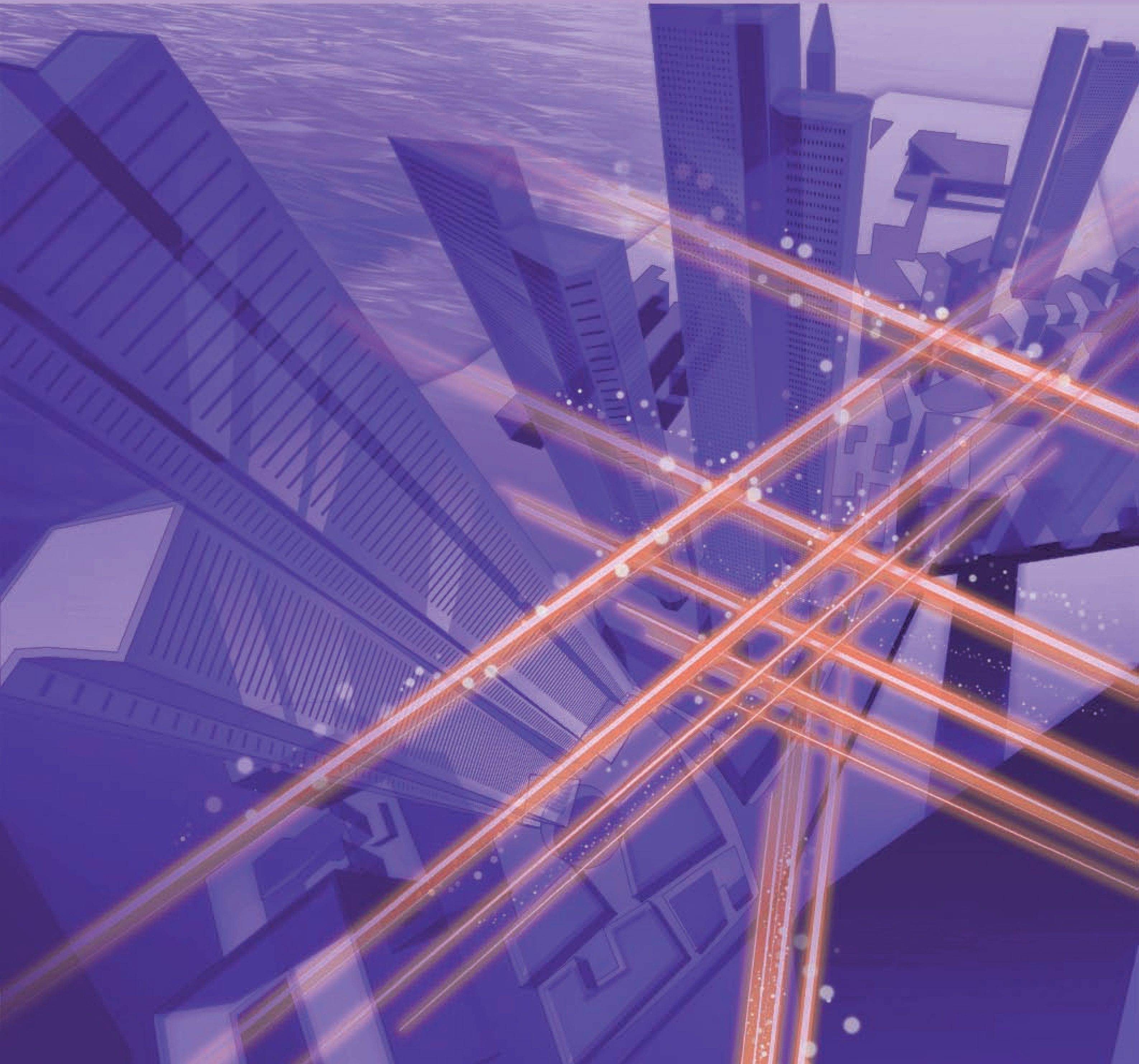
NTT Technical Review

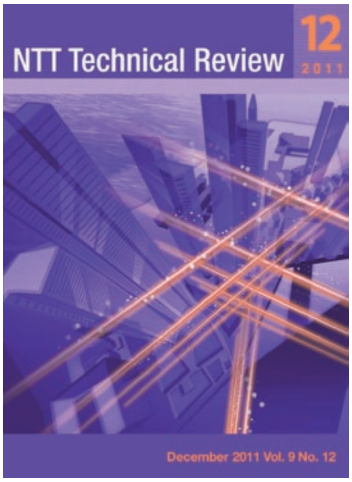
12
2011



December 2011 Vol. 9 No. 12

NTT Technical Review

December 2011 Vol. 9 No. 12



View from the Top

Ryuji Yamada
President and Chief Executive Officer of NTT DOCOMO, Inc.

Feature Articles: Cloud Computing Platform Technologies for Social Infrastructure

NTT Group Initiatives for
Achieving Societal Cloud Infrastructure

Provisioning Infrastructure Supporting Cloud Operations

Large-scale Distributed Data Processing Platform for
Analysis of Big Data

Network Virtualization Technology for Cloud Services

Cloud Traceability (CBoC TRX)

FreeCloud: A Trial Service for OpenStack

Regular Articles

Power-management-circuit Techniques for
Low-power Intermittent LSI Operation in Wireless Applications

Fast Algorithm for Monitoring Data Streams
by Using Hidden Markov Models

Global Standardization Activities

IEC TC86 WG4 Standardization Activities Aimed at
Developing International Standards from Domestic Standards

Practical Field Information about Telecommunication Technologies

Visualization of Problems with Wireless Local Area Networks

External Awards/Papers Published in Technical Journals and Conference Proceedings

External Awards/Papers Published in
Technical Journals and Conference Proceedings

Cultivating Wisdom and Ability by Working Together Toward a Common Objective— ONE docomo: Focusing on Customers, Handsets, the Network, After-sales Service, and Safety & Security

Ryuji Yamada
President and Chief Executive Officer of NTT DOCOMO, Inc.

Abstract

Following the Great East Japan Earthquake of March 2011, NTT DOCOMO worked fervently to restore facilities. With nearly all facilities restored in only about a month and a half, what was it about NTT DOCOMO that helped to achieve such a speedy response? We asked President and Chief Executive Officer Ryuji Yamada to tell us about the ONE docomo in-house campaign to raise customer satisfaction and the diverse services that have been born from this unified spirit at NTT DOCOMO.

Achieving a quick response by getting and sharing all the facts

—Mr. Yamada, please tell us about NTT DOCOMO's response to the Great East Japan Earthquake of March 2011, an unprecedented disaster in the history of Japan.

On hearing the first reports of this earthquake, I realized that it was an extraordinary emergency. The Primary Emergency Response System at NTT DOCOMO was immediately set in motion and I myself went over to the Disaster Countermeasures Headquarters on that day.

Our first task was to check the status of our chain-of-command structure. To achieve a quick response at the time of a disaster, it is essential to obtain a clear understanding of current conditions and share such information. In our case, it was important to understand who would be working on what tasks, such as the restoration of transmission paths and restoration of base stations so that all concerned could proceed with their own duties. To this end, we held meetings three times a day after the earthquake to obtain a detailed view of current conditions. We also set up a system to facilitate quick decisions and action plans for critical matters by having top company leaders come together for meetings.

It was also important for me as president to understand conditions in the field, and I made four visits to the stricken region to get first-hand views. The day after the earthquake, for example, I boarded a helicopter carrying relief supplies, and on the way to Kakuda City near Sendai, I saw with my own eyes the enormous damage caused by the tsunami on top of that caused by the earthquake. Upon seeing all of this, I issued instructions that 95% of our restoration efforts were to be devoted to the Pacific side of the stricken region. I think that this decision helped us mount a quick response.

NTT DOCOMO had no previous experience with such a massive tsunami, and carrying out this restoration was proving to be a huge challenge. We came to realize that this restoration work would have to be a group effort that included affiliated companies. The earthquake disrupted 4900 base stations in the Tohoku region, but nearly all of them were restored by the end of April. In some cases, the damage was so great that base stations and antennas could not be restored at their original sites, so we worked to build up the coverage area by bringing in mobile base stations and installing large-zone base stations.

Working through the night to perform our duties

We received high praise from our customers for our restoration area map. This showed what locations were still without service and which were again providing services, as well as estimates of how many days it would take to restore services in specific areas. Providing this map was actually a bit risky since missing our restoration estimates might lead to complaints and just make things worse, but we felt that providing this information as a convenience for our customers was our first priority.

We worked round-the-clock to develop this system and were able to complete it in only three days. Of course, the state of facilities and service-area restoration was changing day by day, so providing up-to-date information was essential. For this reason, our staff collected new information every evening at around 7 pm and worked through the night to process it so that it could be reflected in the restoration area map in the morning.

Reflecting all lessons learned in future disaster countermeasures

This earthquake reminded us just how crucial telecommunication services are to the safety and security of our customers. There is nothing more important to people during a disaster than being able to contact family members and others they are close to. We used the lessons learned from this devastating earthquake to develop and implement new disaster countermeasures, which we announced on April 28.

For example, we decided to install large-zone base stations at NTT DOCOMO buildings having robust in-house power generation facilities. Base stations of this type will be activated whenever ordinary base stations are put out of commission by a natural disaster. A large-zone base station can cover a circular service area with a radius of approximately seven kilometers, which is far larger than that of ordinary base stations. In fact, installing large-zone base stations at 100 locations throughout Japan would cover about 35% of the population. Given the possibility of major earthquakes occurring in the Tokyo metropolitan area and the Tonankai region of Japan (south-central coast of Honshu island), we decided to prioritize the installation of large-zone base stations in those areas. We have already installed 37 large-zone base stations and plan to install more in order to reach a total of 100 units by the end of December 2011.

We are also moving forward on the development of a voice-file message service. Immediately after the earthquake, the heavy concentration of voice calls on the network caused up to 90% of attempted calls to be restricted, which made it extremely difficult for people to make calls. By contrast, packet-based services like email were relatively easy to use at this time. This situation suggested the possibility of transmitting voice messages by means of packet communications while giving the user the same feel as a normal call when voice traffic becomes congested after a natural disaster. This service allows the user to use an ordinary telephone number to deliver a message and can therefore be used by people who are not familiar with email services. We expect to start providing this service by March 2012.

At NTT DOCOMO, we plan to invest about 23 billion yen (about US$300 million) in the development and implementation of ten new disaster countermeasures, including the ones I just described. Except for a part of those countermeasures, such as the voice-file message service, we aim to complete them by the end of December.

This earthquake taught us the importance of the division of roles between headquarters and the field. While respecting the judgments of field personnel directly involved in disaster response, then headquarters must make important decisions so that NTT DOCOMO can carry out its mission. Headquarters must also strive to look ahead and give instructions in anticipation of future developments. I know that the staff at headquarters works very hard, but we must not let pressures during extremely busy times overwhelm us. In this regard, we began to study the new disaster countermeasures that I just described early in April while facilities were still in the process of being restored, and we were able to make a statement about them at an earnings announcement press conference on April 28. I believe that this quick transition from countermeasure formulation to execution was a result of a united effort by NTT Group companies that understand that the customer is the number-one priority and that we must continue to provide a safe, secure, and trustworthy brand. Working together toward a common objective brings forth wisdom and ability.

Our mission originates out of gratitude to our customers

Disaster countermeasure formulation—and indeed, all of the work that we do at NTT DOCOMO—begins with a sincere feeling of gratitude to our customers. This feeling is essential to achieving true customer satisfaction.

Let me explain this using smartphones as an example. Smartphone usage is growing rapidly, but because they are used differently from conventional mobile phones, we instituted changes in various areas so as not to inconvenience our customers. For example, each of our retail shops has at least one staff member who is a smartphone expert, so that customers with questions or problems can be given authoritative answers. We have also enlarged our smartphone-savvy staff at call centers, which has raised the response rate on smartphone matters from 50 to 90%. Moreover, in the area of service development, we have shifted development resources in order to better handle smartphone needs.

I earlier touched upon the company's chain-of-command structure and the division of roles between headquarters and the field, namely branch offices and retail shops. I can paraphrase these two concepts by saying that *understanding the field* is where our work as a company begins. It is sometimes said that the bigger a company becomes, the more removed it becomes from the field. It is therefore important for headquarters to set up a system to absorb whatever is happening in the field and provide backup support. To this end, key personnel from headquarters must visit the field and make an effort to understand what is actually going on. I myself have been part of a *branch-office caravan* that visits NTT DOCOMO branch offices and retail shops throughout the country. Directors and other top executives at NTT DOCOMO are also assigned different areas to visit. The conclusions reached by the application of logic are correct perhaps about 60% of the time: to make the remaining 40% right, we must talk to field staff in person. Seeing that top executives themselves visit the field to study actual conditions has made a great impression on personnel in various departments at headquarters. They have taken a proactive approach to placing importance on the field, saying that they also want to understand conditions in the field and share information.

On seeing this kind of energetic response by headquarters personnel, I feel that NTT DOCOMO is blessed with highly capable staff having great potential to come up with answers to problems. Once a target has been established, they go to work demonstrating great ability. However, there was a time when we fell short in this regard. During the i-mode era, we may have put too much emphasis on technology. But we learned from this experience, and in J. D. Power's customer satisfaction survey for mobile phone services, we were ranked number one in the personnel sector for two years straight (2010 and 2011) and number one in the corporate sector for three years straight (2009–2011). Our renewed focus on customer satisfaction is also reflected in handset sales. In the previous fiscal year, our customers purchased about 19 million mobile phones.

Customers, handsets, the network, after-sales

service, and safety & security: these five areas are basic to our work at NTT DOCOMO. Looking forward, I want NTT DOCOMO to make great progress while achieving a balance among these five basic elements of our work.

Becoming an integrated service company: NTT DOCOMO's vision for the future

—As President and Chief Executive Officer, what are your present concerns and what do you see as NTT DOCOMO's future vision?

We have been galvanized in various ways through our exchanges with overseas organizations and companies. What really strikes us here is how movements like the spread of smartphones and social networking sites and services have brought about rapid and profound changes in the world. In such a tumultuous era, I believe that building up our research and development (R&D) capabilities is of vital importance. For example, Long Term Evolution (LTE) technology, the basis for the LTE Xi (Crossy) Service that we launched last year on December 24, was first proposed by NTT DOCOMO and approved for standardization by the 3rd Generation Partnership Project (3GPP) in 2004. In this way, we would like to commercialize technology that we have researched and developed through cooperation with various countries and companies.

Moreover, in the mobile field, I feel that an *age of convergence* is coming. The evolution of technology centered on mobile communications should stimulate innovation in diverse business areas through the convergence of industry and services and create new value for users. For example, I can envision the convergence of mobile with various types of equipment such as personal digital assistants and car-navigation devices and the convergence between mobile and medical/healthcare services. We can also consider the convergence between mobile and content (media), convergence between broadcasting and communications, and convergence between the energy and ecology fields, just to name a few possibilities.

Against this background, NTT DOCOMO announced its "Medium-Term Vision 2015—Shaping a Smart Life—" on November 2, 2011. On the basis of this vision, I would like to promote the evolution of services, content, and device operability through an open development environment for a wide variety of devices centered on the smartphone and to pursue services and products that provide our customers with even more enjoyment and convenience.

Furthermore, with the aim of becoming an integrated service company centered on mobile communications, I would like to stimulate innovation and create new markets by forming alliances with other companies and converging mobile with various types of industries and services. Our aim is to provide a truly *smart life* for all by using the DOCOMO cloud to accelerate this evolution of services and the convergence of mobile communications, industry, and services. In this way, daily life and business activities will become more secure, more convenient, and more efficient.

—Mr. Yamada, could you leave us with a message to the researchers that support NTT DOCOMO activities?

Yes, I would like to ask our R&D personnel to execute their work with a sense of urgency. Trends in the mobile phone industry are particularly fast moving, and instead of mistiming the market by pursuing perfectionism in products, I think that it is better to achieve 80% of the product objectives and enter the market at just the right time. Of course, achieving 100% in safety-related matters is an absolute requirement. And we can think of the remaining 20% as requirements that should reflect feedback from customers in later enhancements. Please don't forget that good timing and skillful reading of trends are very important. I have great expectations that many services will be coming forth from NTT DOCOMO R&D.

Interviewee profile

■ Career highlights

Ryuji Yamada received the M.E. degree in telecommunication engineering from Osaka University's Graduate School of Engineering and joined Nippon Telegraph and Telephone Public Corporation (now NTT) in 1973. In 1994, he took the lead in drawing up NTT's strategic plan for transition from voice-oriented services to advanced IP (Internet protocol) services for the Internet age, which continue to underpin the NTT Group's long-term vision. Over the years, he was also a central figure in planning NTT's 1.7-trillion-yen network of nationwide facilities. From July 1999 to June 2004, he held various top managerial positions at NTT WEST, where he played a key role in making the company profitable just three years after NTT's reorganization in 1999. In June 2004, he took up the post of Senior Executive Vice President for NTT, where he oversaw its world-class research center and the development of the company's Next Generation Network (NGN). Since joining NTT DOCOMO as a Senior Executive Vice President, a Member of the Board of Directors, and Managing Director of the Corporate Marketing Division in June 2007, he has contributed greatly to the growth and advancement of the mobile market for corporate customers. He assumed the posts of President and Chief Executive Officer of NTT DOCOMO in June 2008.

NTT Group Initiatives for Achieving Societal Cloud Infrastructure

Atsushi Abe[†], *Tsuyoshi Ochi, Akira Shirahase, and Fusayoshi Kumada*

Abstract

Cloud technology, which is a major trend in the field of information and communications technology, is expected to develop as a core technology for broad support of societal infrastructure in the future. In this article, we overview overall trends in cloud business and technology and then introduce the directions that the NTT Group is pursuing in cloud business, NTT R&D initiatives toward achieving them, and our main technical developments.

1. Introduction

1.1 Cloud business trends

It has already been several years since the term cloud computing came into use. Some people think that it is merely a buzzword, but all major vendors and systems integration providers in Japan are setting policies that have cloud business as a major pillar of their business strategy, and it is recognized as a paradigm shift in the field of information and communications technology (ICT). First, we provide a general overview of cloud business and its two main trends.

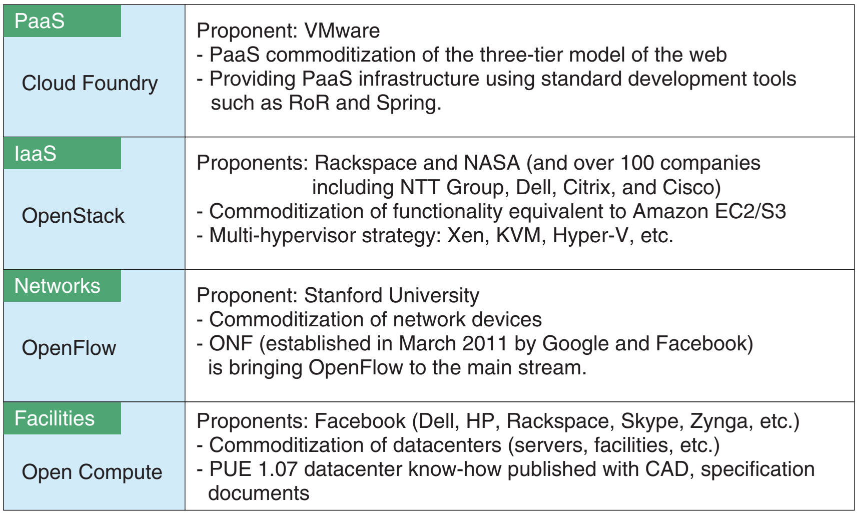
The first trend is related to public clouds, which have been developed by major players in North America such as Google and Amazon. Both are expanding their businesses in the fields of advertising on a web search site and electronic commerce on the Internet, respectively, but in the process, they have invested in computing-resource equipment and have established scale-out* technologies for operating these large-scale systems. These scale-out technologies utilize distributed parallel processing on multiple servers and together they provide a single service (search, product recommendations, etc.). Using the facilities, operational expertise, and technical power developed in building their main business as a base, Google, Amazon, and others have developed public clouds in order to lease out facility resources on a time basis. Google Apps, the Google App Engine, Amazon Web Services, etc. are now being used, not only by consumers and startup companies, but also by ordinary companies and public services, mainly for non-mission-critical systems.

The other main trend in cloud business is toward private clouds. Companies such as VMware have established virtualization technology that allows individual physical servers to be divided into multiple virtual servers. This allows computing resources to be used efficiently and enables enterprise systems to be built with flexibility and speed, at reduced cost. This has facilitated the development of a private cloud business for enterprise and promoted the migration of corporate systems, which were previously provided through on-premises (company operated) systems integration, onto private cloud systems. This movement is still mainly focused on non-mission-critical systems, but it has included some mission-critical systems and there are examples of companies building global private clouds, particularly in manufacturing and distribution industries. Commercial services for hybrid clouds, which combine private and public clouds, have also started to be offered in Japan.

The technologies supporting these trends are by no means new, and the appearance of such cloud services

† NTT Research and Development Planning Department
Chiyoda-ku, Tokyo, 100-8116 Japan

* Scale-out: Increasing the performance of the overall server group by increasing the number of servers in the group.



PaaS Cloud Foundry	Proponent: VMware - PaaS commoditization of the three-tier model of the web - Providing PaaS infrastructure using standard development tools such as RoR and Spring.
IaaS OpenStack	Proponents: Rackspace and NASA (and over 100 companies including NTT Group, Dell, Citrix, and Cisco) - Commoditization of functionality equivalent to Amazon EC2/S3 - Multi-hypervisor strategy: Xen, KVM, Hyper-V, etc.
Networks OpenFlow	Proponent: Stanford University - Commoditization of network devices - ONF (established in March 2011 by Google and Facebook) is bringing OpenFlow to the main stream.
Facilities Open Compute	Proponents: Facebook (Dell, HP, Rackspace, Skype, Zynga, etc.) - Commoditization of datacenters (servers, facilities, etc.) - PUE 1.07 datacenter know-how published with CAD, specification documents

CAD: computer aided design

Fig. 1. Commoditization of cloud technology.

has been made possible by increases in computer and network performance. The performance of servers and networks has improved to the level that they can be used for virtualization and distributed processing, and this has created new needs and planted the seeds for new forms of services. Both these seeds and these needs could also lead to further new trends in the future as well.

1.2 Cloud technology trends

Next, we briefly touch upon some of the latest trends in cloud technology, including the commoditization of technologies for cloud infrastructure, the commercialization of network virtualization technology and mobility of cloud resources, and large-scale data processing technology.

