regarding the lifecycle management of IDs, such as creation, updating, and release, must be enabled. For example, it shall be possible to simultaneously synchronize, partially synchronize, or release synchronization on the creation, updating, or release of consumer information that is managed across different Cloud systems.
- There must be mechanisms to keep consumers' Personally Identifiable Information (PII) confidential. To this end, an access control function must be established that protects consumer information against unauthorized access.
- A back-up authentication function must be installed for cases where the federated authentication service fails due to a disaster.
- Consistency of authentication information must be ensured by means of synchronization between functions managing identity information.

5.8 Network Interworking

This function provides central network management tasks in order to provide the highest possible network quality for interworking Cloud systems. It manages networks by monitoring the flow of each service, and by autonomously changing service flows based on the load level of the network. Also, it should enable energy savings, e.g., through partial shutdowns of network equipment.

5.9 Alternation and Retrieval of Data

This function is to transfer required data to the Cloud service which provides a service at a given point of time. When a consumer receives services provided by a different Cloud system, the function must ensure that data are adequately moved to which services have been delegated in order to cope with a disaster or degradation in service performance. In turn, it must retrieve the data when the own system recovers and becomes capable again to provide the services. Here, it shall be possible that data are autonomously moved along the network so that the consumer can transparently access the substitute Cloud system without further actions.

5.10 Releasing Resources

Finally, this function's task is to judge that the recovery of a service is no longer needed. It makes its decision based on monitoring results and eventually releases unnecessary resources, after disaster recovery or load distribution has been adopted. In particular, this means that it is possible to shut down virtual machines or applications that were activated when the secured resources began to be used, to update resource management information, and to completely delete or collect transferred data. Further, it shall be possible to release networks after servers and storage units have been released and to collect any workload remaining in other Cloud systems.

6 Reliability and Security Remain an Open Issue

While in the case of disasters reliability and security of Inter-Cloud is based upon the redundancy of resources, the unsolved issue in Cloud Computing – and therefore to an even higher degree in Inter-Cloud Computing – is the provision of reliable security and compliance guarantees with regard to business and consumer data (Mather et al. 2009). With respect to business the issue is the guarantee of an isolation property. An isolation property is fulfilled if there are no information flows between different users of the Cloud at the same time. The term isolation describes the requirement that service providers must guarantee that applications and data of different parties running on the same service Cloud are kept isolated from each other. This requirement cannot be fulfilled with classical security strategies, since they can merely regulate access control. Currently, isolation is achieved solely with access control mechanisms. This, however, is not sufficient as it does not provide "end-to-end" guarantees and neglects, for instance, implicit information flows (Ristenpart et al. 2009). Technically, the same problem arises when protecting customer data against the misuse by the Cloud providers. While the Japanese example shows that there are applications without such strong security requirements, other application domains require much greater security than we have today. To this end, further efforts must be made to develop reliable and formally-founded techniques that can ensure consumers that their data are stored, used, and processed only in compliance with applicable regulations and user-defined policies. Solving this issue is a central prerequisite for unleashing the positive potential of Cloud Computing – and therefore Inter-Cloud Computing – for business, society, and administration.

References

Bernstein D, Ludvigson E, Sankar K, Diamond S, Morrow M (2009) Blueprint for the intercloud – protocols and formats for cloud computing interoperability. In: International conference on internet and web applications and services, pp 328–336

Celesti A, Tusa F, Villari M, Puliafito A (2010) How to enhance cloud architectures to enable cross-federation. In: Proceedings of the 3rd IEEE international conference on cloud computing (IEEE Cloud 2010), Miami

DMTF (2010) Architecture for managing clouds – a white paper from the open cloud standards incubator 1.0. Distributed management task force, Inc. http://www.dmtf.org/standards/published_documents/DSP-IS0102_1.0.pdf. Accessed 2011-02-14

DMTF (2009) Interoperable clouds – a white paper from the open cloud standards incubator 1.0. Distributed management task force, Inc. http://www.dmtf.org/sites/default/files/standards/documents/DSP-IS0101_1.0.0.pdf. Accessed 2011-02-14

Mather T, Kumaraswamy S, Latif S (2009) Cloud security and privacy: an enterprise perspective on risks and compliance. O'Reilly, Köln

Ristenpart T, Tromer E, Shacham H, Savage S (2009) Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds. In: 16th ACM conference on computer and communications security (CCS'09). ACM, New York

Use Cases

CSA (2009) Security guidance for critical areas of focus in cloud computing V2.1. http://www.cloudsecurityalliance.org/csaguide.pdf. Accessed 2011-02-14

DMTF – Interoperable clouds. White paper from the open cloud standards incubator.

http://www.dmtf.org/sites/default/files/standards/documents/DSP-IS0101_1.0.0.pdf. Accessed 2011-02-14

SLA

GSA (2009) US Federal Cloud Computing Initiative RFQ. http://www.scribd.com/doc/17914883/US-Federal-Cloud-Computing-Initiative-RFQ-GSA. Accessed 2011-02-15

JEITA (2008) SLA guideline for IT systems in the private sector. http://home.jeita.or.jp/is/committee/solution/guideline/080131/index.html. Accessed 2011-02-14

Ministry of Internal Affairs and Communications (2008) SLA guidelines for SaaS. http://www.meti.go.jp/press/20080121004/20080121004.html. Accessed 2011-02-14

Office of Prime Minister of Japan (2010) Guideline for e-Government. http://www.kantei.go.jp/jp/singi/it2/guide/index.html. Accessed 2011-02-15

Ministry of Internal Affairs and Communications (2010) Guideline for the development of business continuity plan (BCP) for the ICT departments of municipalities. http://www.soumu.go.jp/menu_news/s-news/2008/080821_3.html. Accessed 2011-02-15

Ministry of Internal Affairs and Communications – Promotion of e-Government (2010). http://www.soumu.go.jp/main_sosiki/gyoukan/kanri/a_01.htm. Accessed 2011-02-15

Interfaces and functional structures

DMTF – Distributed Management Task Force (2010) Architecture for managing clouds. http://www.dmtf.org/standards/published_documents/DSP-IS0102_1.0.pdf. Accessed 2011-02-14

OGF (2010) Open cloud computing interface working group. http://www.occi-wg.org. Accessed 2011-02-15