(1) Commoditization of technologies for cloud infrastructure

Conventionally, products used in the cloud industry have been centered on vendor-specific technologies, but there has recently been an accelerating trend toward open standards and commoditization. Open community initiatives are receiving particular attention with Open Compute, led by Facebook, at the facilities layer; OpenFlow from the Open Network Foundation (ONF) at the network layer; OpenStack from Rackspace, NASA, and others at the infrastructure-as-a-service (IaaS) layer; and even Cloud Foundry from VMware and others at the platform-as-a-service (PaaS) layer, as shown in **Fig. 1**. Commoditization at each of these layers is expected to develop further in the future. Open technologies concentrate knowledge that exceeds the frameworks of individual companies and other organizations, so it will be increasingly important to embrace these trends correctly and apply them skillfully.

(2) Commercialization of network virtualization technology and mobility of cloud resources.

Virtual server and virtual storage services have already been offered using the virtualization technologies, as exemplified by services such as Amazon EC2 and S3. In addition to these, it has recently become possible to offer virtualized network equipment, such as switches and routers, and even virtualized firewalls and load balancers. Also very significant is the rapid commercialization of such network virtualization using OpenFlow, with the open technology discussed above, as the configuration-and-control protocol for virtual network equipment. A significant aspect of this network virtualization is the improvement in the *mobility* of cloud resources (**Fig. 2**). It is currently possible to migrate individual servers within the same location or to a different location in a relatively short time, either according to a plan or after a fault has occurred. However, for the entire system including the network to be usable continuously, additional operational preparations are required, such as configuration of the destination network configuration in advance. By contrast, network virtualization technology enables network setup

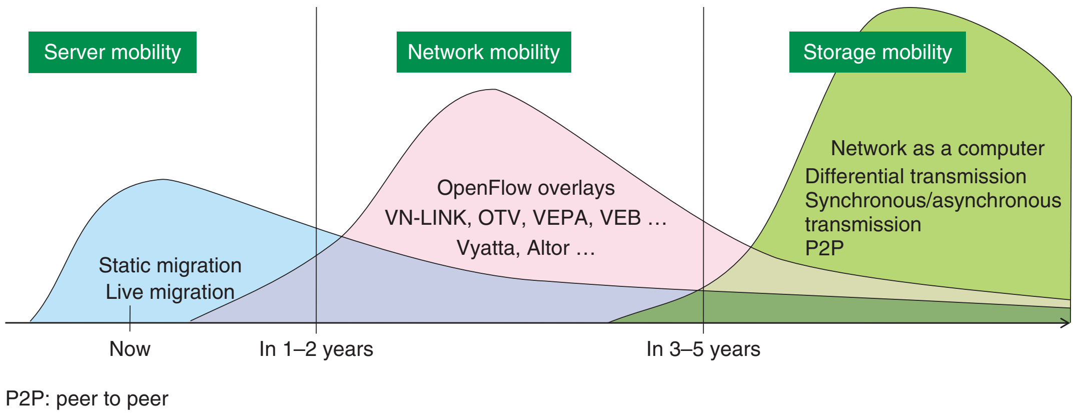


Fig. 2. Virtualization and mobility.

and modification to be done rapidly and on demand, so migration becomes much easier. For storage migration, it is technically difficult to migrate large volumes of data instantly owing to network bandwidth limitations, so operational issues must also be handled by, for example, copying the data beforehand and updating it whenever there are changes.

(3) Large-scale data processing technology

In the past, products such as business intelligence analysis software and data warehouse appliances were used as decision-making support tools for marketing, production, and other areas in enterprise. Recently, however, open source software (OSS) called Hadoop, which is equivalent to Google's technology used in its search processing and can process even larger-scale data, has been gaining attention and is being used more and more. Moreover, with the spread of sensor technology and embedded systems, an increasing variety of data is expected to be gathered in the future. Use of this huge volume of data in the public and consumer fields as well as enterprise is also being expected [1]. Research and development (R&D) is also progressing on realtime analysis of such sensor data.

2. NTT's cloud strategy

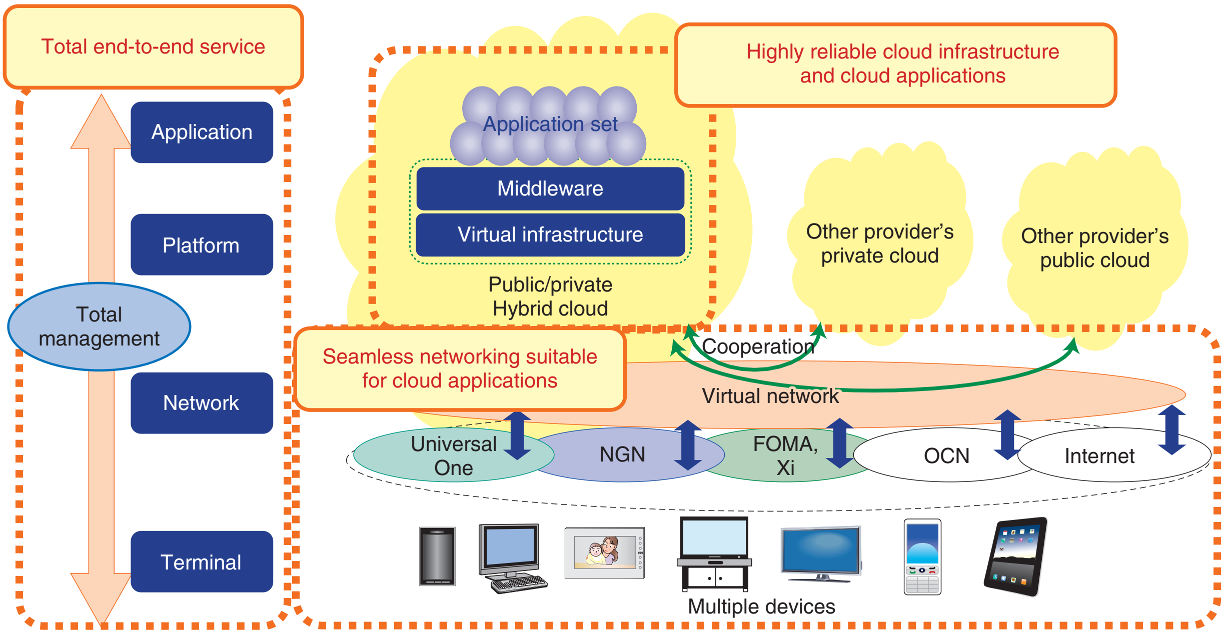
2.1 NTT's objectives for cloud services

In 2008, the NTT Group established a SaaS-over-NGN task force with participation from NTT Group companies (100 members from 16 companies) to start and grow the software-as-a-service (SaaS) business and provide a SaaS infrastructure to service providers (NGN: Next Generation Network). The scope of this task force was further expanded in spring of 2010, from SaaS to the cloud, incorporating the changing conditions in technology and cloud business. The group was renamed the SaaS/Cloud task force to promote and expand the cloud business within the NTT Group overall. Within this task force, the following basic policies were decided regarding NTT's objectives in cloud services.

- Provide a safe, secure, and highly reliable cloud
- Provide optimized services combining applications, platforms, networks, and terminals
- Make a cloud as a societal infrastructure through openness and collaboration

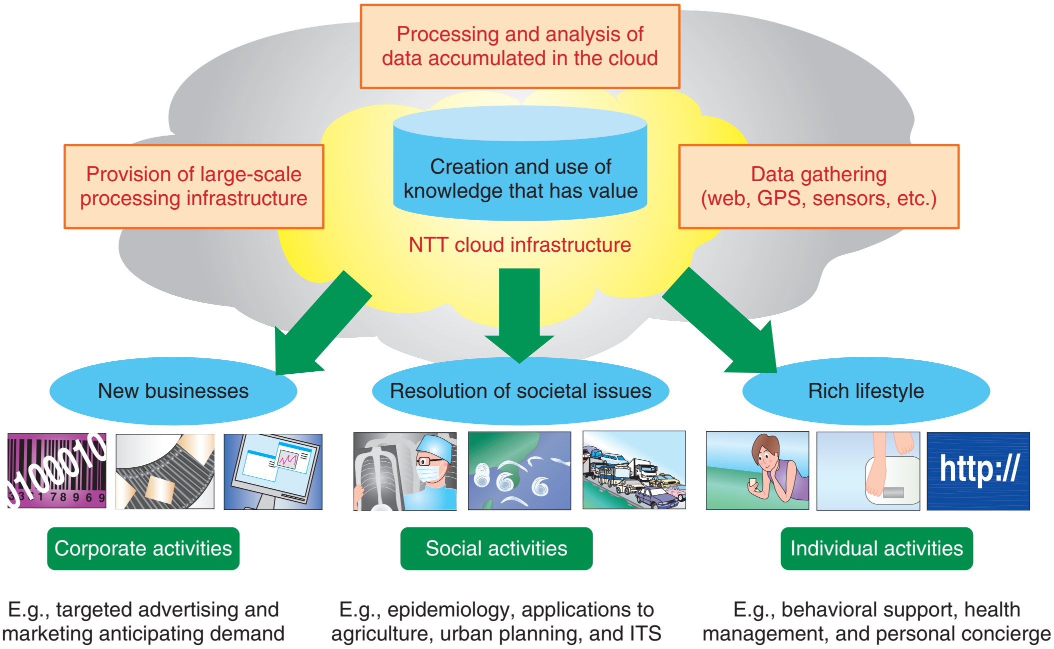
NTT's objectives in cloud services, based on these basic policies, are shown in **Figs. 3** and **4**. They are supported by the following four pillars. We are advancing development of services to maximize the strengths of each NTT Group company and the strength of the whole group.

1) Provide total operation and one-stop services, including networking.
2) Provide highly reliable cloud infrastructure and cloud applications.
3) Provide seamless networking suitable for cloud applications.
4) Create valuable knowledge and solve societal issues through the accumulation and analysis of data in the cloud.



OCN: open computer network

Fig. 3. NTT Group cloud objectives.


Processing and analysis of data accumulated in the cloud
Provision of large-scale processing infrastructure
Creation and use of knowledge that has value
Data gathering (web, GPS, sensors, etc.)
NTT cloud infrastructure
New businesses
Resolution of societal issues
Rich lifestyle
http://
Corporate activities
Social activities
Individual activities
E.g., targeted advertising and marketing anticipating demand
E.g., epidemiology, applications to agriculture, urban planning, and ITS
E.g., behavioral support, health management, and personal concierge

GPS: global positioning system
ITS: intelligent transportation systems

Fig. 4. Creating new value through use of large-scale data.

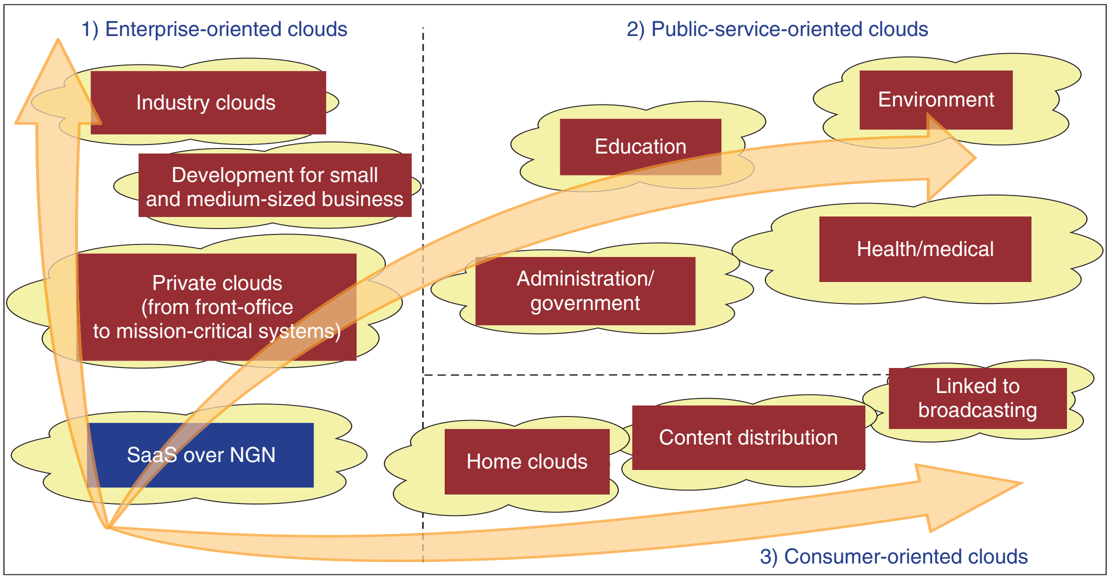

1) Enterprise-oriented clouds
2) Public-service-oriented clouds
Industry clouds
Development for small and medium-sized business
Private clouds (from front-office to mission-critical systems)
SaaS over NGN
Environment
Education
Health/medical
Administration/ government
Linked to broadcasting
Content distribution
Home clouds
3) Consumer-oriented clouds

Fig. 5. Fields for development of NTT Group cloud businesses.

2.2 Development of cloud businesses

As shown in **Fig. 5**, NTT Group cloud services are being developed for enterprise, public, and consumer services, with the issues in each field being resolved.

1) Enterprise-oriented clouds: We are contributing to the expansion of cloud applications from front-end systems, which are currently the major area, to include mission-critical systems and we are expanding our customers' systems globally. We are also expanding development to include small and medium-sized businesses and industry-specific clouds.
2) Public-service-oriented clouds: We are expanding mainly in the areas of healthcare, regional government, and education, as reflected in government directions for the utilization of cloud technology.
3) Consumer-oriented clouds: We are developing these with a focus on home clouds and content distribution services, which can utilize NTT's strengths.

3. NTT R&D initiative policies

To promote cloud activity based on its cloud strategy, NTT is advancing the following policies and developing the overall cloud business of the NTT Group and the technology needed for it.

3.1 Business development

In business development, NTT is working to integrate the strengths of each group company and create incremental added value by promoting cooperation among them within the three fields discussed earlier: enterprise, public, and consumer-oriented services. We are also promoting cooperation with strategic partnerships across group companies. For example, with Microsoft, we have an initiative to develop a hybrid cloud that cooperates with Windows Azure. We also plan to further expand our cloud business through collaboration with new vendors, service providers, and other business partners.

3.2 Technical development

Technical development on the cloud infrastructure needed to develop businesses can be divided broadly into initiatives for cost reduction and initiatives for differentiation. More specifically, the cloud infrastructure consists of *execution systems*, which perform the actual processing, *operations/management systems*, which manage the execution systems, and *security systems*, which handle security for the execution and operations/management systems, as shown in **Fig. 6**. Each of these is further divided into major functional groups, and we have initiatives in each of these areas, advancing development for cost reduction utilizing commoditized products and for differentiation from competitors.

For parts of the execution system in particular, from

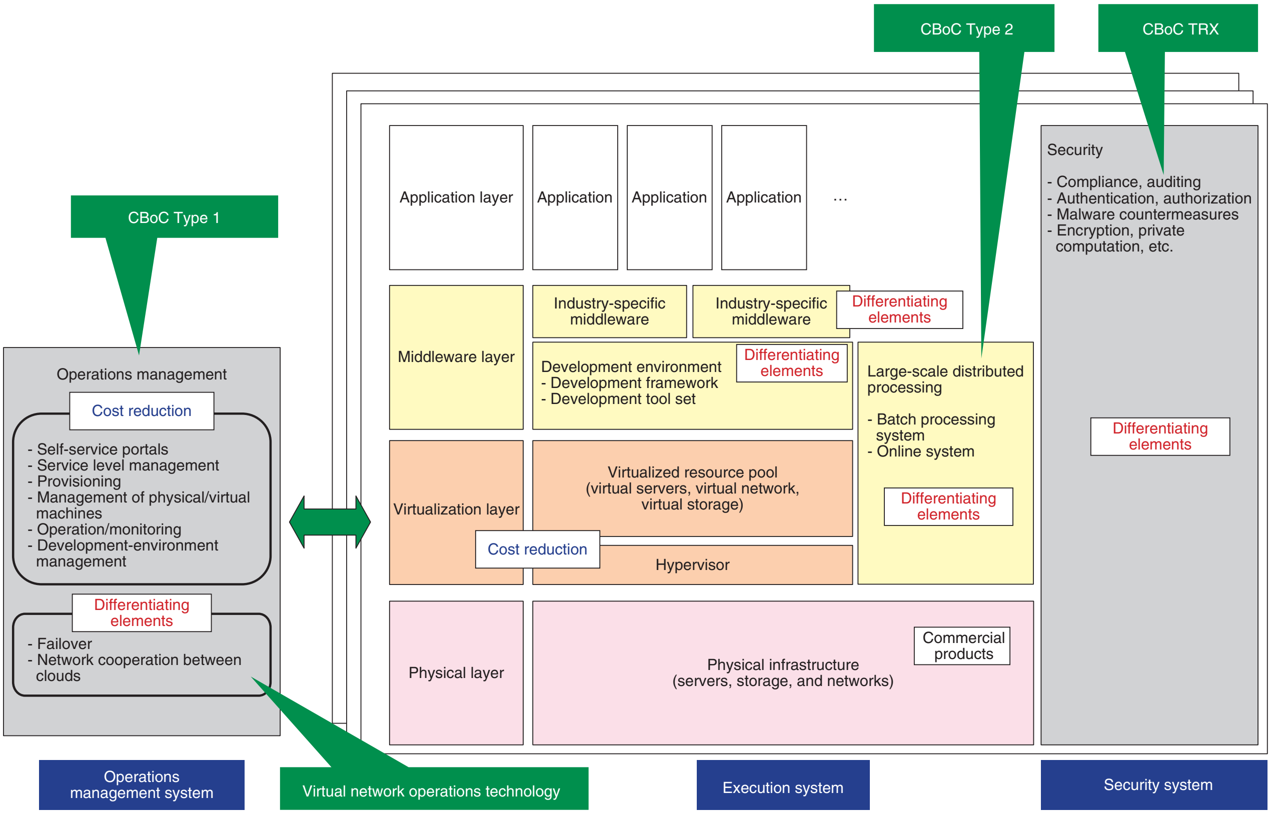

CBoC Type 2
CBoC TRX
CBoC Type 1
Application layer
Application
Application
Application
…
Security
- Compliance, auditing
- Authentication, authorization
- Malware countermeasures
- Encryption, private computation, etc.
Operations management
Cost reduction
- Self-service portals
- Service level management
- Provisioning
- Management of physical/virtual machines
- Operation/monitoring
- Development-environment management
Differentiating elements
- Failover
- Network cooperation between clouds
Middleware layer
Industry-specific middleware
Industry-specific middleware
Differentiating elements
Development environment
- Development framework
- Development tool set
Differentiating elements
Large-scale distributed processing
- Batch processing system
- Online system
Differentiating elements
Differentiating elements
Virtualization layer
Virtualized resource pool (virtual servers, virtual network, virtual storage)
Cost reduction
Hypervisor
Physical layer
Physical infrastructure (servers, storage, and networks)
Commercial products
Operations management system
Virtual network operations technology
Execution system
Security system

Fig. 6. Cloud infrastructure development policy.

the facilities layer to the PaaS layer, we can use open technologies as discussed above, and there are several technologies and specifications that are expected to become commoditized in the future. We will use these proactively, and by proposing technical specifications and providing source code from NTT, we will also contribute to these open communities. We are also promoting standardization activities needed for cooperation among clouds through the Global Inter-Cloud Technology Forum (GICTF) and in other ways.

On the other hand, in R&D for differentiation, we have our own development initiatives centered on virtual network operations technology, large-scale data distributed processing technology, cloud security technology, etc. and we are implementing applications of these leading-edge technologies in our cloud businesses.

3.3 Promotion of openness and collaboration

We have developed a cloud environment for R&D (R&D cloud) as a place for rapidly and smoothly promoting business- and technology-development initiatives in order to accelerate the cycle of selecting requirements and issues from the business side and providing technologies to resolve the issues from the technical development side.

4. Technical development overview

At NTT Information Sharing Platform Laboratories, we are developing technology covering the technical development policies and trends discussed above. Below, we give an overview of the main initiatives. The name Common IT Bases over Cloud Computing (CBoC) used below is NTT's name for this development (IT: information technology).

4.1 Technical development for cost reduction in commoditized technical field

(1) Cloud infrastructure operation and management technology (CBoC Type 1)

Although many commercial and OSS products are already available in the field of cloud computing, technical development in the field is advancing very rapidly and many more products are expected to be available in the future. Consequently, one feature of CBoC Type 1 is that its architecture allows the execution system stack to be built flexibly using the best commercial and OSS products available at the time. We are also working to establish an operations technology for managing effective use of each type of resource (servers, storage, networks, etc.) being provided on the cloud to enable resources to be increased or decreased according to the load and to implement a mechanism for life cycle management, from construction to renewal of the cloud infrastructure. This will handle the whole cycle of provisioning, operations monitoring, and capacity planning.

(2) OpenStack initiatives

One of our initiatives with open technology is our proactive participation in the OpenStack community. OpenStack is an open-source cloud infrastructure technology providing functions equivalent to Amazon's EC2/S3. It has an extremely active community: the main proponents are Rackspace and NASA, but there are over 100 other companies participating, including NTT DATA, DOCOMO Innovations, and NTT from the NTT Group. OpenStack is very likely to become an industry standard in the future. In anticipation of using this software commercially, we are proposing and working on additional functions and improvements that will be required for such use and we are ensuring and improving software quality. We are also participating in the FreeCloud project, which is operating an OpenStack trial service; through this project, we are contributing to improving the operability of OpenStack.

4.2 Technical development for differentiation

(1) Large-scale distributed processing infrastructure (CBoC Type 2)

Large-scale data analysis technology is expected to be one of the core technologies for accumulating and analyzing data in the cloud, creating valuable knowledge, and resolving societal issues. These large-scale data analysis systems are composed mainly of 1) data gathering functions, 2) data storage and management functions (large-scale distributed processing infrastructure), and 3) data analysis functions. NTT is studying technology for designing and operating the overall architecture of these large-scale data-analysis systems. It is working particularly on a large-scale distributed processing infrastructure as a differentiating technology and also working to develop technology for reliability, operability, and performance.

(2) Virtual network operations technology

In cloud datacenters, with the trend toward higher concentration and multi-tenant usages through server virtualization, designing and making changes to networks within datacenters has become extremely complex. Because of this, as discussed earlier, technology that allows virtual networks to be configured on demand is attracting attention. NTT is integrating network virtualization technology into cloud systems and developing technology for operating total cloud systems. This allows network migration linked with server migration to be done automatically, and this can be applied, for example, as a disaster recover solution or as a means to reduce power consumption by optimizing server resource deployment.

(3) Traceability infrastructure (CBoC TRX)

With current cloud services, it is impossible to know what equipment is actually providing the service or what its operational state is, and this lack of transparency looks like a barrier to many customers from the security and operational perspectives. In response to this, we are developing technology that will provide an audit trail of what actually happened, together with visualization and certification of its safety and security, by reproducing and displaying the sequence of events in an easy-to-understand manner from various types of log data related to the human operation, content migration, and processes and the related systems distributed throughout the cloud.

4.3 NTT R&D cloud initiatives

We have consolidated the R&D cloud, which consists of well over 1000 servers, connecting R&D centers and providing a place to promote collaboration within NTT laboratories and the NTT Group. The main initiatives being undertaken with this R&D cloud are as follows.

- Building a comprehensive testbed fully utilizing laboratory technologies, from datacenter facilities to application infrastructure, and establishing comprehensive technology for implementing radically increased productivity, energy conservation, and operational efficiency through cloud technology.
- Providing a location for collaborative testing among group companies and creating prototypes of new services and solutions through testing of the cloud stack.
- Providing a place where researchers themselves can use products developed in the laboratory on

the R&D cloud, which will promote further improvements to laboratory products.

This article has given an overview of technical development in NTT. The other Feature Articles in this issue give details of each technology.

5. Conclusion

Cloud technology is a core field of ICT and will be an important element supporting societal infrastructure in the future. The NTT Group is contributing to the creation of a cloud that will support this societal infrastructure through group-wide initiatives and the implementation of its advanced technologies.

Reference

[1] N. Uji, "Kuraudo ga kaeru sekai: The Cloud Revolution: How Cloud Computing is Transforming Both Business and Society," Nikkei Publishing Inc., Aug. 2011 (in Japanese).

Atsushi Abe

Senior Research Engineer, R&D Produce Group, NTT Research and Development Planning Department.

He joined NTT in 1995 and worked on traffic engineering of ATM networks and the development of IP VPN systems. He also engaged in the development of the Next Generation Network and commercialization of R&D products. Since moving to NTT Research and Development Planning Department in 2010, he has been engaged in R&D strategy planning for cloud computing business.

Tsuyoshi Ochi

Manager, R&D Produce Group, NTT Research and Development Planning Department.

He joined NTT Packet Communication Division in 1993 and engaged in the development of X.25 packet switching systems. After moving to NTT Communications in 2000, he worked on the development of remote access systems for IP VPN services. Since moving to NTT Research and Development Planning Department in 2009, he has been engaged in promoting cloud computing business.

Akira Shirahase

Senior Manager, R&D Produce Group, NTT Research and Development Planning Department.

He joined NTT in 1992 and worked as a systems engineer in corporate sales and marketing, especially for Internet systems. He was involved in starting up the Internet datacenter business at NTT Smart Connect and in business development at NTT WEST. Since moving to NTT Research and Development Planning Department in 2010 he has been engaged in promoting cloud computing business for the entire NTT Group.

Fusayoshi Kumada

Vice President, R&D Produce Group, NTT Research and Development Planning Department.

He joined NTT in 1985. He moved to NTT Data Communications as a manager in 1992. From 1996 to 2008, he was a senior manager in NTT DATA and engaged in corporate business, logistics business, and corporate business consulting and marketing. Since moving to NTT Research and Development Planning Department in 2008, he has been engaged in promoting Home ICT services, Healthcare ICT services, and cloud computing business.

Provisioning Infrastructure Supporting Cloud Operations

Kenichi Sato[†], *Hideki Hayashi, and Ken Ojiri*

Abstract

With cloud technology, computing resources are provided dynamically in response to requests from users, so the function that manages normal operation of these resources (the management system) is very important. In this article, we describe a provisioning infrastructure, which is one of the main components of the management system within Common IT Bases over Cloud Computing (CBoC), a cloud infrastructure system being developed by the NTT Information Sharing Platform Laboratories (IT: information technology).

1. Introduction

NTT Information Sharing Platform Laboratories is conducting research and development (R&D) on a cloud for large-scale distributed processing that can act as societal infrastructure. This cloud is composed of an execution system and a management system (**Fig. 1**). The execution system is the set of functions providing specific computing resources to users, including virtual machines and storage, and the management system is the set of functions that manage the operation of the execution system so that services (computing resources) can be provided appropriately to the users. In this article, we describe the cloud management system.

This cloud can be regarded from several perspectives, including those of the cloud operator, the cloud services provider, and the cloud services users. In this article, we treat the cloud management system, so we focus on the operator's perspective and the services provider's perspective.

2. Cloud operation model

The cloud operations management task cycle, as conceived by NTT, is shown in **Fig. 2**.

In response to user requests, provisioning functions control the execution system to provide computing resources. Monitoring functions check the health of the computing resources being provided (whether continuous service is being provided) and the service level (whether the service quality is appropriate). The capacity planning function calculates the appropriate amount of resources on the basis of monitoring results and creates configuration change instructions for the provisioning functions. Initial provisioning is done on the basis of estimates calculated by the applications provider, so as the services being provided on the cloud mature, computing resources could become either excessive or insufficient compared with requirements owing to inaccuracies in the estimates. The purpose of the capacity planning function is to make these adjustments appropriately, but at present there is no well-defined method of accomplishing this. In current operation, the user re-calculates the estimate on the basis of the monitoring results.

Note that this cycle of provisioning, monitoring, and capacity planning is not a new concept with cloud computing. It is basically the same as has been done for various services in the past. In fact, such adjustments were done every few months or years in the past, but with cloud technology, they can now be done every few minutes or tens of minutes. The automation of this operational cycle has made it possible to operate with fine adjustments being made to equipment according to demand, which was difficult to do in the past.

Below, we describe mainly the provisioning function, which is one of the functions in the cloud operations management model.

[†] NTT Information Sharing Platform Laboratories
Musashino-shi, 180-8585 Japan

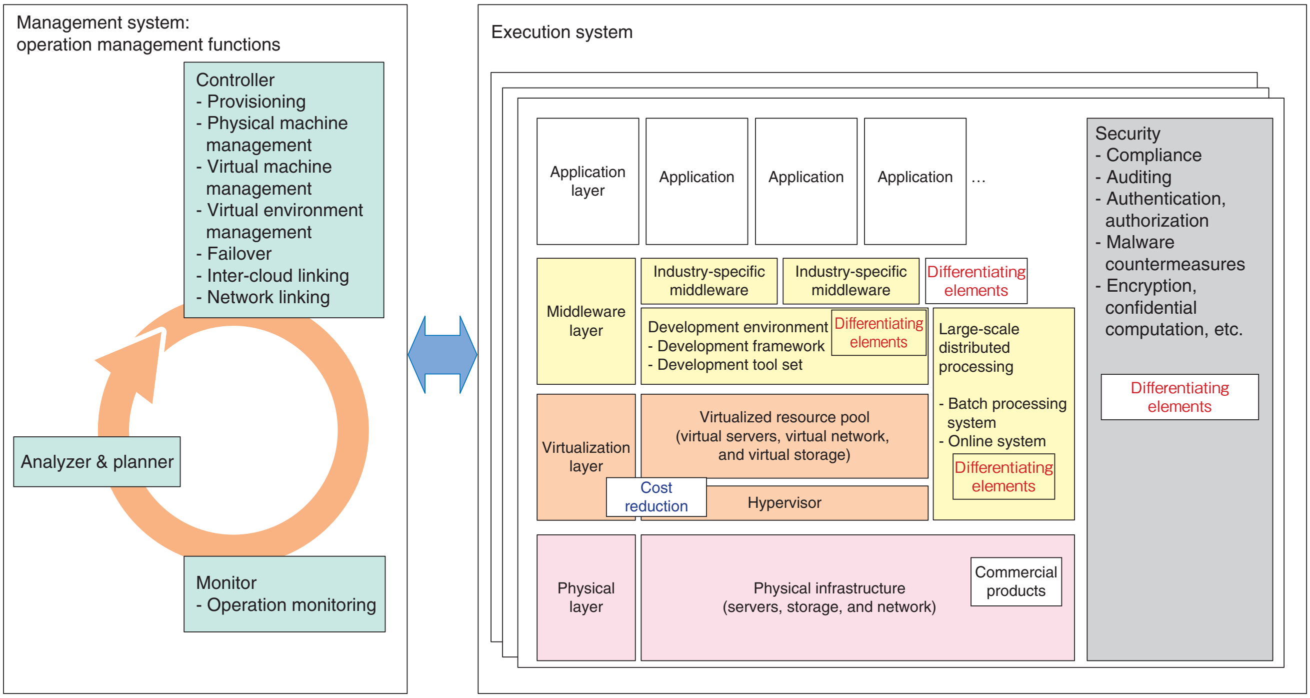


Fig. 1. Cloud configuration.

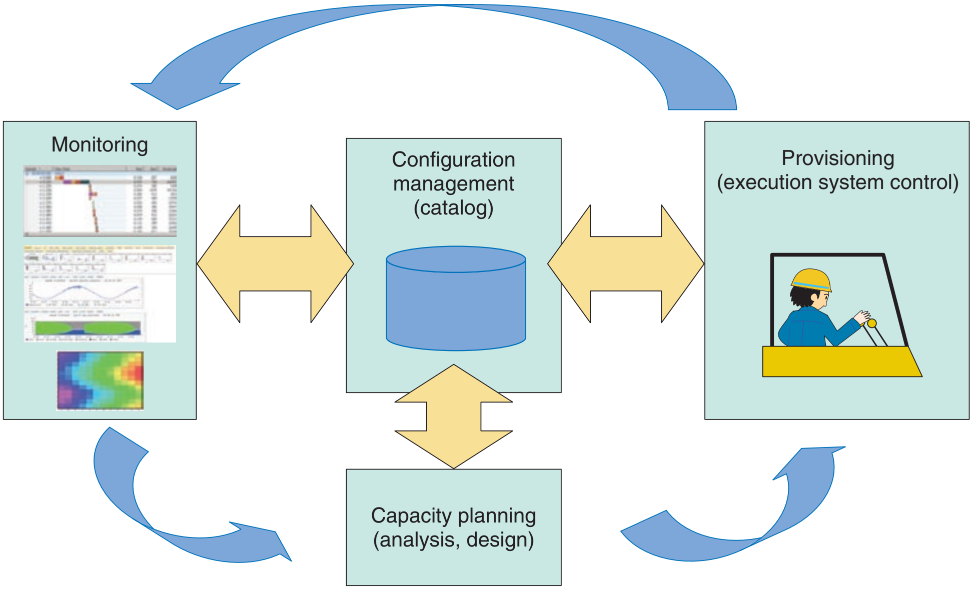

Monitoring
Configuration management (catalog)
Provisioning (execution system control)
Capacity planning (analysis, design)

Fig. 2. Cloud operations-management task cycle.

3. CBoC provisioning infrastructure

Next, we describe the Common IT Bases over Cloud Computing (CBoC) provisioning infrastructure being developed by NTT (IT: information technology).

3.1 Development goals

There are already several cloud services in existence, including that of the NTT Group. On-demand self-service [1] is a widely accepted property of clouds, performing resource provisioning according to instructions received from users, so the management system, and in particular the provisioning function, is a basic feature of a cloud service.

These existing cloud services provide various types of services at various layers, according to business needs. This trend is expected to accelerate in the future [2].

The provisioning functions in existing services can be considered to optimize specific computing resources according to their properties. On the other hand, the speed with which the provisioning function can be developed when providing a new computing resource must not become an obstacle to the business.

Thus, with the CBoC provisioning infrastructure, we have abstracted the control model so that various computing resources can be supported very quickly, helping to accelerate business development. The operation of various computing resources can also be optimized daily (automatically in the future) through the operational cycle discussed above.

3.2 Central concepts

Below, we describe the central concepts implemented by the features of the CBoC provisioning infrastructure, which allow various computing resources to be added quickly and applications to be integrated easily.

3.2.1 Resource abstraction and connection between resources

One question that arises is whether particular computing resources will behave differently, though we talk about all resource types together. However, from a provisioning perspective, the important things are the fixed operations: create, allocate, initialize, activate, deactivate, finish, free, and delete. If we consider a specific example, these fixed operations apply to both virtual machines (kernel-based virtual machine (KVM) and VMware) and virtual local area networks (VLANs), allowing them to be handled in a unified way by the provisioning infrastructure. The CBoC provisioning infrastructure uses this concept to abstract all of the objects that it handles as resources (of course, in addition to the fixed operations, it also provides ways to handle attributes and operations particular to a given resource).

Another question that arises is whether the process of connecting a virtual machine to a LAN can be abstracted. However, connections between resources are also abstracted, including definitions for a connection's source and destination as well as connection operations, and these definitions regulate what connections and what connection operations are possible. In this way, virtual machines and VLANs can connect, and higher-layer connections, such as those between an application and a database, can also be automated.

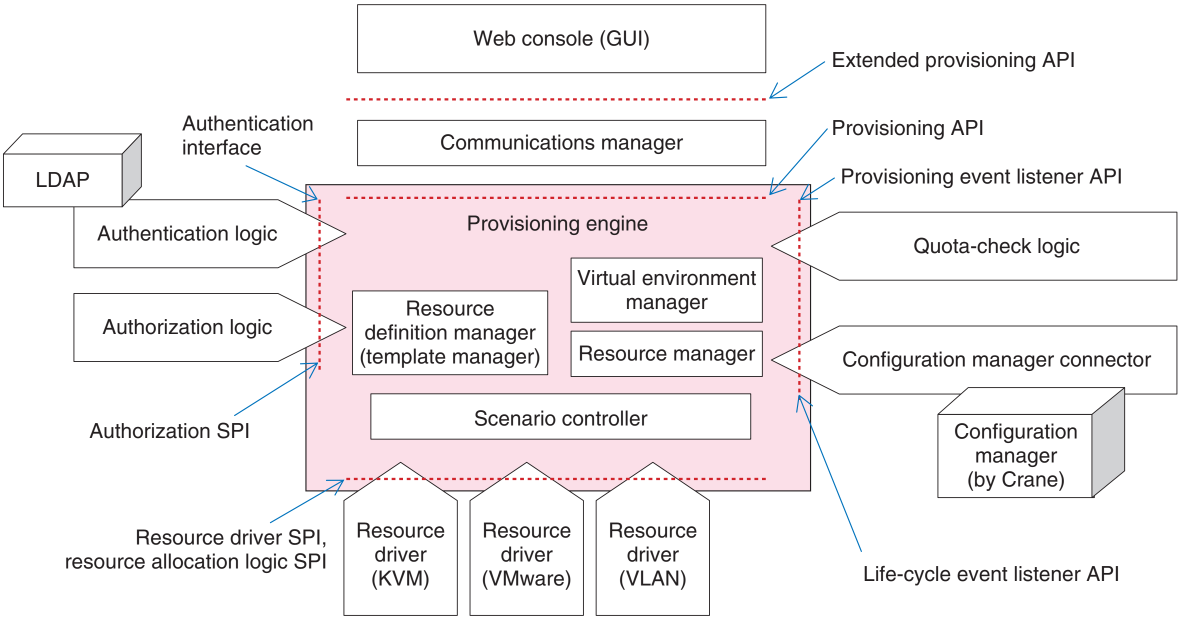
This abstraction enables new resources to be added easily to the CBoC provisioning infrastructure. By creating a driver that operates the execution system and by writing template data (resource definitions, connection definitions, and connection operation definitions), one can integrate a new resource into the system.

The development required when new resources are added is also a concern. It is true that creating a resource driver involves development. However, a resource re-definition procedure can be used, even for small-scale cases. Resource re-definition is a function for creating new template data from the existing resource status and its template data, without writing programs or creating template data from scratch. This function makes it easy to perform operations such as creating a new template of a virtual machine with installed applications from a virtual machine with only the installed operating system.

3.2.2 Virtual environment operation, asynchronous scenario processing

In many cases with cloud services like Amazon's, the focus is on handling individual virtual machines. On the other hand, in ordinary system development, it is instead more common to bundle multiple processors and treat them as a single system or development environment. The CBoC provisioning infrastructure uses a virtual environment approach, modeling the system being provided by the user. When a system is being developed or a service is being provided, multiple resources can be conveniently bundled in a virtual environment, enabling batch operations (start, end, backup, etc.).

Resource operations are basically considered to be asynchronous. Operations on real resources such as



API: application programming interface
GUI: graphical user interface
LDAP: Light Directory Access Protocol
SPI: service provider interface

Fig. 3. CBoC provisioning infrastructure components.

launching a virtual machine usually take time, so processing is done asynchronously with respect to user requests.

Asynchronous processing is relatively easy to implement for individual virtual machine operations, but for batch operations in a virtual environment with multiple virtual machines, it is more complex. As an example, what would need to be done in a state where one virtual machine was stopped, another virtual machine was assigned but was not processing yet, and you want to transition them to a completed state all at once? With the CBoC provisioning infrastructure, operating procedures are generated automatically, taking into account the current state of resources bundled in the virtual environment and the state of connections between them.

3.2.3 CBoC provisioning infrastructure configuration

Keeping in mind the need to accelerate business development, the various components of the CBoC provisioning infrastructure were designed to be pluggable, with well-defined interfaces, so that only the required parts for a given application need be developed or modified. These components are shown in **Fig. 3**. The types of interfaces and their applications are also listed in the **Table 1**.

4. Application examples and effect

Next, we describe some examples of applying the CBoC provisioning infrastructure and the effect of these applications.

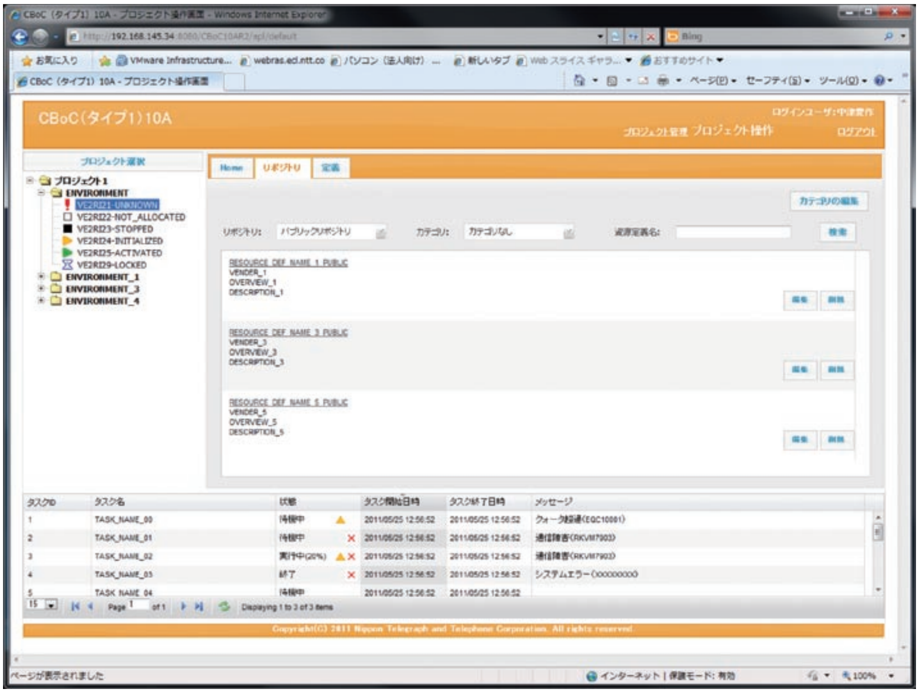
4.1 Application to the R&D cloud

Starting in October 2011, we plan to apply the CBoC provisioning infrastructure to the R&D cloud as a development environment lending service components. The configuration for the R&D cloud includes authentication linked with the Open Light Directory Access Protocol (OpenLDAP), project management, role-based authorization, and group quota checking. It uses a web-based console screen (**Fig. 4**) and allows users to build and operate their own development and testing environments using virtual machines and VLANs.

Prior to the R&D cloud, NTT was operating a development environment lending service manually. Experience with it showed that the actual work time required was approximately 5.5 hours, from applying to use the system to actually beginning to use it, and that other work was also required, such as 8.5 hours to create a virtual machine template. The CBoC provisioning infrastructure should reduce this work significantly. Since these figures are based on actual

Table 1. CBoC provisioning-infrastructure external interfaces.

Interface	Application/objective
Resource SPI	A new resource driver should conform to this interface when being created and integrated.
Resource allocation logic SPI	Modules conforming to this interface can control resource allocation (allocating virtual machines to physical machines etc.).
Authentication interface	When the authentication method used for accessing the provisioning infrastructure is changed, the authentication module must satisfy this interface.
Authorization SPI	When a change is made to the method by which operators authorize operations, an authorization module implementing this interface must be created and integrated. The authorization module checks who, in what role, is authorized to perform what actions.
Provisioning API	Basic API provided by the CBoC provisioning infrastructure. Allows creation and deletion of resources, querying and changing of resource states, and other actions.
Extended provisioning API	Provides the provisioning API as a web service. Bundles primitive API requests according to application requirements.
Provisioning event listener API	To enable the introduction of processing before or after the execution of a provisioning API routine, an event listener can be implemented using this interface.
Life-cycle event listener API	To enable the introduction of any additional processing when the state of a resource changes, an event listener can be implemented using this interface.



- プロジェクト1
  - ENVIRONMENT
    - VE2RI21-UNKNOWN
    - VE2RI22-NOT_ALLOCATED
    - VE2RI23-STOPPED
    - VE2RI24-INITIALIZED
    - VE2RI25-ACTIVATED
    - VE2RI29-LOCKED
  - ENVIRONMENT_1
  - ENVIRONMENT_3
  - ENVIRONMENT_4

RESOURCE_DEF_NAME_1_PUBLIC
VENDER_1
OVERVIEW_1
DESCRIPTION_1

RESOURCE_DEF_NAME_3_PUBLIC
VENDER_3
OVERVIEW_3
DESCRIPTION_3

RESOURCE_DEF_NAME_5_PUBLIC
VENDER_5
OVERVIEW_5
DESCRIPTION_5

タスクID	タスク名	状態	タスク開始日時	タスク終了日時	メッセージ
1	TASK_NAME_00	待機中	2011/05/25 12:56:52	2011/05/25 12:56:52	クォータ超過(EQC10001)
2	TASK_NAME_01	待機中	2011/05/25 12:56:52	2011/05/25 12:56:52	通信障害(RKVM7903)
3	TASK_NAME_02	実行中(20%)	2011/05/25 12:56:52	2011/05/25 12:56:52	通信障害(RKVM7903)
4	TASK_NAME_03	終了	2011/05/25 12:56:52	2011/05/25 12:56:52	システムエラー(xxxxxxxx)
5	TASK_NAME_04	待機中	2011/05/25 12:56:52	2011/05/25 12:56:52	

Page 1 of 1 Displaying 1 to 3 of 3 items

Copyright(C) 2011 Nippon Telegraph and Telephone Corporation. All rights reserved.

ページが表示されました インターネット | 保護モード: 有効 100%

Fig. 4. Web console for the R&D cloud.

working time, the waiting time for users will be reduced even more, which will increase user satisfaction.

4.2 Evaluating ease with which resource drivers for products on the market can be created

As part of the process of evaluating virtual network products in NTT, we have created a resource driver that handles the product controller and have linked it with a KVM driver for evaluation. The evaluator was not particularly familiar with virtual network products, but was still able to create the resource driver in about one person-month, including testing and integrating the virtual network product into the self-provisioning environment.

Considering that this resource driver was produced internally by laboratory staff, without the quality control of a commercial product, that there is currently no development guide, and that there are many other variable factors, this evaluation result shows that it is easy to create resource drivers. It also shows that it will be easy to quickly integrate various new computing resources provided in the future into the system.

5. Future directions

(1) Linking with capacity planning

There is no definitive method for capacity planning yet, but in the future this area will expand beyond simply the so-called performance optimization to include aspects like controlling power consumption in response to the demands of society. Linking this with the provisioning function is also important, and we intend to pursue this R&D in a unified way.

(2) Enhancing network functions

Network system functions must be enhanced in order to provide the type of cloud infrastructure expected from a network carrier. This requires us to advance network driver development, centered on the laboratory's core technologies, and expand into the area of performance optimization, including that of the network.

(3) Inter-cloud linking

Cloud linking from the perspective of distributing resources is also important considering the effects of disasters. When clouds are linked, the management of large amounts of widely distributed data and cooperative behavior over a wide-area network are very important. Such wide-ranging cooperation is an important area, even in the laboratory. In the future, we will continue to work on these elemental technologies.

(4) Commodity function initiatives

Virtual machines and VLANs, which were the initial cloud resources, are now familiar, and this area will continue to become commoditized in the future. In commodity fields, influential open source software such as OpenStack has appeared, and this may contribute to a unification of operational interfaces in the future. For the CBoC provisioning infrastructure, resource drivers have been developed according to standardized interfaces, so we leave integration as commodity resources to organizations like OpenStack and focus our R&D efforts on differentiating functions.

References

[1] P. Mell and T. Grance, "The NIST Definition of Cloud Computing," NIST, Vol. 53, No. 6, p. 50, 2009.

[2] A. Abe, T. Ochi, A. Shirahase, and F. Kumada, "NTT Group Initiatives for Achieving Societal Cloud Infrastructure," NTT Technical Review, Vol. 9, No. 12, 2011.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201112fa1.html

Kenichi Sato

Senior Research Engineer, Supervisor, Cloud Computing SE Project, NTT Information Sharing Platform Laboratories.

He received the B.E. and M.E. degrees in precision mechanical engineering from the University of Tokyo in 1991 and 1993, respectively. He joined NTT Network Information Systems Laboratories in 1993. He studied intelligent agent communication including agent communication platforms and agent application development platforms. After that, he joined an electronic payment system trial project. From 2002 to 2005, he developed and operated an ASP (application service provider) service (Business goo), targeted at small-office home-office businesses. In 2005, he returned to the Information Sharing Platform Laboratories and engaged in the development of a high reliability transaction processing monitor on Linux. Since 2008, he has been studying cloud computing systems, including cloud distributed data management, cloud application frameworks, and cloud operation platforms.

Hideki Hayashi

Senior Research Engineer, Supervisor, Development Project Leader, Cloud Computing SE Project, NTT Information Sharing Platform Laboratories.

He received the B.E. and M.E. degrees in electrical engineering from Tokyo Institute of Technology in 1987 and 1989, respectively. He joined NTT Telecommunication Networks Laboratory in 1989 and studied network control technology. He developed ATM network integration manager systems, security gateway systems, and authentication authorization accounting systems. He is currently studying CBoC Type 1. He is a member of the Institute of Electronics, Information and Communication Engineers.

Ken Ojiri

Research Engineer, Cloud Computing SE Project, NTT Information Sharing Platform Laboratories.

He received the B.E. and M.E. degrees in communication engineering from Osaka University in 1994 and 1996, respectively. He joined NTT Network Service Systems Laboratories in 1996 and developed intelligent network systems until 2002. From 2002 to 2010, he developed, provided, and operated identity management systems for web-based Internet services in NTT Information Sharing Platform Laboratories and NTT Resonant Inc. He is currently studying and developing CBoC provisioning infrastructure.

Large-scale Distributed Data Processing Platform for Analysis of Big Data

Mitsukazu Washisaka, Eiji Nakamura, Takeshi Takakura, Satoru Yoshida, and Seiji Tomita†

Abstract

Cloud technology can scale horizontally (scale out) through the addition of servers to handle more data and enables the analysis of large volumes of data (known as big data). This has led to innovation in the form of new services that create added value such as providing recommendations on the basis of attributes extracted from the big data. In this article, we describe technology and services implemented with large-scale data processing, the development of a large-scale distributed data processing platform, and technology fostered by the program design and system level testing/verification.

1. Introduction

One application area for cloud computing is large-scale data processing. The analysis of voluminous data within a realistic time requires abundant computer resources, including central processing units, disk space for data storage, and network bandwidth. A lot of researchers have been interested in scale-out (horizontal scaling) technology and have started research and development (R&D) activities as a way to deal with large-volume data analysis by adding blade/rack servers to increase processing speed and expand data capacity. NTT Information Sharing Platform Laboratories has also been developing a large-scale distributed data processing platform called CBoC Type 2 (CBoC: Common IT Bases over Cloud Computing; IT: information technology) [1].

In this article, we first describe services that use large-scale data processing and then introduce technology for achieving the high degree of availability required of a large-scale distributed data processing platform as well as testing technology for evaluating applicability.

2. Advanced services through large-scale data processing

It is difficult to manage the data generated in the web, IT systems, etc. (e.g., transaction logs, sensor logs, and life logs) and other data that continues to increase explosively in volume. Analysis of such voluminous data (known as big data) in a conventional manner becomes exponentially costly even when the data is collected by the system, so the data has been either stored wastefully or discarded. The advent of scale-out technology, however, has reduced the cost of constructing systems for processing large-scale data, and new advanced services such as personalization based on analytical results are now possible.

Large-scale data processing enables the use of diverse types of big data in a cloud environment in order to create mash-up[*1] services, as shown in **Fig. 1.** A large-scale distributed data processing platform collects and stores the big data produced by IT systems or the Internet. By analyzing such large volumes of data, one can acquire new knowledge and expertise

† NTT Information Sharing Platform Laboratories
Musashino-shi, 180-8585 Japan

[*1] Mash-up: A new service constructed by combining multiple application programming interfaces.

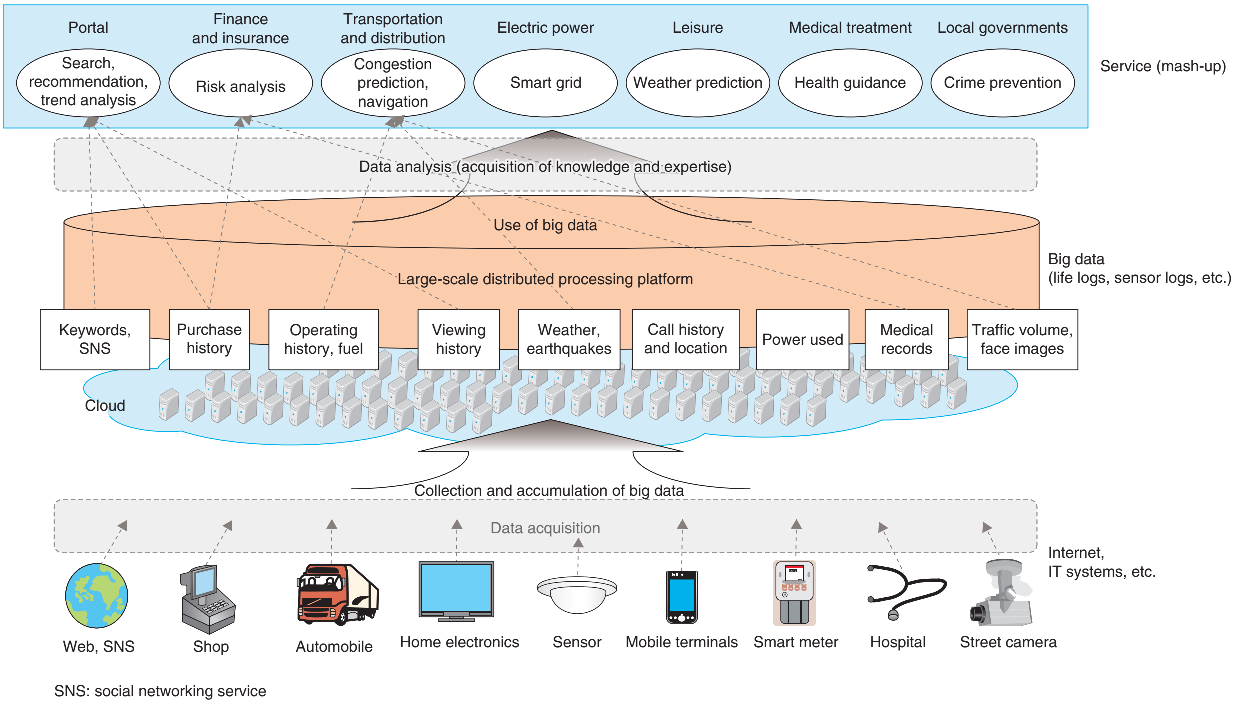


Fig. 1. Large-scale distributed data processing platform and application to services.

and create new mash-up services. A large-scale distributed data processing platform is expected to serve as a platform for creating knowledge on which to base advanced services for customers.

3. Development of CBoC Type 2

The conventional approach to the management of various types of data is to use a relational database management system (RDBMS)*2, but for big data, a technology known as NoSQL (not only SQL (structured query language))*3 is a cost-effective approach. NoSQL technology implements scale-out by relaxing the guarantee of data consistency that is essential for transaction processing in a RDBMS. Thus, while an RDBMS is suited to the analysis of structured data, the NoSQL approach, which is based on BASE (basically available, soft states, eventual consistency), is suited to the processing of big data that is less structured, such as natural language data.

Typical examples in this field include the service platforms provided by Google and Amazon [2], [3] and the open-source software Hadoop [4]. Hadoop users are increasing in number, and systems on the scale of thousands of servers have been reported to be in operation. However, some problems regarding introduction to mainstream systems that require continuous operation remain; one example is the inability to switch servers online when a management server fails. We therefore took up the challenge of developing CBoC Type 2 to improve the reliability, operability, and maintainability of a large-scale distributed data processing platform.

4. Improving fault tolerance in large-scale distributed data processing platforms

CBoC Type 2 comprises three distributed processing subsystems: a distributed file system for storing big data on many blade/rack servers in a distributed manner, a distributed table subsystem for managing big data as structured data, and a distributed lock subsystem, which provides a basic function that allows a high degree of settlement in these distributed systems (**Fig. 2**).

In the development of CBoC Type 2, particular effort was made to improve fault tolerance, which is

*2 RDBMS: A database management system that features data representation in the form of two-dimensional tables.

*3 NoSQL: Processing requests submitted to an RDBMS are often written in SQL, but NoSQL refers to a language designed as *not being an RDBMS*.

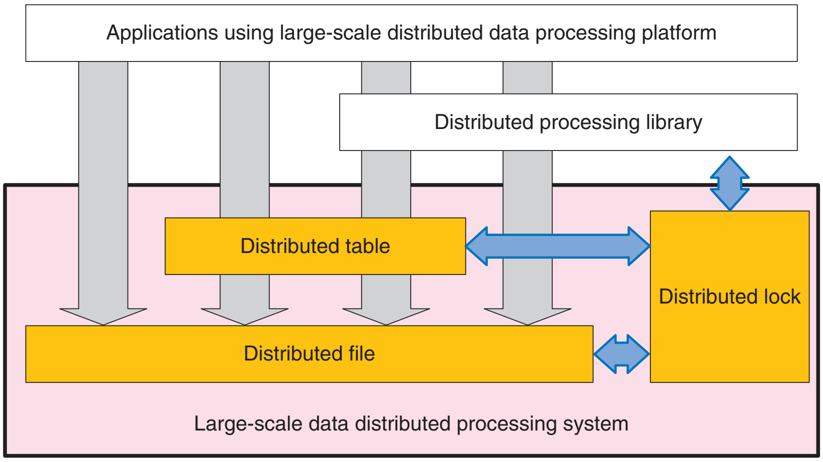

Applications using large-scale distributed data processing platform
Distributed processing library
Distributed table
Distributed lock
Distributed file
Large-scale data distributed processing system

Fig. 2. CBoC Type 2 software structure.

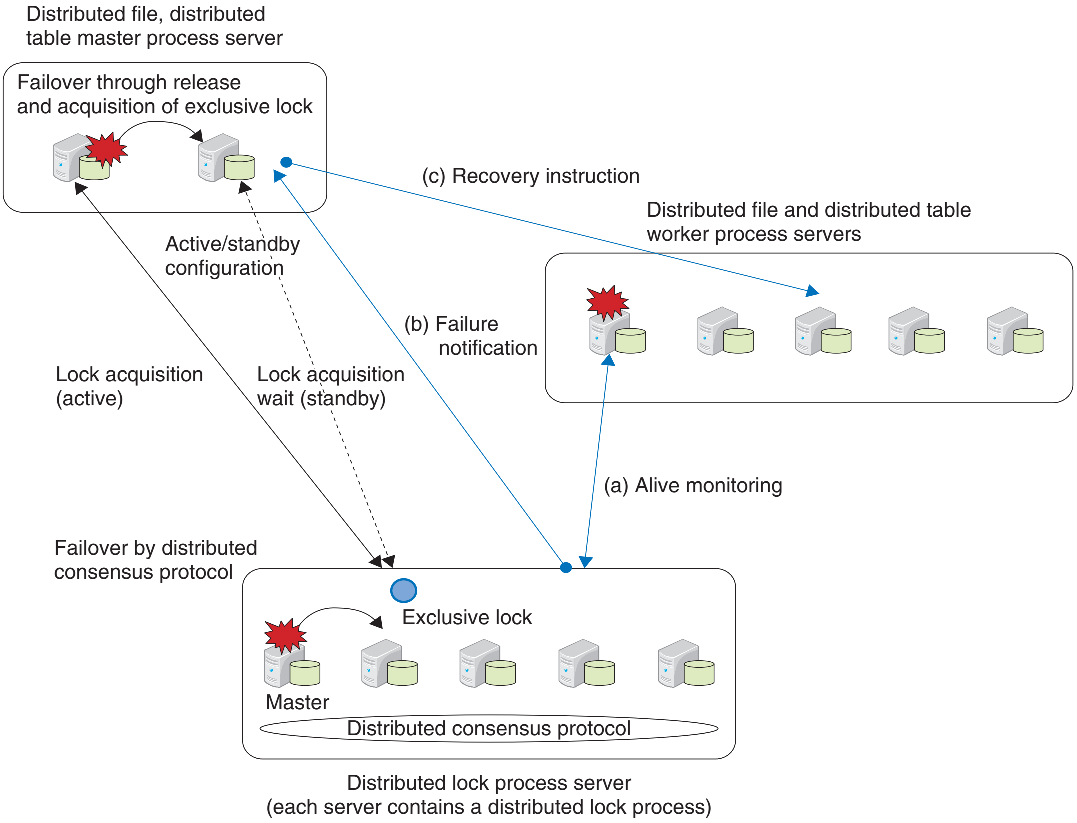

Distributed file, distributed table master process server
Failover through release and acquisition of exclusive lock
(c) Recovery instruction
Distributed file and distributed table worker process servers
Active/standby configuration
(b) Failure notification
Lock acquisition (active)
Lock acquisition wait (standby)
(a) Alive monitoring
Failover by distributed consensus protocol
Exclusive lock
Master
Distributed consensus protocol
Distributed lock process server (each server contains a distributed lock process)

Fig. 3. CBoC Type 2 failure recovery processing.

important for managing scale out. In systems that use many blade/rack servers, the probability of overall system failure is high, even if the failure rate of individual servers is low. Thus, ensuring hardware failure tolerance is a major problem. In CBoC Type 2, the distributed lock function provides the basis for fault tolerance in the distributed file and distributed table subsystems (**Fig. 3**).

4.1 Fault tolerance of distributed locks

A distributed lock works to set one of the five processess on different servers to be the master process. The five server processes communicate via a distributed consensus protocol to form a consensus among backup processes on servers [5]. If the master process server fails, a new master is selected from the remaining backup process servers by a distributed consensus protocol, and the new master takes over the previous master's processing (failover). By doing so, it maintains the distributed lock process with respect to the whole system [1].

4.2 Fault tolerance of distributed files and distributed tables

Distributed files and distributed tables consist of many worker processes that process requests from applications and two master processes that control the worker processes. The active master and standby master processes are distinguished by using the exclusive lock[*4] function of the distributed lock. The master that succeeds in acquiring an exclusive lock becomes the active master and the master that failed to acquire the lock becomes the standby master. A distributed lock monitors the life and death of each master process; if the active master fails, the exclusive lock is released. The standby master can then acquire the exclusive lock and become the active master (failover).

The distributed lock also monitors the life and death of worker processes ((a) in Fig. 3). If a worker process fails, the failure is reported to the active master (b). The active master issues an instruction to another running worker process to recover the data (c) that was being managed by the failed worker process and the running worker process takes up the processing that was being performed by the failed worker process. To prevent data loss due to distributed file failure, the data is made redundant and the same data is managed by multiple worker processes.

4.3 Fault tolerance considering the network

The network configuration of the servers also has a strong effect on fault tolerance. CBoC Type 2 uses a tree network configuration in which multiple edge switches are subordinate to a core switch and each edge switch connects to multiple servers. The master processes of the distributed file and distributed table system are positioned under different edge switches and the redundant data of distributed files is managed by worker processes that are under different edge switches.

In that way, CBoC Type 2 implements fault tolerance such that the whole system does not go down when a single unit of hardware (such as a server or switch) fails, even if an edge switch failure results in multiple servers being removed from the system at the same time.

5. System testing in large-scale distributed systems

Large-scale distributed systems designed for the analysis of big data may comprise from several tens to several hundreds of servers, or even thousands in some cases, depending on the scale of the data to be processed and the nature of the processing. The number of variations of state transitions in such distributed systems is huge, so system testing is correspondingly more important. Therefore, system testing in CBoC Type 2 is designed to allow the construction of a testing environment that involves thousands of servers and implementation of testing that assumes actual service use cases.

The difficulties faced by testing in the construction of a large-scale distributed data processing platform include 1) automation of testing environment construction, 2) faster registration of big data, and 3) more efficient confirmation of test results. CBoC Type 2 deals with those problems in the following ways.

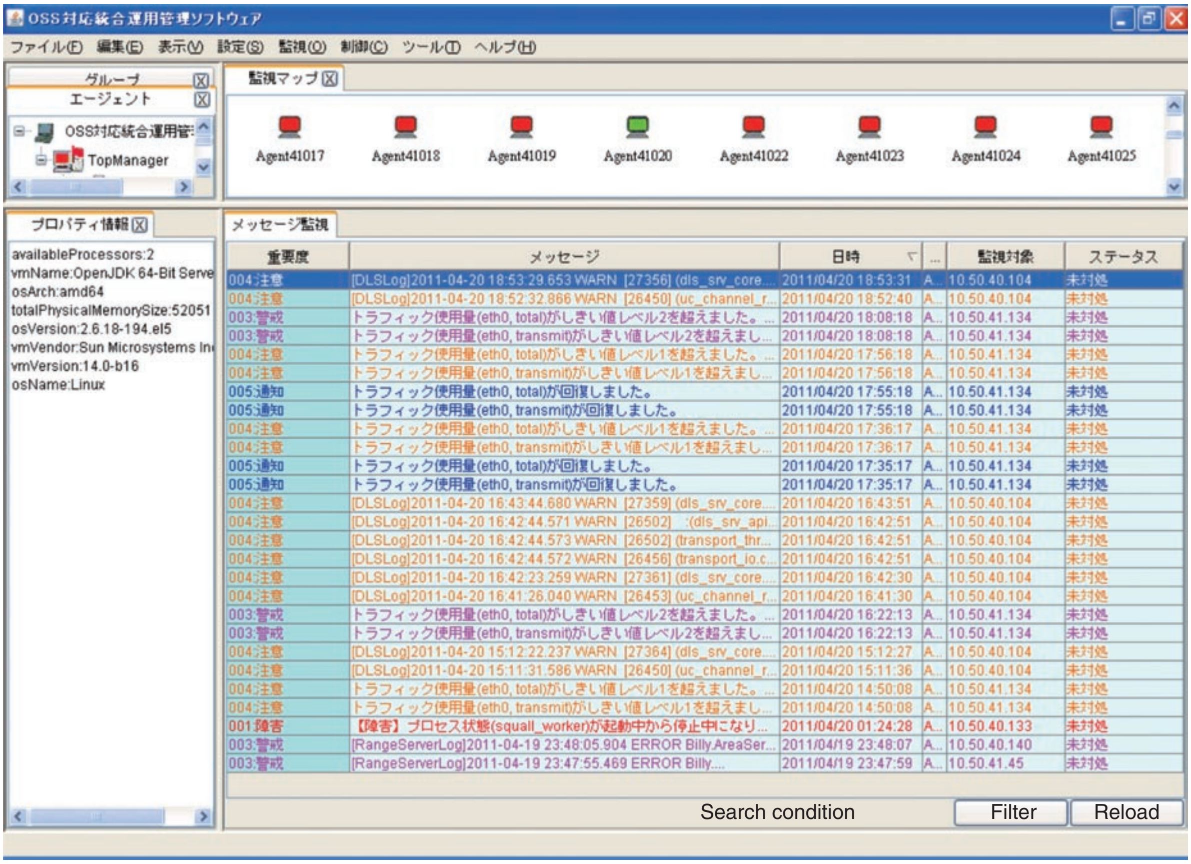
5.1 Automation of testing environment construction

The difficulty of constructing testing environments can probably be imagined by simply considering the work involved in installing software on several hundred servers. System management automation tools can be used effectively in the construction of such an environment. Environment construction and maintenance can be made more efficient by using system management automation tools such as Puppet [6] for unified management of installed operating systems, middleware, application programs and various settings, as well as for the management of the installation process.

5.2 Faster entry of big data

To check for stable system operation, we developed a data entry tool that can construct various data

[*4] Exclusive lock: A mechanism that limits the number of processes that can modify data to one to avoid problems caused by modification of the same data by multiple processes.



```
availableProcessors:2
vmName:OpenJDK 64-Bit Serve
osArch:amd64
totalPhysicalMemorySize:52051
osVersion:2.6.18-194.el5
vmVendor:Sun Microsystems In
vmVersion:14.0-b16
osName:Linux
```

重要度	メッセージ	日時	...	監視対象	ステータス
004:注意	[DLSLog]2011-04-20 18:53:29.653 WARN [27356] (dls_srv_core....	2011/04/20 18:53:31	A...	10.50.40.104	未対処
004:注意	[DLSLog]2011-04-20 18:52:32.866 WARN [26450] (uc_channel_r...	2011/04/20 18:52:40	A...	10.50.40.104	未対処
003:警戒	トラフィック使用量(eth0, total)がしきい値レベル2を超えました。 ...	2011/04/20 18:08:18	A...	10.50.41.134	未対処
003:警戒	トラフィック使用量(eth0, transmit)がしきい値レベル2を超えまし...	2011/04/20 18:08:18	A...	10.50.41.134	未対処
004:注意	トラフィック使用量(eth0, total)がしきい値レベル1を超えました。 ...	2011/04/20 17:56:18	A...	10.50.41.134	未対処
004:注意	トラフィック使用量(eth0, transmit)がしきい値レベル1を超えまし...	2011/04/20 17:56:18	A...	10.50.41.134	未対処
005:通知	トラフィック使用量(eth0, total)が回復しました。	2011/04/20 17:55:18	A...	10.50.41.134	未対処
005:通知	トラフィック使用量(eth0, transmit)が回復しました。	2011/04/20 17:55:18	A...	10.50.41.134	未対処
004:注意	トラフィック使用量(eth0, total)がしきい値レベル1を超えました。 ...	2011/04/20 17:36:17	A...	10.50.41.134	未対処
004:注意	トラフィック使用量(eth0, transmit)がしきい値レベル1を超えまし...	2011/04/20 17:36:17	A...	10.50.41.134	未対処
005:通知	トラフィック使用量(eth0, total)が回復しました。	2011/04/20 17:35:17	A...	10.50.41.134	未対処
005:通知	トラフィック使用量(eth0, transmit)が回復しました。	2011/04/20 17:35:17	A...	10.50.41.134	未対処
004:注意	[DLSLog]2011-04-20 16:43:44.680 WARN [27359] (dls_srv_core....	2011/04/20 16:43:51	A...	10.50.40.104	未対処
004:注意	[DLSLog]2011-04-20 16:42:44.571 WARN [26502] :(dls_srv_api...	2011/04/20 16:42:51	A...	10.50.40.104	未対処
004:注意	[DLSLog]2011-04-20 16:42:44.573 WARN [26502] (transport_thr...	2011/04/20 16:42:51	A...	10.50.40.104	未対処
004:注意	[DLSLog]2011-04-20 16:42:44.572 WARN [26456] (transport_io.c...	2011/04/20 16:42:51	A...	10.50.40.104	未対処
004:注意	[DLSLog]2011-04-20 16:42:23.259 WARN [27361] (dls_srv_core....	2011/04/20 16:42:30	A...	10.50.40.104	未対処
004:注意	[DLSLog]2011-04-20 16:41:26.040 WARN [26453] (uc_channel_r...	2011/04/20 16:41:30	A...	10.50.40.104	未対処
003:警戒	トラフィック使用量(eth0, total)がしきい値レベル2を超えました。 ...	2011/04/20 16:22:13	A...	10.50.41.134	未対処
003:警戒	トラフィック使用量(eth0, transmit)がしきい値レベル2を超えまし...	2011/04/20 16:22:13	A...	10.50.41.134	未対処
004:注意	[DLSLog]2011-04-20 15:12:22.237 WARN [27364] (dls_srv_core....	2011/04/20 15:12:27	A...	10.50.40.104	未対処
004:注意	[DLSLog]2011-04-20 15:11:31.586 WARN [26450] (uc_channel_r...	2011/04/20 15:11:36	A...	10.50.40.104	未対処
004:注意	トラフィック使用量(eth0, total)がしきい値レベル1を超えました。 ...	2011/04/20 14:50:08	A...	10.50.41.134	未対処
004:注意	トラフィック使用量(eth0, transmit)がしきい値レベル1を超えまし...	2011/04/20 14:50:08	A...	10.50.41.134	未対処
001:障害	【障害】プロセス状態(squall_worker)が起動中から停止中になり...	2011/04/20 01:24:28	A...	10.50.40.133	未対処
003:警戒	[RangeServerLog]2011-04-19 23:48:05.904 ERROR Billy.AreaSer...	2011/04/19 23:48:07	A...	10.50.40.140	未対処
003:警戒	[RangeServerLog]2011-04-19 23:47:55.469 ERROR Billy....	2011/04/19 23:47:59	A...	10.50.41.45	未対処

Fig. 4. Screenshot from the integrated operation and management software (Crane).

storage states in a short time; this function is essential for testing when a large amount of data has been entered. The tool incorporates functions for restarting from a state midway through data entry and for automatically adding previously entered data, etc.; these functions greatly reduce the preparation time required for testing.

5.3 More efficient validation of test results

The items involved in the validation of test results include the input and output results as seen by the application, the data entry state, and whether or not an error occurred. For large-scale testing environments in particular, realtime observation and visualization of error occurrence and server process states made possible by the monitoring of system operation and error log and other such records is important. For this purpose, NTT's Crane technology [7] or other similar integrated operation management software is used for efficient confirmation of test results. The operation management screen of Crane is shown in **Fig. 4**.

By proceeding with testing while using related technology and tools in addition to the three subsystems for distributed files, distributed tables, and distributed locks in this way, we are building up the expertise in system operation needed for large-scale environments in order to bring CBoC Type 2 to a highly practical level of development.

6. Future development

In developing CBoC Type 2, we encountered many difficulties in running programs in a large-scale environment and we learned much during the development and testing phases. By gaining use experience for various kinds of assumed services through the application of CBoC Type 2 to NTT R&D Cloud, the cloud environment for NTT's R&D centers, we will identify the common functions and performance requirements for the platform and increase applicability to specific needs to develop CBoC Type 2 into a large-scale distributed data processing platform that has high reliability, operability, and maintainability.

References

[1] T. Takakura, K. Sora, Y. Amagai, M. Washisaka, and S. Tomita, "Implementing Large-scale Distributed Processing Systems with CBoC," NTT Technical Journal, Vol. 21, No. 9, pp. 80–83, 2009 (in Japanese).

[2] Google App Engine. http://code.google.com/intl/en/appengine/docs/whatisgoogleappengine.html

[3] Amazon web services. http://aws.amazon.com/.

[4] Hadoop. http://hadoop.apache.org/

[5] L. Lamport, "Paxos Made Simple," ACM SIGACT News, Vol. 32, No. 4, pp. 18–25, 2001.

[6] Puppet. http://www.puppetlabs.com/

[7] NTT Open Source Software Center, Crane (in Japanese). https://www.oss.ecl.ntt.co.jp/ossc/oss/r_crane.html

Mitsukazu Washisaka

Senior Research Engineer, Supervisor, Distributed Data Processing Platform SE Project, NTT Information Sharing Platform Laboratories.

He received the B.S. and M.S. degrees in information and computer science engineering from Osaka University in 1985 and 1987, respectively. He joined NTT Basic Research Laboratories in 1987. He has been engaged in R&D of wide-area IP networks and their applications.

Eiji Nakamura

Senior Research Engineer, Supervisor, Distributed Data Processing Platform SE Project, NTT Information Sharing Platform Laboratories.

He received the B.E. and M.E. degrees in nuclear engineering from Hokkaido University in 1986 and 1988, respectively. He joined NTT in 1988. He has been engaged in R&D of facsimile communication systems and operation systems for home gateway devices. He is currently studying big data processing and management systems.

Takeshi Takakura

Senior Research Engineer, Supervisor, Distributed Data Processing Platform SE Project, NTT Information Sharing Platform Laboratories.

He received the B.E. and M.E. degrees in material physics engineering from Osaka University in 1990 and 1992, respectively. He joined NTT in 1992 and engaged in R&D in the Network Information Systems Laboratories, where he studied multimedia database systems and information processing systems. He received the 1997 Best Paper Award for a Young Researcher of IPSJ (Information Processing Society of Japan) National Convention. He is currently studying big data processing and management systems.

Satoru Yoshida

Senior Research Engineer, Supervisor, Distributed Data Processing Platform SE Project, NTT Information Sharing Platform Laboratories.

He received the B.S. and M.S. degrees in condensed matter physics engineering from Tokyo Institute of Technology in 1987 and 1989, respectively. He joined NTT Applied Electronics Laboratories in 1989. He is currently engaged in R&D of cloud computing systems. He is a member of the Institute of Electronics, Information and Communication Engineers.

Seiji Tomita

Senior Research Engineer, Supervisor, Distributed Data Processing Platform SE Project, NTT Information Sharing Platform Laboratories.

He received the B.S. and M.S. degrees in electronics from Kyushu University, Fukuoka, in 1983 and 1985, respectively. He joined the Yokosuka Electrical Communications Laboratories of Nippon Telegraph and Telephone Public Corporation (now NTT) in 1985. He has been engaged in R&D of system software in computer systems such as operating systems, communication software, transaction monitors, and database management systems. His current interest is big data processing and management systems.

Network Virtualization Technology for Cloud Services

Hideo Kitazume[†]*, Takaaki Koyama, Toshiharu Kishi, and Tomoko Inoue*

Abstract

The network virtualization technology needed to effectively and efficiently construct and operate a cloud is already in place. In this article, we introduce the trends in the latest network virtualization technology for the cloud environment and its fields of application.

1. Introduction

In recent years, the development of server virtualization technology has led to changes in the cloud services environment, such as more efficient usage of physical servers (high aggregation), sharing of hardware resources such as server and network (multitenacy), and the need for practical migration. For networks within datacenters, attention has been drawn to problems such as the explosive increase in the number of medial access control (MAC) addresses and virtual local area networks (VLANs), the construction of layer 2 (L2) networks across offices, and network migration.

Technology for solving these problems has taken two major directions: architectures that extend existing technology and network equipment commoditization through virtualization. The former uses fast, high-capacity L2 switches ranging from 40GbE (40-Gbit/s Ethernet) ones to 100GbE (100-Gbit/s Ethernet) ones to make flat connections over multiple switches in one hop; these switches are operated and managed as one very large logical L2 switch. Although this approach is expected to be widely used in large, next-generation datacenter networks, there are problems such as strong dependence on the switch vendor, insufficient interworking with network equipment other than switches (e.g., firewalls and load balancers), and the same high construction cost as in the past. The latter approach, on the other hand, establishes a logical network independently of the physical network by logically integrating functions such as those of network equipment other than switches into a standard switch called an OpenFlow switch. As a result, a carrier can expect lower construction costs for datacenter networks using commodity network equipment. OpenFlow was initiated by the Open Network Foundation (ONF), a promotional organization that has been active in moving the technology toward a practical stage. ONF has focused on the flexibility of network programmability (reconfiguration of the network in connection with applications), which is considered to be a powerful network virtualization technology for the cloud environment.

In this article, we explain the need for network virtualization as well as the technology itself and its application areas.

2. Need for network virtualization

As described above, multitenancy requires the ability to provide an isolated network for each cloud services user, together with network flexibility and agile construction and configuration changes for high aggregation and migration needs. There are, however, difficult problems in satisfying those requirements with existing VLANs and products.

(1) Difficulty of network design and management in datacenters

In datacenters, the tagged VLAN is generally used to isolate each user's network, but there are two

[†] NTT Information Sharing Platform Laboratories
Musashino-shi, 180-8585 Japan

problems with this approach. One is the limited capacity of the tagged VLAN. This technique attaches one of 4094 VLAN identifiers (VLAN-IDs) to packets, so the maximum number of isolatable networks is 4094. A VLAN-ID must be unique over the entire datacenter network, so it is impossible to accommodate more than 4094 users at the same time. The other problem is the use of proprietary specifications by the vendors of the products used in datacenters. For the products currently used in datacenters, most vendors use proprietary specifications in both network design and setup. Network design, for example, may be premised on a virtual chassis* being used for switch products, in which multiple switches are seen as a single switch, and on the tagged VLAN being used for communication within server products. For setup methods, individual vendors have proprietary control protocol specifications etc. It is thus necessary to assign the VLAN-ID to match the vendor specifications, and the control protocol used must also match the vendor specifications. Furthermore, as the numbers of servers and switches in datacenters increase to more than hundreds of units, the use of products from multiple vendors to avoid vendor lock-in and reduce costs increases the difficulty of datacenter network design and management.

(2) Agile automatic changes in network configuration in cooperation with migration

For cloud services, virtual machine migration among multiple datacenters in times measured in hours is necessary in order to take advantage of nighttime electricity rates and to respond to disasters. Specifically, services can be migrated without interruption on a virtual machine that uses TCP (transmission control protocol), UDP (user datagram protocol), or other protocols, and the network configuration can be changed in cooperation with the migration within a very short time.

As a step toward solving the above two problems, a new standard network virtualization technology that does not use VLANs and that eliminates vendor dependence is needed for cloud services.

3. OpenFlow and ONF

OpenFlow originated as technology for an academic network at Stanford University in 2008, but it is now being studied by the OpenFlow Switch Consortium as a network control technology.

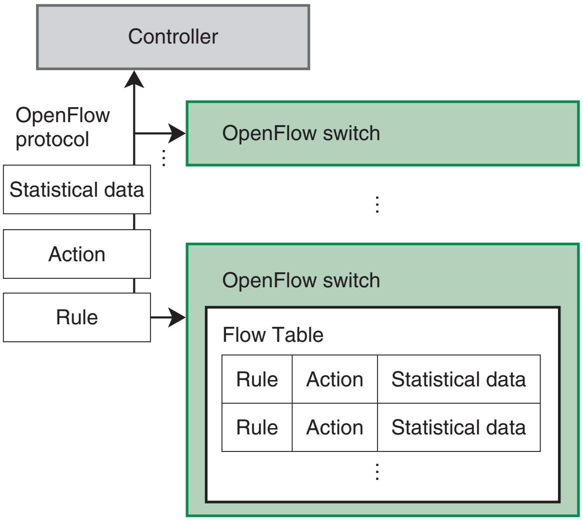


Fig. 1. Switch control in OpenFlow.

3.1 Controller-data separation

The basic concept of OpenFlow is that a controller performs central control by distributing programs to switches that conform to the OpenFlow specifications. Each switch then operates according to the program (**Fig. 1**). Unlike the conventional Internet, in which the various types of network equipment exchange path information and select paths autonomously, the controller performs all control centrally, and each switch operates according to the instructions given to it. This scheme is referred to as controller-data separation.

3.2 Control mechanism

OpenFlow control is specified as combinations of rules and actions. Rules identify the packets to be processed. It is thus possible to specify the evaluation conditions for L1–L4 header contents for packets whose TCP port number is 80, for example. Actions specify operations to be performed on the packets that match the rules. Specifically, it is possible to rewrite the header so that the packets are transferred to different ports or to specify that they are to be discarded, etc. For example, an action can specify that packets that arrive at a particular port number are to be discarded. Another way to regard this is that the controller can change a switch into a router, firewall, or load balancer as needed by sending a simple program to it. The flexibility of being able to control anything by programs is one reason that OpenFlow has been attracting attention.

* Chassis: In this context, chassis covers case, cabinet, and chassis.

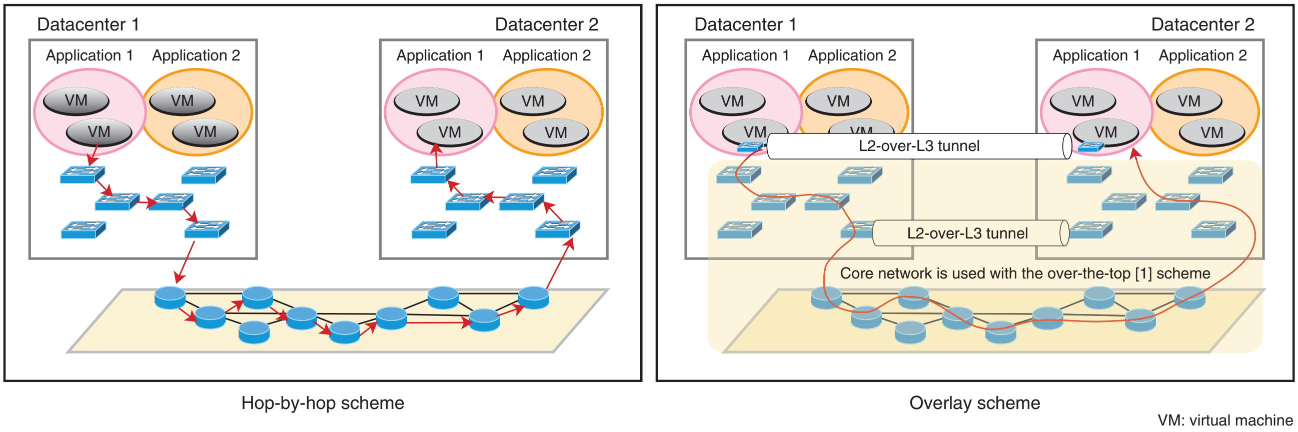


Fig. 2. Hop-by-hop and overlay schemes.

3.3 ONF

ONF is an organization of member companies such as Google, Facebook, and Yahoo that began promoting OpenFlow in March, 2011. NTT is also a participant. An interesting feature of ONF is that the board members are representatives of companies that operate large-scale datacenters, which is to say they are network user companies rather than network vendors. With the appearance of ONF, OpenFlow has taken on a commercial aspect in addition to its previous image of an academic system and it is now attracting much attention.

4. Use of OpenFlow

Considering routing, there are two schemes for using OpenFlow: hop-by-hop and overlay (**Fig. 2**).

(1) Hop-by-hop

In the hop-by-hop routing scheme, the controller knows all of the switches and designs paths service-by-service. Each switch operates according to instructions so as to repeatedly forward packets in a relay scheme that delivers the packets to their final destination. Although this scheme makes free use of the advantages of OpenFlow, each switch must hold all of the path information, so scalability may be a problem. It is suitable for the construction of small-scale networks, but application to large-scale networks requires measures against path congestion etc.

(2) Overlay scheme

In the overlay scheme, the controller does not control all of the paths, but uses the tunneling technique described later to control the communicating end points, a practice that is referred to as edge networking. With this scheme, the controller and the various switches need to know only the source and destination of the communication; the path is handled by the conventional routing mechanism.

This approach enables the amount of routing data to be managed to be kept down to a realistic level, even for a large-scale network. Early introduction to actual services is expected to be more feasible for the overlay scheme than for the hop-by-hop scheme. Next, we explain the implementation of an overlay virtual network with L2-over-L3 tunneling, which is one kind of overlay scheme.

5. Overlay virtual network

The overlay virtual network is implemented by encapsulating users' L2 frames inside L3 packets to achieve L2-over-L3 tunneling (**Fig. 3**). The three main points are explained below.

(1) L2-over-L3 tunnel

In this technique, an OpenFlow switch within a hypervisor is equipped with an L2-over-L3 tunnel endpoint function and a tunnel is established between two hypervisors. The connection between the on-premises environment (locally operated) and a hypervisor is also established by setting up an OpenFlow switch-based virtual gateway that has a tunnel endpoint function. The use of VLANs is abandoned to avoid VLAN-ID exhaustion and to escape from vendor dependence in VLAN-ID design.

(2) User isolation

User isolation is implemented with a function that assigns a user ID to each user and encapsulates the

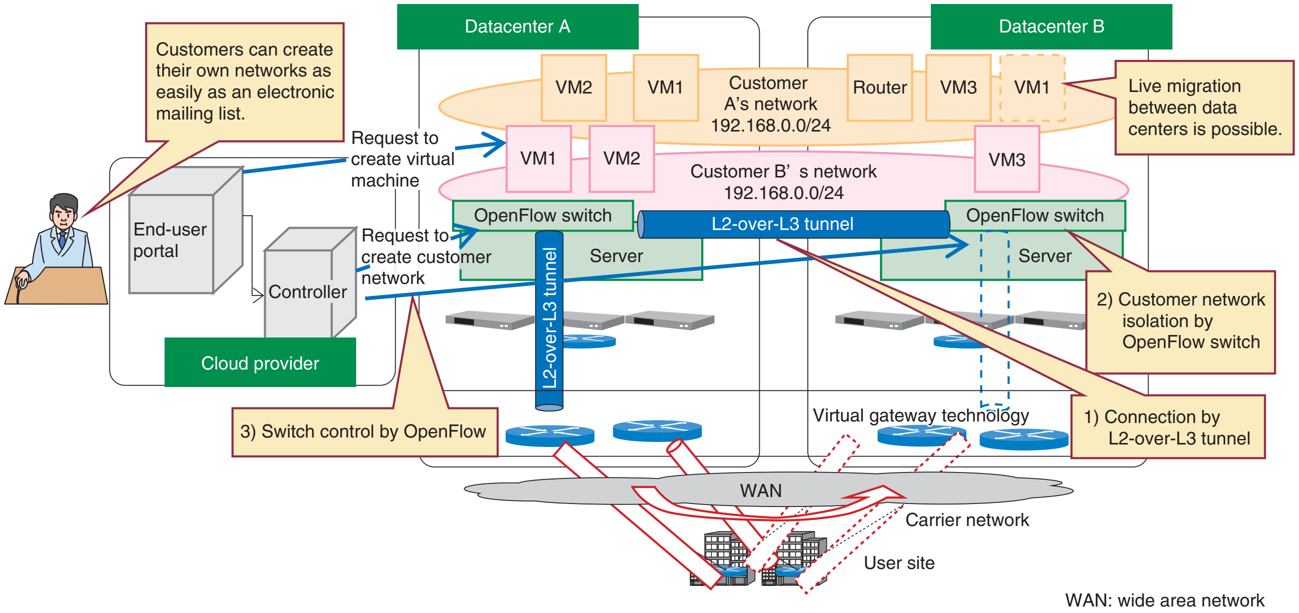

Customers can create their own networks as easily as an electronic mailing list.
Datacenter A
Datacenter B
VM2
VM1
Customer A's network 192.168.0.0/24
Router
VM3
VM1
Live migration between data centers is possible.
Request to create virtual machine
VM1
VM2
Customer B' s network 192.168.0.0/24
VM3
End-user portal
OpenFlow switch
L2-over-L3 tunnel
OpenFlow switch
Request to create customer network
Server
Server
Controller
L2-over-L3 tunnel
2) Customer network isolation by OpenFlow switch
Cloud provider
3) Switch control by OpenFlow
Virtual gateway technology
1) Connection by L2-over-L3 tunnel
WAN
Carrier network
User site
WAN: wide area network

Fig. 3. Overlay virtual network.

data within the tunnel endpoint function of the OpenFlow switch. The switching is then performed by the OpenFlow switch's switching function using both the virtual and physical interfaces and the user ID.

(3) Standard switch control protocol

The use of OpenFlow for the switch control protocol makes it possible to develop a (hardware) controller that can control the products of multiple vendors. That allows a reduction in equipment costs for servers and switches through multivendor sourcing. The development of this hardware also enables reductions in maintenance and operation costs.

6. Application areas

Next, we describe a few fields of application for overlay virtual networks as virtual network technology.

6.1 Disaster recovery and business continuity plans

After the Great East Japan Earthquake on March 11, 2011, disaster recovery and business continuity plans that involve the backup of data used in offices have gained attention. Disaster recovery countermeasures require remote copying of data among distant offices and migration between virtual machines at different locations. The conventional movement of virtual machines involves the connection of special-purpose machines between offices and requires full network setup to be performed at both locations. That took months to accomplish in the past. By contrast, virtual network functions enable end users themselves to perform a live migration of a virtual machine to a remote location in minutes, which enables smooth disaster recovery.

NTT Information Sharing Platform Laboratories has constructed a logical network and remote live migration function by using virtual network control technology software to establish L2-over-L3 tunneling between cloud environments at the NTT Musashino Research and Development Center and the NTT Atsugi Research and Development Center. Evaluation tests have confirmed that smooth disaster recovery measures can be implemented in this manner.

6.2 Power consumption reduction

After the Great East Japan Earthquake in March 2011, which disrupted the electricity supply, attention turned to ways of reducing the power consumption of cloud services. The number of physical servers, and thus power consumption, can be reduced by concentrating the processing achieved by server virtualization using cloud services.

Nevertheless, the situation surrounding virtual machine use is changing over time, so having virtual

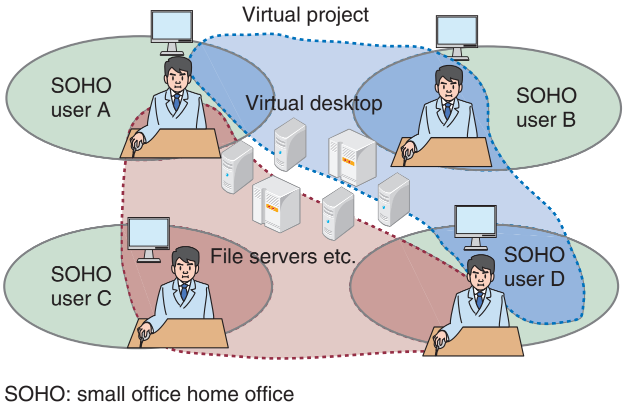


Fig. 4. Virtual desktop service.

servers running on a single physical server is not necessarily the optimum arrangement of physical servers and virtual servers.

Partial movement of a virtual machine among physical servers according to the virtual machine operation state enables concentrated processing that is always optimal to be achieved and power can be conserved by powering down empty physical servers. Furthermore, using the virtual network for migration between remote locations allows flexible operation, such as partial movement of virtual machines to areas that have a large surplus of power.

6.3 Desktop as a service

Desktop as a service (DaaS) puts the user desktop environment in the cloud so that inexpensive personal computers or smart phones can be used for the user environment while maintaining the same high degree of operability provided by a local desktop environment. Furthermore, the provision of new services by using DaaS in combination with virtual networks is being studied.

For example, it would be possible to place the desktop environments of corporate employees of affiliated companies in the cloud and also build logical networks between arbitrary employee desktops on demand. By setting up shared servers and chat servers, etc. on such logical networks, one could easily set up a shared space for projects that involve the employees of multiple organizations or related companies (**Fig. 4**).

In the past, it has been necessary to set up virtual desktops and a VPN for each project, and the end users had to access the virtual desktop of each particular project. With the combination of DaaS and virtual networks, on the other hand, the virtual desktops can be collected together for each user, and the end users only need to switch among the virtual desktops of the projects in which they are participating on demand.

7. Concluding remarks

We have introduced the network virtualization technology needed for the cloud environment. NTT Information Sharing Platform Laboratories is working toward the early implementation of overlay virtual network technology and its incorporation into CBoC (Common IT Bases over Cloud Computing) Type 1. In future work, we plan to investigate interworking between network virtualization technology within datacenters and VPN services in broadband networks, as well as gateway technology for maintaining quality and service level agreement guarantees.

Reference

[1] Over-the-top.
http://en.wikipedia.org/wiki/Over-the-top_content

Hideo Kitazume

Senior Research Engineer, Supervisor, Network Security Project, NTT Information Sharing Platform Laboratories.

He received the B.E. and M.E. degrees in computer science from Gunma University in 1987 and 1989, respectively. He joined NTT in 1989 and engaged in R&D of an ATM-LAN system, ATM traffic control studies, and the development of a global networking service platform. From 1998 to 2010, he worked in NTT EAST and engaged in the development, design, and operation of IP-VPN services. He is currently working on R&D of virtual networking technologies for cloud systems. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and the Operations Research Society of Japan.

Takaaki Koyama

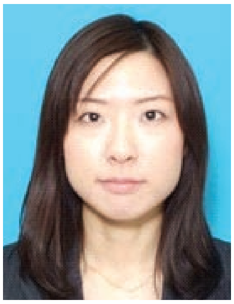
Senior Research Engineer, Network Security Project, NTT Information Sharing Platform Laboratories.

He received the B.E. and M.E. degrees in media and governance from Keio University, Kanagawa, in 1994 and 1996, respectively. He joined NTT Software Laboratory in 1996 and has been studying software CALS. Since 1999, he has been studying GMN-CL, which is a kind of IP-VPN technology and developing some network equipment. Recently, he has been interested in enterprise cloud network systems. He is a member of the Information Processing Society of Japan.

Toshiharu Kishi

Researcher, Secure Networking System Group, Network Security Project, NTT Information Sharing Platform Laboratories.

He received the B.E. and M.E. degrees in medical electronics from Chiba University in 2007 and 2009, respectively. He joined NTT Information Sharing Platform Laboratories in 2009 and worked on threat analysis of web applications. Since 2011, he has been interested in enterprise cloud network system and studying the architecture and construction of virtual networks in a cloud environment. He is a member of IEICE.



Tomoko Inoue

Researcher, Network Security Project, NTT Information Sharing Platform Laboratories.

She received the B.A. degree in literature from Ritsumeikan University, Kyoto, in 2003 and the M.A. degree in informatics from Kyoto University in 2005. She joined NTT WEST in 2005 and moved to NTT Information Sharing Platform Laboratories in 2011. Since 2011, she has been interested in enterprise cloud network systems and is studying the architecture and construction of virtual networks in a cloud environment.

Cloud Traceability (CBoC TRX)

Shinichi Nakahara†*, Naoto Fujiki, and Shigehiko Ushijima*

Abstract

In this article, we introduce a cloud forensics service for cloud security that can visualize a sequence of operations and the lifecycle of data and a traceability platform for implementing this service. This visualization contributes to overcoming the security worries and concerns about data leakage and unintended deletion that the majority of potential cloud users have over the inability to see the progress or operation of cloud services. The infrastructure has an architecture that allows various functions to be added or substituted, allowing large traceable logs of various types to be linked together to suit users' needs.

1. Introduction

For many potential new users, the inability to know the operational state of cloud services or where the service and data themselves actually reside is a barrier to introducing such services [1]. Therefore, it is important to inform users about what kind of processing the cloud service is performing and who performed what human actions and where their data is located [2], [3]. A traceability platform overcomes these non-transparency concerns by reproducing and displaying the chain of events from log information indicating human operations, file transfers, and process activity as well as information from related systems such as authentication and equipment management systems.

2. Cloud forensics

The cloud forensics service (**Fig. 1**) is a value-added service aimed at allowing anxious users to receive cloud services with confidence, as well as letting cloud services providers confidently provide services that will satisfy users, by giving the cloud system the ability to explain who (or what) did what, as well as when, where, in what way, and with what result. Specifically, the cloud forensics service maintains evidence (logs) of events occurring in the cloud and links them together according the users' or operators' intentions to provide event verification and visualization. For example, it enables checking of the lifecycle of a user file (creation, editing, transfer, and deletion) (from the user's perspective), checking of which operator performed which operations on a user's environment configuration (operator's perspective), and verification of whether the system is operating according to regulations (auditor's perspective).

For this purpose, the system must maintain traceable event information in the 4W1H1R (when, where, who, what, how, and what result) format.

3. Traceability platform architecture

To implement such cloud forensics, we have started developing the Common IT Bases over Cloud Computing (CBoC) cloud traceability system (CBoC TRX), which gathers, stores, links together, and refers to logs of events occurring in the cloud (IT: information technology). The system architecture consists mainly of a log-trace section, which gathers and normalizes large volumes of logs output from servers and access points; the log security section, which maintains the security of the logs; and the log operation management section, which performs log provisioning, which enables different logs to be linked and traced (**Fig. 2**).

† NTT Information Sharing Platform Laboratories
Musashino-shi, 180-8585 Japan

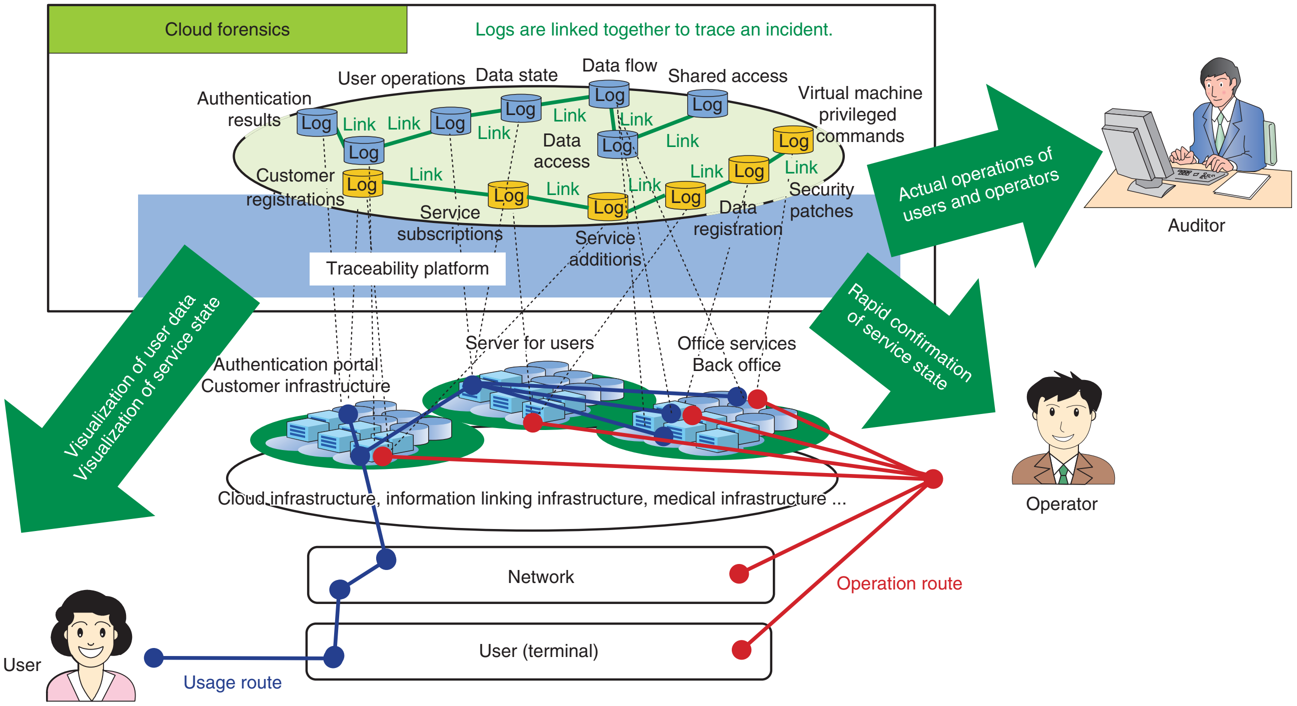


Fig. 1. Cloud forensics.

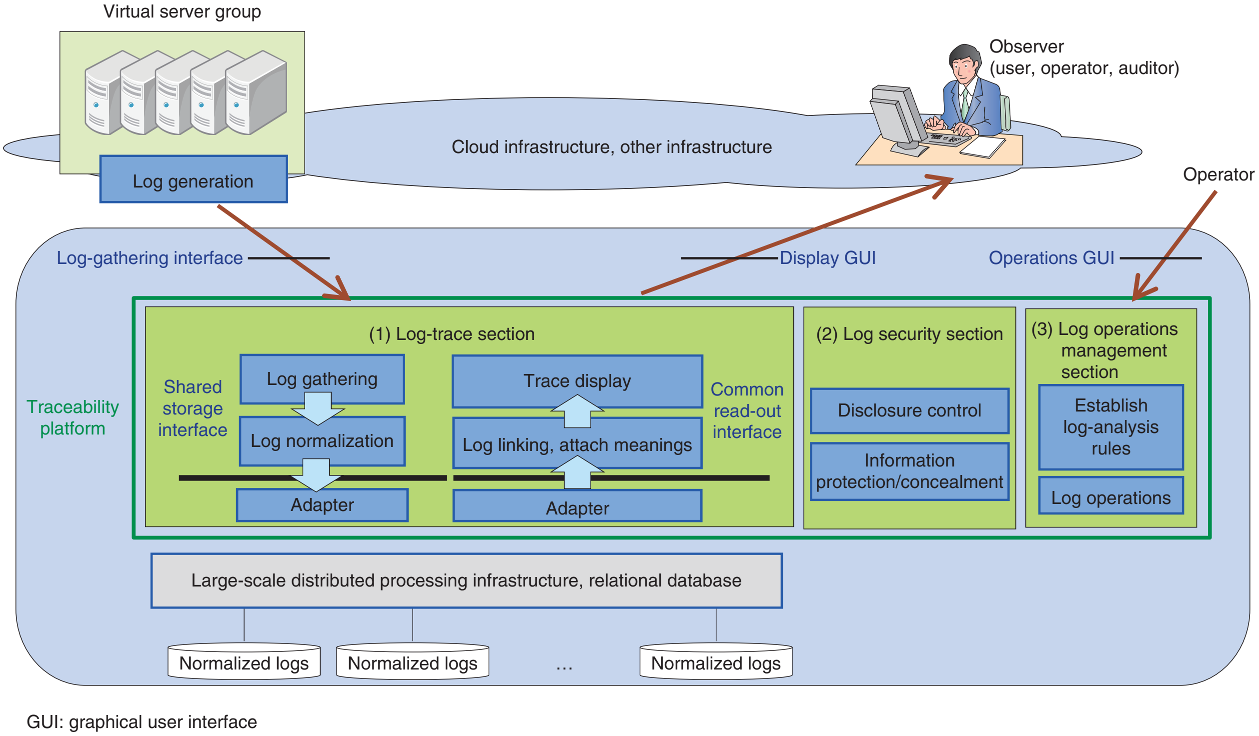


Fig. 2. Traceability platform architecture.

The log-trace section processes, normalizes (mapping to 4W1H1R format), and stores the large volume of input logs. To maintain storage and data processing scalability, the storage of large logs can be linked to large-scale distributed processing infrastructures such as CBoC Type 2 or Hadoop[*1].

To avoid dependence on a particular infrastructure when linking with large-scale distributed processing infrastructures in particular, we decided on a common access interface (storing and reading out) for programs being used, and infrastructures are accessed through adapters supporting this interface. The log-trace section and the log security section were designed with a reusable component architecture, with uniform access interfaces between the functional blocks, and with independence at the launch and unit operation level being maintained.

4. Infrastructure features

Building log security, scalability, and log normalization into the log infrastructure brings the following features to the traceability platform and improves usability.

4.1 Log security

Log evidence needs to be admissible. To improve log admissibility, three measures are taken: (1) the accuracy of terminal operation log timestamps for the Windows operating system is improved by linking to an NTP (network time protocol) server, (2) input logs (primary logs) have file-level anti-interpolation protection provided by a tamper detection function, and (3) input logs are normalized and complemented to ensure that there is sufficient information to reconstruct actual events.

Evidence of any leakage of log data or tracing information is needed. To overcome one of the concerns about cloud technology, which is that information could leak to other users or operators, stored logs can be encrypted. Log data disclosure is also restricted according to the viewer's authorization, and user names and other private information are not displayed except to the actual person.

4.2 Scalability

To process the large volume of logs for cloud services, we implement scalability in the collection and parallel processing of log data using Flume (open source software) for the process of gathering logs from many locations, and we link to systems such as CBoC Type 2, Map/Reduce, and the Hadoop Distributed File System for storage and search processing.

4.3 Log normalization

Input log data (called log messages, since they are normally input as datagrams) is normalized by separating it into elements and mapping each element to a 4W1H1R category. If the information is insufficient for tracing at this point, the data is complemented with information from other logs or other pre-defined external data. Log normalization has three effects.

(1) Logs can be linked to support a variety of trace requirements.

By clearly defining the meaning of individual log message elements, we enable a single log message to be used for mapping trace-service requests for various purposes. In this way, conventional logs output for specific purposes can be used to reproduce actual events.

(2) Trace scopes can be extended by adding additional logs.

New logs are normalized when they are added to the traceability platform, so if there are log elements with the same meaning as elements already stored in the infrastructure, or if it is possible to add logs from a new time period or information about new services or functions in order to define how the logs are connected, the scope of the trace can be extended temporally or spatially.

(3) Enhance adaptability to key-value store

A fast search can be implemented for keys such as user or file names by storing data in the key-value store (KVS[*2]) format and by using 4W1H1R elements as key data.

4.4 Integration into the log infrastructure

To enable the traceability platform to handle logs in any existing format, we made it possible in the log operations management section to define parsing rules for logs and rules mapping from log elements to 4W1H1R elements. For this purpose, we provided a graphical user interface, which makes it easy to integrate new logs into the infrastructure, which already has normalized log data.

[*1] Hadoop: A Java software framework for distributed processing of large-scale data.

[*2] KVS is a format commonly used for Internet searching and other applications.

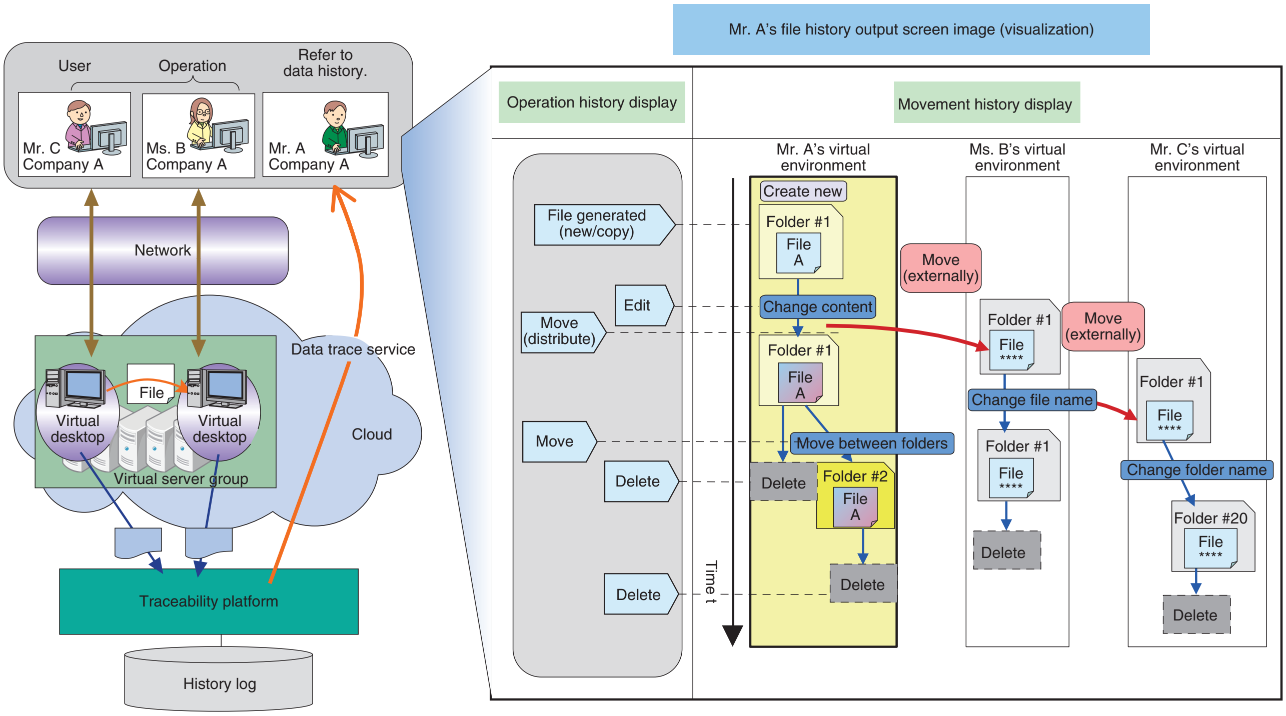


Fig. 3. Data trace concept.

5. Trace service concept

5.1 Data tracing

The log of a user's operations at a terminal is summarized and managed by CBoC TRX in 4W1H1R format, revealing what operations the user did on what files (data) (**Fig. 3**). This allows the record of what operations—creating, editing, copying, moving, or deleting—were done when and by whom on a given file.

The user can visually check the historical sequence of file creations, deletions, and transfers to other people even though the data is entrusted to the cloud operator, so the history of a file can be checked easily, including its source, where it went, or whether it was deleted according to requirements. In principle, for files passed to other people, the history is displayed with personal information such as file names anonymized, e.g., letters replaced by meaningless characters such as *, but it is also possible to suppress any display of this information.

5.2 Operation tracing

As with data tracing, terminal operation logs (4W1H1R) are also collected. The system summarizes and links together both user and operator logs, so the relationships between operations of both can also be seen (**Fig. 4**).

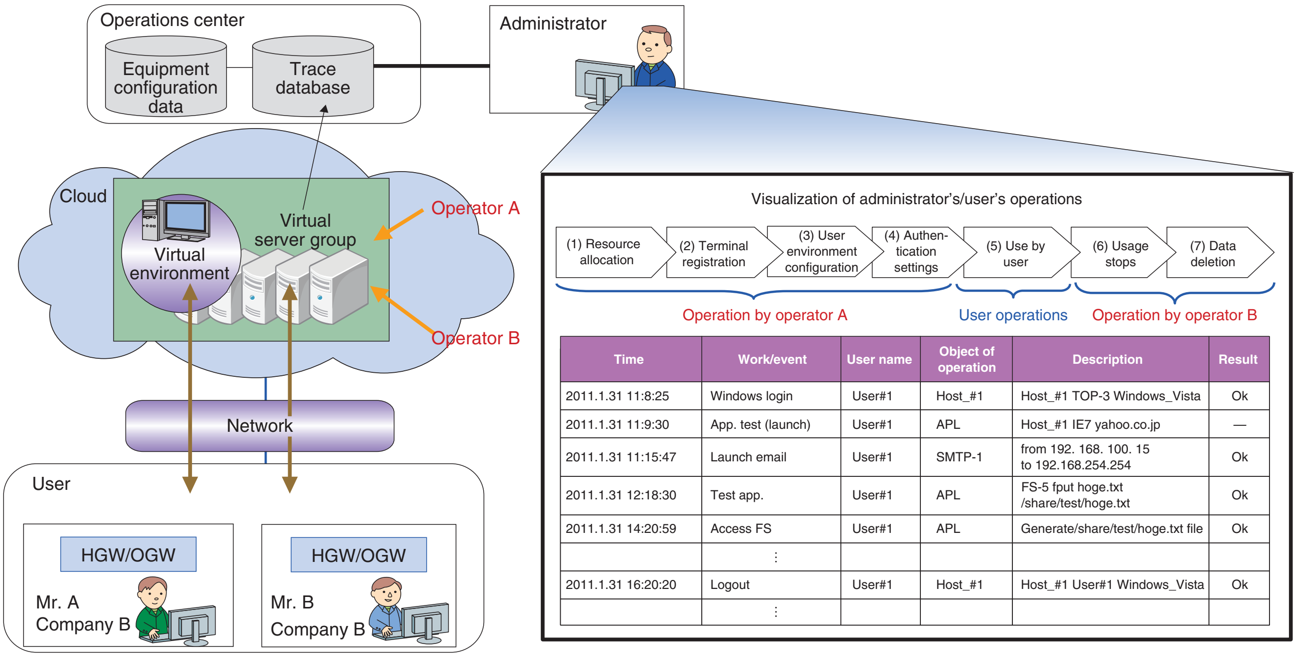
Normally, it is impossible to link user logs with operator logs, but adding additional information that can be parsed and linked, such as operation logs from the same virtual machine or computer, enables the sequence of operations to be reconstructed despite the barrier of independence between user and operator.

Even if a cloud user's environment is configured by different operators, the multiple logs from individual operators are linked, and the sequence of operations can be recognized. This also applies when a service system is composed of multiple services and servers. Using the common information in the logs enables operations spanning people, services, and systems to be rapidly linked together and understood.

6. Future initiatives

6.1 Strengthen security

The security functions of CBoC TRX were implemented from the perspective of making the logs admissible and preventing information leaks, but we



Time	Work/event	User name	Object of operation	Description	Result
2011.1.31 11:8:25	Windows login	User#1	Host_#1	Host_#1 TOP-3 Windows_Vista	Ok
2011.1.31 11:9:30	App. test (launch)	User#1	APL	Host_#1 IE7 yahoo.co.jp	—
2011.1.31 11:15:47	Launch email	User#1	SMTP-1	from 192. 168. 100. 15 to 192.168.254.254	Ok
2011.1.31 12:18:30	Test app.	User#1	APL	FS-5 fput hoge.txt /share/test/hoge.txt	Ok
2011.1.31 14:20:59	Access FS	User#1	APL	Generate/share/test/hoge.txt file	Ok
	⋮				
2011.1.31 16:20:20	Logout	User#1	Host_#1	Host_#1 User#1 Windows_Vista	Ok
	⋮				

HGW: home gateway
OGW: office gateway
App.: application
FS: file system

HGW: home gateway
OGW: office gateway
SMTP: simple mail transfer protocol
TOP: user login account name

Fig. 4. Operations traceability.

plan to strengthen them further in the future taking into consideration issues regarding the use of public information and protection of private or sensitive information and using additional technology such as long-term digital signatures and evidence record syntax.

6.2 More active use of logs

We will make improvements in log linking through techniques such as (1) resolving the various IDs (identities) of the same person on different services and using them to link logs and (2) integrating logs with other logs by recognizing the configuration between a physical device and logical devices. This is expected to meet the accountability obligations of cloud service providers and improve operational efficiency.

References

[1] Ministry of Economy, Trade and Industry, "Survey Report on Information Security Audits of Cloud Services," Jan. 2010 (in Japanese).

[2] S. Nakahara and H. Ishimoto, "A Study on the Requirements of Accountable Cloud Services and Log Management," Proc. of APSITT 2010, pp. 1–6, Kuching, Malaysia, 2010.

[3] Hewlett Packard report.
http://www.hpl.hp.com/techreports/2011/HPL-2011-38.pdf

Shinichi Nakahara

Senior Research Engineer, Supervisor, Security Management SE Project, NTT Information Sharing Platform Laboratories.

He received the B.E. and M.E. degrees in electrical engineering from Osaka University in 1984 and 1986, respectively. He joined NTT Yokosuka Laboratories in 1986, where he engaged in research on operating systems. He is currently studying cloud security and data & operation traceability in the cloud. He is a member of IEEE and the Information Processing Society of Japan (IPSJ).

Naoto Fujiki

Senior Research Engineer, Security Management SE Project, NTT Information Sharing Platform Laboratories.

He received the B.E. and M.E. degrees in precision engineering from Niigata University in 1989 and 1991, respectively. He joined NTT in 1991 and studied information sharing systems, network operation systems, and network security. He is currently studying traceability systems. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and IPSJ.

Shigehiko Ushijima

Senior Research Engineer, Supervisor, Communication Platform SE Project, NTT Information Sharing Platform Laboratories.

He received the B.E. and M.E. degrees in electronic engineering from Keio University, Kanagawa, in 1986 and 1988, respectively. He joined NTT Communication Switching Laboratories in 1988. His recent research area is cloud computing architecture and security. He is a member of IEICE.

FreeCloud: A Trial Service for OpenStack

Nachi Ueno[†], *Hisaharu Ishii*, *Keisuke Tagami*, *and Koji Iida*

Abstract

This article describes NTT's work on the OpenStack trial service FreeCloud, which is intended to improve the operability of OpenStack and expand the set of operation tools.

1. Introduction

OpenStack [1] is a cloud management platform that has been attracting attention. It is open source software for virtual machine infrastructure in the form of infrastructure as a service (IaaS). NTT is participating in the FreeCloud project, which is operating FreeCloud as a trial service for OpenStack[*1]. The FreeCloud project is managed by the OpenStack community, which currently has over 60 corporate participants, including NTT DATA and the NTT R&D Laboratory Group. With over 70 active developers, the community is engaged in vigorous activity, so it is highly likely that OpenStack will become an industry standard for IaaS open source software.

2. Toward higher operability

The objectives of FreeCloud are to construct a DevOps[*2] system, expand the set of operation tools, and define a reference architecture that combines hardware configuration and setup standards to improve OpenStack operability.

2.1 DevOps

The insufficiency of operation manuals and operating tools is a problem for OpenStack. To solve that problem, the FreeCloud project is working on a development method called DevOps. This maintains tighter coupling between the operation and development teams by adding functions while the system is operating. The DevOps method permits early feedback of operational problems to the developers. The trends in DevOps are (1) resource virtualization through cloud computing and IaaS and (2) automation of many operations by using the Puppet middleware setup automation tool [2] and tools such as Chef [3]. This reduction in the distance between development and operation has attracted the attention of many engineers. The creation of a DevOps system in the FreeCloud project can improve operability through clear documentation of the operating procedures and feedback from operations to the development community.

2.2 Implementation of efficient operation tools

The FreeCloud project is developing OpenStack operation tools. The specific tools planned for development include an autoinstall function and a monitoring function. The OpenStack automatic installation system comprises Cobbler, a tool that automates operating system (OS) installation [4], and Puppet. The OpenStack monitoring function uses the notification function that was implemented in the September 2011 release of OpenStack [5].

2.3 Reference architecture

OpenStack has various settings to support many different kinds of virtualization software. The

† NTT Information Sharing Platform Laboratories
Musashino-shi, 180-8585 Japan

*1 Since this topic is still under study, the final details may differ somewhat.

*2 DevOps: A coined term that combines development and operations.

FreeCloud project provides for the sharing of a reference architecture and service operation by the community and for the preparation of community standard operating manuals and operation tools.

3. Operation by the community

Volunteer developers from the OpenStack community are operating FreeCloud. The tool set developed in the FreeCloud project is also planned to be open to the OpenStack community. Community management of FreeCloud will make it possible to produce industry standard manuals and tools for OpenStack operation.

4. Free offering to users

FreeCloud is intended to be made available to any user without charge. It is offered free because it is a trial version, so it is difficult to guarantee commercial quality. It is open to any user for two reasons: as a way to expand the community and as a way to reproduce the environment for providing actual services in which it is impossible to assume a user base. The plan is to select a certain number of users from among applicants and let each one use a virtual machine service for about one week.

5. System configuration

The FreeCloud project currently has two setup patterns for FreeCloud, both of which are planned to be included in the next OpenStack release.

The first FreeCloud setup pattern assumes a public cloud environment that provides IaaS to general users. The public cloud settings are the Ubuntu (Linux) OS and Xen virtualization software, which are widely used in the U.S. market. For the network setup as well, all of the virtual machines exist on the same network.

The second FreeCloud setup pattern assumes a private cloud that provides IaaS within an enterprise. It uses Red Hat Enterprise Linux 6 or an equivalent OS and kernel-based virtual machine (KVM) virtualization software, which has a large share of the market in Japan. A virtual local area network function is used to isolate virtual user networks.

Both patterns include functions such as ones for starting up virtual machines from a web application program interface (API)[*3] or web graphical user interface (GUI)[*4]. Two web APIs are provided. One is OpenStack's own API and the other is compatible with Amazon EC2, which is compatible with the IaaS offered by Amazon. For the web GUI, two versions are planned: OpenStack Dashboard [6] and the Clanavi interface being developed by DOCOMO Communications Laboratories USA, Inc. [7].

6. FreeCloud project management system

Discussions within the FreeCloud project as well as the materials and tools are published on the project's website [8]. FreeCloud is being managed by members registered on that website [9]. The leader of the OpenStack virtual machine image[*5] management function development project and the leader of the OpenStack quality improvement project are currently participating as committee members in addition to NTT members. The weekly development meetings are conducted online. Because the entire process is public, anyone who is interested can participate and comment.

7. Schedule

The FreeCloud release is expected to be announced and the service inaugurated at OpenStack Design Summit 2011, in October 2011. We encourage anyone who is interested to try the service. The OpenStack Design Summit will bring together the companies that are developing OpenStack and companies that may plan to use it. The previous event in San Jose was attended by over 400 persons and was very successful. At the Design Summit, we also plan to enlist sponsors to support the continuation of FreeCloud.

8. Concluding remarks

The FreeCloud project is intended to improve the operability of OpenStack through actual operation. The expected results are operating procedures, an expanded set of operation tools, and a reference architecture definition.

[*3] Web API: Application programming interface that can be used via the web.

[*4] Web GUI: Graphical user interface that can be used via the web.

[*5] Virtual machine image: Data for generating a virtual machine.

References

[1] OpenStack. http://www.openstack.org/

[2] Puppet. http://www.puppetlabs.com/
[3] Chef. http://www.opscode.com/chef/
[4] Cobbler. https://fedorahosted.org/cobbler/
[5] OpenStack notification function. https://blueprints.launchpad.net/nova/+spec/notification-system
[6] Openstack Dashboard. https://github.com/openstack/horizon
[7] Clanavi. http://drupal.org/project/clanavi
[8] Freecloud website. https://launchpad.net/freecloud
[9] Freecloud admins. https://launchpad.net/~freecloud-admins

Nachi Ueno

Research Engineer, NTT Information Sharing Platform Laboratories.

He received the B.E. and M.Sc. degrees from Waseda University, Tokyo, in 2004 and 2006, respectively. He joined NTT Information Sharing Platform Laboratories in 2006 and has been researching identity management technology and cloud computing technology.

Keisuke Tagami

Research Engineer, NTT Information Sharing Platform Laboratories.

He received the B.Physics degree from Tokyo Metropolitan University and M.Eng. degree from Tokyo Institute of Technology in 2004 and 2006, respectively. He joined NTT DATA in 2006 and designed and developed enterprise systems of network and distributed system. He joined NTT Information Sharing Platform Laboratories in 2011 and has been researching cloud computing technology.

Hisaharu Ishii

Research Engineer, NTT Information Sharing Platform Laboratories.

He received the B.E. and M.Sc. degrees from Yokohama National University, Kanagawa, in 2006 and 2008, respectively. He joined NTT Information Sharing Platform Laboratories in 2008. His research interests include data mining and cloud computing.

Koji Iida

Senior Research Engineer, NTT Information Sharing Platform Laboratories.

He received the B.E. and M.Sc. degrees from Keio University in 1993 and 1995, respectively. He joined NTT Information Platform Laboratories in 1995 and studied enterprise communication middleware and distributed object technologies. He joined NTT Information Sharing Platform Laboratories in 2007 and has been researching identity management technology and cloud computing technology.

Power-management-circuit Techniques for Low-power Intermittent LSI Operation in Wireless Applications

Mamoru Ugajin[†], *Toshishige Shimamura, Akihiro Yamagishi, Kenji Suzuki, and Mitsuru Harada*

Abstract

Power-management-circuit techniques for low-power intermittent LSI (large-scale integrated circuit) operation are discussed for various wireless transceiver designs. For an RFID (radio-frequency identification) application, the most important issue is reducing the power of the clock timer; we used an analog RC (resistor-capacitor) timer circuit with megaohm resistors. For a sensor node that has a power generator, the ability to operate with a nanowatt-level power supply is essential; we devised a two-stage power management circuit with a subnanowatt voltage-detection circuit. For wide-area-ubiquitous-network transceivers, the power-management circuit should supply 100 mA of current in the active mode and reduce the leakage current to the nanoampere level; we developed a regulator circuit with a reversely biased nMOS/pMOS cascode switch to reduce the leakage current and with a DC-DC converter for supplying sufficient current without any increase in power consumption.

1. Introduction

The number of wireless applications using small terminals, such as the wide area ubiquitous network (WAUN [1]–[2]), wireless sensor networks [3]–[6], and active RFID (radio-frequency identification) [7]–[12], has been increasing. Small wireless terminals must have very low power consumption and small power supplies such as coin batteries or small power generators [3]–[6]. In such wireless systems, intermittent operation is one of the key techniques for reducing the power consumption of the terminals [13], [14]. The average power consumption of the terminals can be reduced by decreasing the activity ratio of the intermittent operation. However, the method of controlling intermittent operation varies depending on the power consumption of the transmitters, preciseness of the transmission timing, power supply capabilities, and so on.

In this article, we describe power-management-circuit techniques for low-power intermittent LSI (large-scale integrated circuit) operation in power-management-circuit designs for three different applications: RFID devices, sensor nodes, and WAUN transceivers. For RFID, the most important issue is reducing the power of the clock timer; for this purpose, we use an analog RC (resistor-capacitor) timer circuit with megaohm resistors, all-digital RF transmission, and automatically tuned RF pulse generation [15]–[17]. For an ultrasmall sensor node, a nanowatt-order power supply is essential; we have made one by using a two-stage power management circuit with a subnanowatt voltage detection circuit [18]–[24]. For WAUN transceivers, the power-management circuit must be able to supply 100 mA of current in the active mode and reduce the leakage current to the nanoampere level; we have developed regulator circuits with a reversely biased nMOS/pMOS (positive/negative-type metal oxide semiconductor) cascode switch to reduce the leakage current to the nanoampere level and with a DC-DC converter

[†] NTT Microsystem Integration Laboratories
Atsugi-shi, 243-0198 Japan

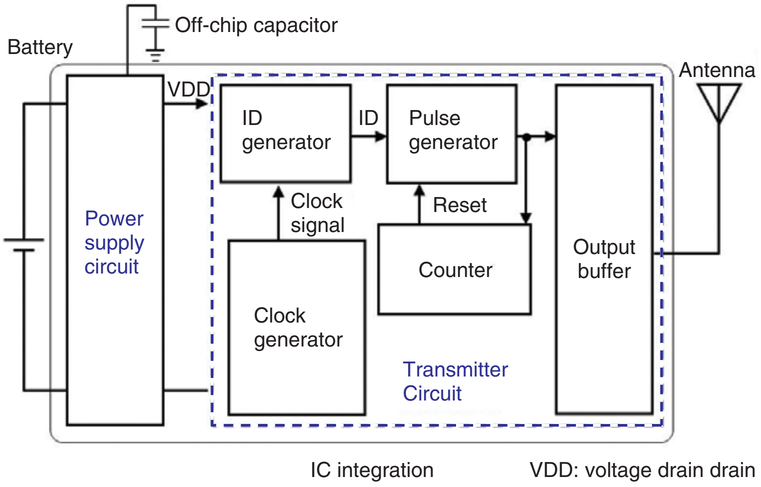


Fig. 1. Block diagram of the active RFID IC.

to supply sufficient current without any increase in power consumption [25], [26]. Our techniques for these applications are described in sections 2–4, respectively.

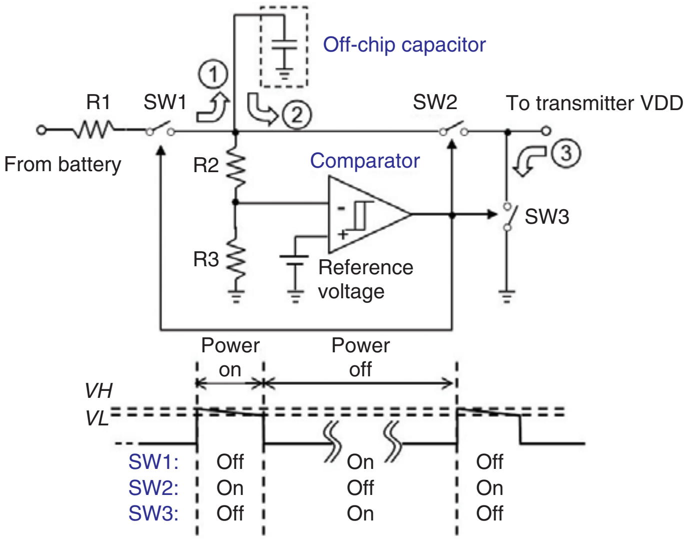
2. Micro-power active-RFID LSI

For an RFID application, the power-supply circuit usually has a timer clock (or processor) and provides power to the transmitter intermittently. However, a timer circuit that uses a crystal oscillator consumes a considerable amount of power. Our power reduction solution uses an analog RC timer circuit with megaohm resistors. Battery voltage fluctuations are stabilized through the use of all-digital RF transmission and automatically tuned RF pulse generation; as a result, the integrated circuit (IC) is suitable for practical use. A block diagram of the active RFID IC is shown in **Fig. 1**. The IC is divided into two major blocks: a power-supply circuit block and a transmitter circuit block. The battery and antenna are external components. The power-supply circuit controls the transmitter's power and its on and off states. When power is provided by the power supply circuit, the transmitter starts to operate and it sends the device's ID (identification) to a receiver.

The architecture of the power-supply circuit, which consists of three switches, a comparator, and an off-chip capacitor, is shown in **Fig. 2**. Initially, switches SW1 and SW3 are on, and SW2 is off. The energy from the battery is stored in a storage capacitor to obtain a sufficient voltage level for transmitter operation. The comparator senses the change in the level and compares it with a reference voltage. When the voltage reaches the high threshold voltage (*VH*) level, the comparator turns on SW2 and turns off SW1 and SW3. Then DC power is provided to the transmitter from the capacitor, and the circuit starts to operate and it sends the ID. During this operation, the energy stored in the capacitor is consumed and the capacitor's voltage level decreases. When the voltage reaches the low-threshold-voltage (*VL*) level, the comparator turns off SW2 and turns on SW1 and SW3. The battery starts to recharge the capacitor. At this point in the timing, the transmitter immediately stops operating because SW3 shorts the VDD (voltage drain drain) line to ground. This prevents unstable IC operation by avoiding operation at an insufficient voltage. This cycle is repeated until the battery runs down. This intermittent operation is effective for achieving a long lifetime in an active RFID tag. The typical operating voltage of an external battery is 3.4 V. *VH* and *VL* are 2.1 and 1.9 V, respectively. The total average current drawn from the battery is 1.6 μA. The comparator and voltage-reference circuit are the only components operating continually in the IC, and they consume about half the total average current.

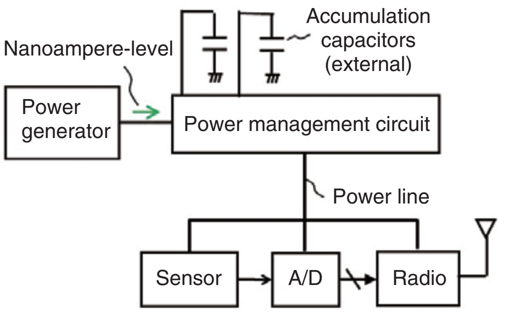
3. Nanowatt wireless sensor nodes

Wireless sensor nodes can be powered by small batteries (e.g., coin batteries) or small power generators. One popular approach is energy harvesting, in which



SW: switch
$V_{ref.}$: reference voltage

Fig. 2. Architecture of power supply circuit.



A/D: analog-to-digital converter

Fig. 3. Target architecture of the nanowatt sensor node.

energy is captured from external sources such as solar power, thermal energy, or kinetic energy (e.g., movement of a human body carrying the sensor). For a sensor node with a power generator, one of the most important issues is size reduction: sensor nodes as small as dust grains are expected [5]. However, the amount of power that a power generator can generate is almost proportional to its size [6]. Dust-sized sensor nodes must operate with nanoampere-level power generators [18]–[24]. Since the generated power is much less than that needed for radio transmission, the generated energy must be accumulated by the sensor node until it has enough for a transmission burst (intermittent operation) [3]. However, since the power management circuit runs continuously not intermittently, the power management circuit's power consumption cannot exceed the generated power. Thus, the power management circuit, which monitors the accumulated energy and controls the current flowing into the radio block, is a key circuit affecting the minimum current that must be accumulated.

The target sensor-node architecture is depicted in **Fig. 3**. The architecture has four key elements: a power generator, power management circuit, vibration sensor, and radio block. The sensor nodes should be smaller than 1 mm^3 [6]. Thus, the power generator outputs nanowatt-level power and the power management circuit accumulates that energy in capacitors. When enough energy to operate the radio has been accumulated, the power management circuit supplies power to the sensing circuit, analog-to-digital converter, and radio block.

For a design in which energy from a generator is accumulated and supplied to the radio block, there are two issues to resolve. One, as mentioned above, is that the total power consumption of the power management circuit must be less than the generated power. The total power consumption consists mainly of the power consumed for monitoring the voltage and controlling the switches and the leakage current

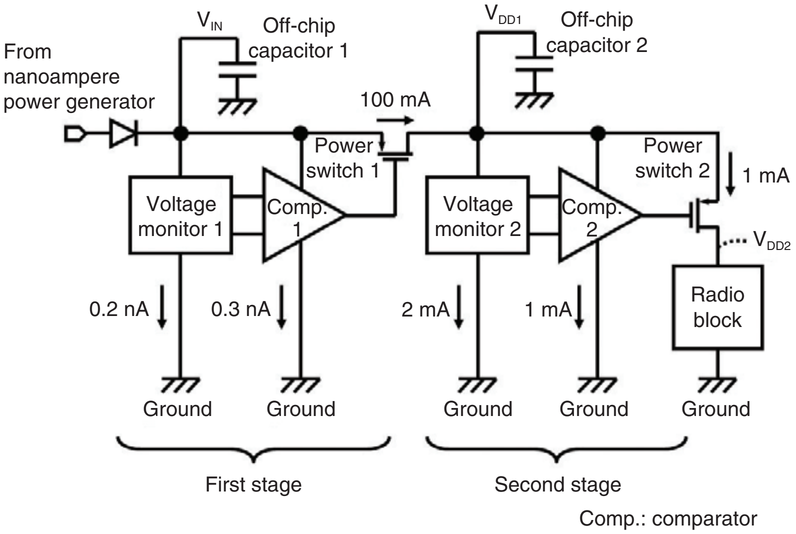


Fig. 4. Two-stage power management architecture.

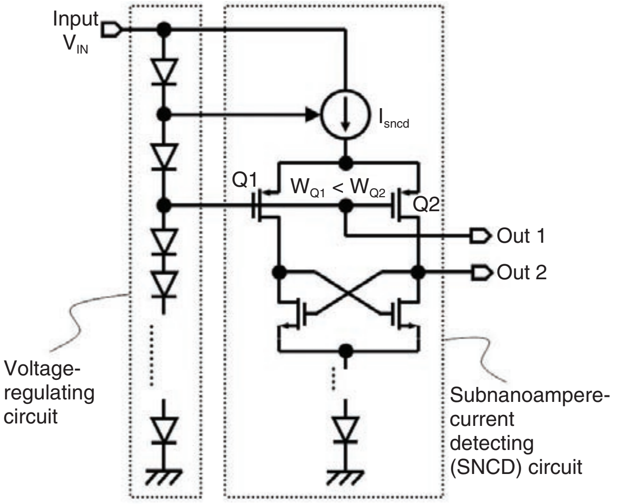

Input V_{IN}
I_{sncd}
Q1
$W_{Q1} < W_{Q2}$
Q2
Out 1
Out 2
Voltage-regulating circuit
Subnanoampere-current detecting (SNCD) circuit

Fig. 5. Voltage-monitoring circuit in the first stage.

flowing through the power switch transistor. The other issue is that the transition time of the power switch should be sufficiently shorter than the operation time of the radio block. This is because the output voltage of the management circuit is too low for radio operation during the transition period, so during this period the output power is wasted. Such waste should be minimized so that most of the accumulated energy is used for radio operation.

A simulation showed that the sum of the switch leakage current and comparator current is 19 nA for a pMOS switch or 11 nA for an nMOS switch [24]. Since these values are both greater than 1 nA, a single-stage power management circuit cannot accumulate energy from a nanoampere power generator. Therefore, we designed a two-stage power management architecture, as shown in **Fig. 4**. This architecture can resolve both of the issues described above. The first stage accumulates the energy from a nanoampere power generator and supplies about 100 µA of current to the second stage. The second stage accumulates the 100-µA current from the first stage and supplies about 1 mA of current to the radio block.

The voltage-monitoring circuit in the first stage is shown in **Fig. 5**. The current consumption of the first stage must be much lower than that of the second stage because the first stage must operate continuously whereas the second stage operates intermittently. With a previously proposed technique, the low-power band-gap reference circuit consumes about 0.2 µW [27]; therefore, the remaining available power is too low to enable the power management block to operate with nanowatt-level power. This is because calculations based on Ohm's law show that the resistors in the band-gap reference circuit operating at a voltage of 2 V would need to be more than 10 GΩ for subnanoampere operation. 10-GΩ resistors would require several square centimeters of chip area, which would be almost impossible in the available area on the chip.

A subnanoampere-current-detecting (SNCD) circuit is used for the voltage-monitoring circuit in the first stage, as shown in Fig. 5. The voltage-monitoring circuit features subnanoampere operation, with voltage regulation through a series-connection of diode-connected MOS field-effect transistors (MOSFETs) and with positive-feedback amplification due to a cross-coupled transistor pair. The diode-connected MOSFETs in the voltage-regulating circuit are six pMOSFETs and three nMOSFETs; those in the SNCD circuit are three nMOSFETs. The threshold voltages of the pMOSFETs and nMOSFETs are -0.33 V and 0.30 V, respectively. The simulated transition characteristics of the input voltage V_{IN} and the outputs of the voltage-monitoring circuit in the first stage (Out 1 and Out 2) are shown in **Fig. 6**. The input voltage V_{IN} is the voltage of the accumulated energy. Initially, the diode-connected MOSFETs are in the off state. When V_{IN} applies the threshold voltage to each diode, a subnanoampere current is generated in the voltage-regulating circuit. This current is mirrored to the current source I_{sncd} in the SNCD circuit. This causes a current difference in MOSFETs Q1 and Q2 in the designed channel width: WQ2 > WQ1 (Fig. 5). This current difference is amplified and converted into a voltage signal by the cross-coupled transistor pair and the relationship between voltages Out 1 and Out 2 is inverted. This differential signal from the voltage-monitoring circuit is converted into a pulse by the hysteresis comparator, which means that the voltage of the accumulated energy reaches the voltage determined by the number of diode-connected

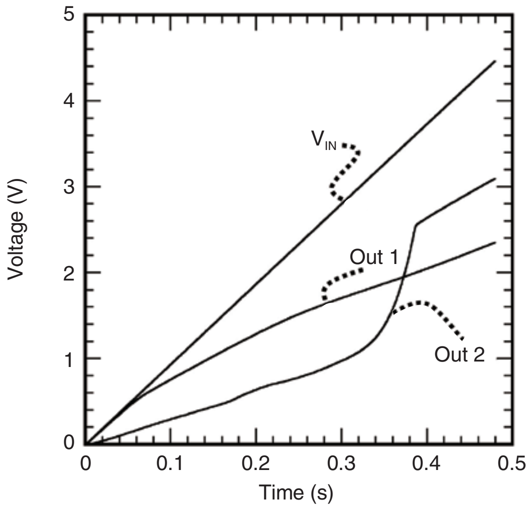


Fig. 6. Simulated transition characteristics of the input voltage V_{IN} and the outputs of the voltage-monitoring circuit in the first stage (Out 1 and Out 2).

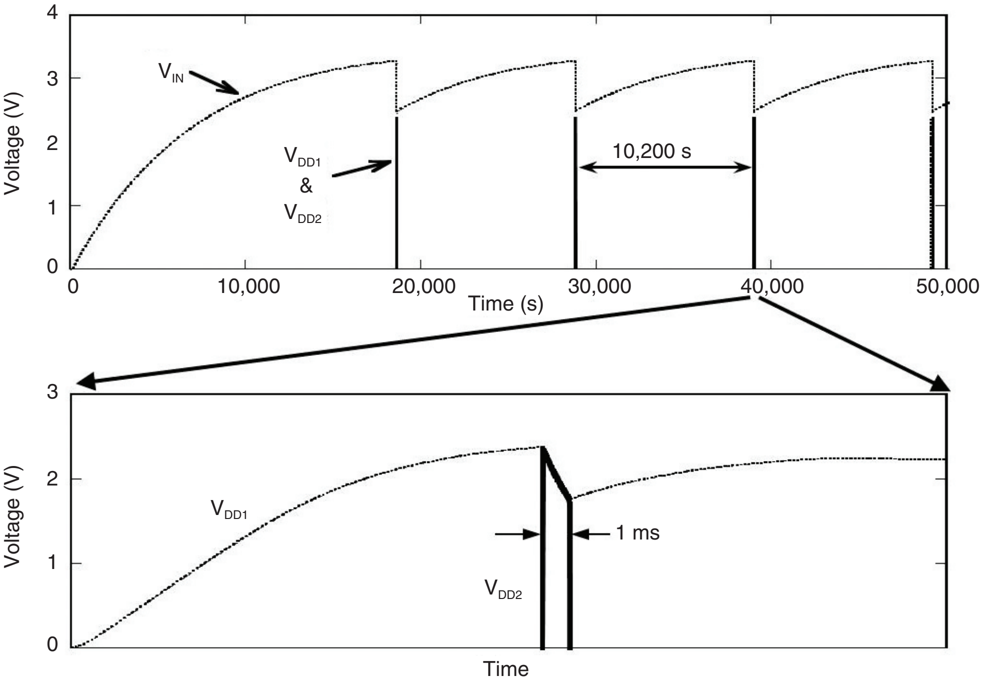


Fig. 7. Simulated transition characteristics of the subnanoampere two-stage power management circuit.

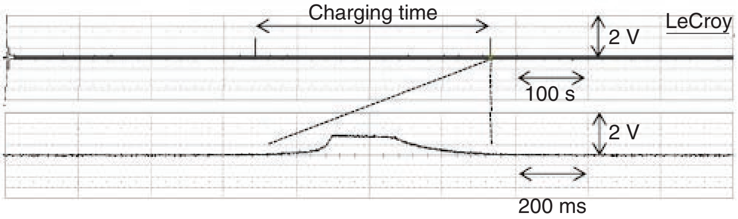


Fig. 8. Measurement results for the charging waveform of V_{DD1}. In this measurement, 1-nA DC was input to the rectifier.

MOSFETs. Thus, our circuit can monitor the voltage of the accumulated energy and regulate it using low power consumption at a subnanoampere current.

The simulated transition characteristics of the subnanoampere two-stage power management circuit are shown in **Fig. 7**. The input voltage V_{IN}, the first supply voltage V_{DD1}, and the second supply voltage V_{DD2} were evaluated at the points depicted in Fig. 4. In the simulation, the input current was 1 nA and the values for the first and second off-chip capacitors were 2.2 µF and 0.7 µF, respectively. The interval between intermittent bursts of the final accumulated output (V_{DD2}) was about three hours. Our on/off keying (OOK) transmitter [15]–[17] can transmit 1000 bits of data using the V_{DD2} output.

To confirm the effectiveness of our circuit techniques, we fabricated a test chip using the 0.35-µm complementary MOS (CMOS) process. To evaluate the characteristics of the voltage-monitoring circuit, we measured the time taken to charge the 2.2-µF accumulation capacitor by using a DC current source. Charging time was obtained from the waveform in **Fig. 8**. In this measurement, DC current of 1 nA was input to the rectifier, and the voltage at the measurement point of the switch controlled by the voltage monitoring circuit (V_{DD1} in Fig. 4) was measured with a digital oscilloscope having impedance of 1 MΩ. The accumulation capacitance was chosen to be 2.2 µF, which is the amount necessary to operate our OOK transmitter for one data transmission. The period of the measured pulse corresponds to the charging time, as shown in the upper waveform.

In **Fig. 9**, the solid line shows the measured charging time for our power management circuit, which includes a voltage detection circuit, and the dashed line is the calculated charging time for a previously reported voltage detection circuit [27]. If this previous circuit is used, the accumulation capacitor cannot be charged when the generated current is less than 0.1 µA.

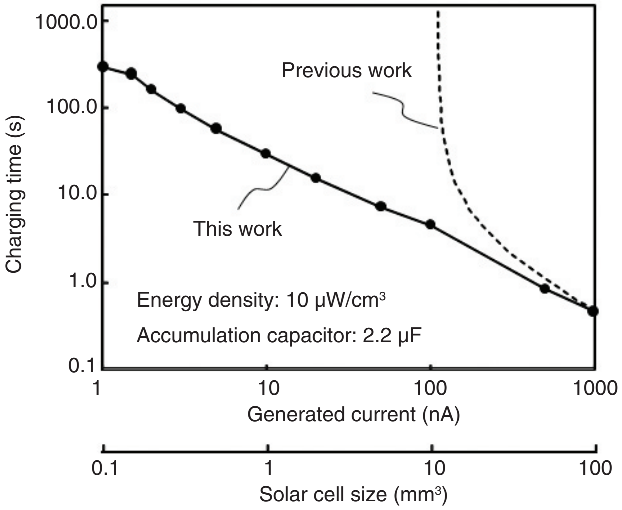


Fig. 9. Measured and calculated charging time for the reported circuit.

On the other hand, for our circuit, the accumulation capacitor can be charged with a 1-nA current, which corresponds to a solar cell less than 0.1 mm^3 in size when the solar energy density is 10 $\mu W/cm^3$. This means that the size of the power source needed for our circuit is 1/100 of that for the circuit in [27].

4. WAUN transceivers

Terminals used in a WAUN must have small power supplies. To reduce power consumption, the wireless terminals operate intermittently with a very low activity ratio. Power reduction with a very low activity ratio strongly depends on power consumption in the standby mode. Thus, we use two key techniques for reducing the standby current: a special power switch for low leakage current and a minimized number of active blocks in standby mode. The standby current of the terminals is less than 5 μA, which is about 1/100 of that for other applications, such as ZigBee and PHS (personal handy-phone system). One of the most important circuits for low-standby-current performance is a power switch that can supply 100 mA in the active mode and reduce the leakage current to the nanoampere level in the standby mode.

The off-state leakage current of MOSFETs is determined mainly by the subthreshold leakage, gate-induced drain leakage, and junction leakage currents [28]. MOSFETs based on semiconductor-on-insulator (SOI) technologies have a smaller subthreshold swing and smaller junction area than those based on bulk technologies. Thus, the subthreshold leakage and junction leakage currents of power switches can be aggressively reduced by using SOI technologies [29]–[30]. The voltage between the gate and source (gate-source) voltage dependences of the off-state leakage current with a constant drain-source voltage for pMOS power switches using SOI and bulk technologies are illustrated in **Fig. 10**. The leakage current of SOI MOSFETs at the gate-source voltage of 0 V is about 1/100 that of bulk MOSFETs. This means that just using SOI technologies can reduce the leakage current satisfactorily.

The architecture of the power switch and regulator circuits with CMOS/SOI technology is shown in **Fig. 11**. The power switch has a conventional architecture and consists of two pMOSFETs. A depletion-mode n-MOSFET (D-MOS) is used in the regulator circuit. This transistor has a negative threshold voltage and can control a large current by means of a low gate-source voltage, so it is useful for supplying stable current when the available battery voltage is low. A stable current supply requires an applied voltage that is sufficiently higher than the threshold voltage between the transistor's gate and source. This is difficult to achieve if the source voltage is low and the threshold voltage is positive, but easy if the threshold voltage is negative.

The standby-mode biases of four types of power switches are illustrated in **Fig. 12**. In each power switch, two transistors are connected in a cascode configuration. The gate biases of the nMOSFET and pMOSFET are the ground level and power-supply level, respectively. The intermediate potential (V_X) values of the power switches are chosen to ensure that the leakage currents of the cascode transistors are equal. In a type-A power switch, the gate-source voltage of the upper pMOSFET is 0 V and that of the lower pMOSFET is more than 0 V. Thus, the V_X value

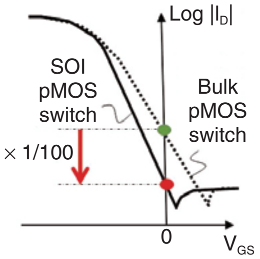


Fig. 10. Gate/source-voltage dependence of the off leakage current with a constant drain/source voltage for PMOS power switches with SOI and bulk technologies.

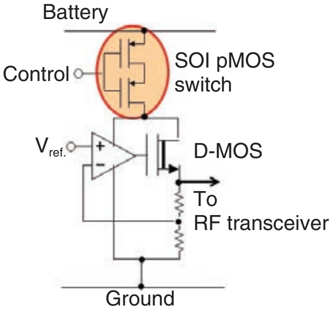


Fig. 11. Architecture of the power switch and regulator circuits with CMOS/SOI technology.

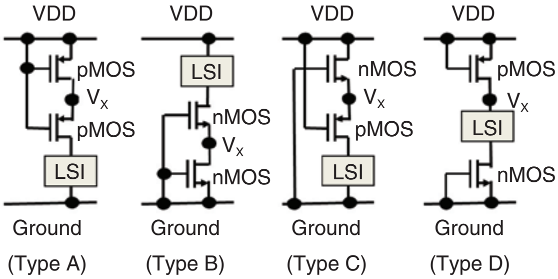


Fig. 12. Standby-mode biases of four types of power switches. V_X is the intermediate potential of the power switches.

Table 1. Intermediate potential (V_X) and gate/source voltages of the four power-switch architectures.

Type	V_X	V_{GS} (upper MOS)	V_{GS} (lower MOS)
A	< VDD/2	0	Weak reverse bias
B	> VDD/2	Weak reverse bias	0
C	~ VDD/2	Reverse bias of VDD/2	Reverse bias of VDD/2
D	~ VDD/2	0	0

exceeds half of VDD. Similarly, the gate-source voltages (V_{GS}) and V_X values of type-B, -C, and -D switches can be estimated; the values are given in **Table 1**. From the conditions and values in Table 1, the type-C power switch is expected to have the lowest leakage current among the four power-switch architectures. This is because the gates of both transistors are reversely biased (i.e., the gate-source voltage of the upper nMOSFET is less than 0 V and that of the lower pMOSFET is more than 0 V) and the V_X value is about half of VDD. Furthermore, the leakage current is expected to be aggressively reduced compared with the other power switches, as shown in **Fig. 13**.

A typical architecture of a bulk regulator circuit with the new power-switch configuration is depicted in **Fig. 14**. An enhancement-mode nMOSFET is used for a power switch between the LSI and VDD. Thus, the voltage drop* across the nMOSFET is larger than the transistor's threshold voltage because the voltage drop is equal to its gate-source voltage. This voltage drop is a big problem for low-voltage operation. Therefore, we use a DC-DC converter to raise the gate bias of the nMOSFET. The maximum gate bias of the nMOSFET is designed to be about twice the value of VDD in the active mode. On the other hand, the gate bias of the pMOSFET is fixed to the ground level. As a result, the regulator can supply sufficient current with a low battery voltage in the active mode.

Since the DC-DC converter is used only for the regulator circuit, a small converter is sufficient for this application. The small DC-DC converter operates only in the active mode. Thus, it does not increase the average power consumption of a regulator that has a low activity ratio. The leakage current in the standby mode is reduced by the new power switch in the main current path and by small switches in the regulator control paths. The gate width of the small switches is a thousand times smaller than that of the power switch in the main path. Thus, the total leakage current is not increased by the small switches in the regulator control paths.

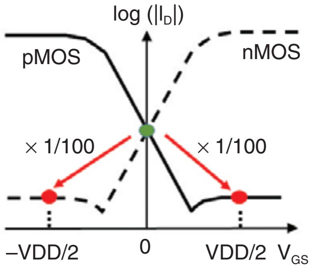


Fig. 13. Comparison of bias condition and leakage current between conventional and the new power switches.

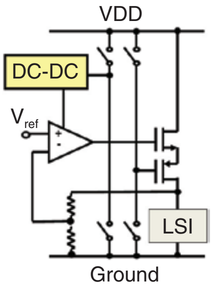


Fig. 14. Typical architecture of a bulk regulator with the new power switch. The DC-DC converter raises the gate bias of the nMOSFET.

The maximum supply current and leakage current of the new power switch were measured using a TEG (test element group) circuit. The TEG architecture is shown in **Fig. 15**. The DC-DC converter makes the gate bias of the nMOSFET twice VDD in the active mode and zero in the standby mode. The gate bias of the pMOSFET is zero in the active mode and VDD in the standby mode. Currents in main path I_1 and control path I_2 can be measured individually.

The measured current characteristics of I_1 are shown in **Fig. 16**. The horizontal axis is the voltage drop across the power switches and the vertical axis is supply current I_1. For comparison, the supply cur-

* When a DC-DC converter is not used, the maximum gate potential of the nMOSFET is VDD: the gate-source voltage and drain-source voltage (which is the same as the voltage drop produced by the nMOSFET) become equal. Moreover, during current flows, the gate-source voltage is always larger than the threshold voltage. That is why the voltage drop is larger than the threshold voltage.

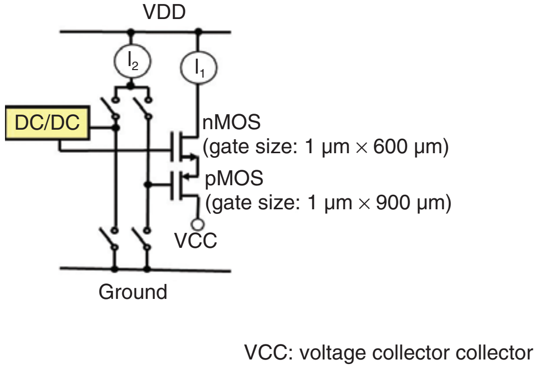

VDD
I2
I1
DC/DC
nMOS
(gate size: 1 µm × 600 µm)
pMOS
(gate size: 1 µm × 900 µm)
VCC
Ground
VCC: voltage collector collector

Fig. 15. TEG architecture of the new power switch for measuring supply and leakage currents. Currents in main path I_1 and control path I_2 can be measured individually.

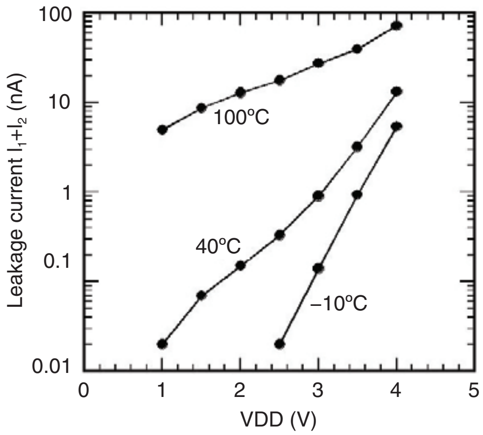


Fig. 17. Measured leakage current versus VDD characteristics in three temperature conditions. The total leakage current of I_1 and I_2 is plotted.

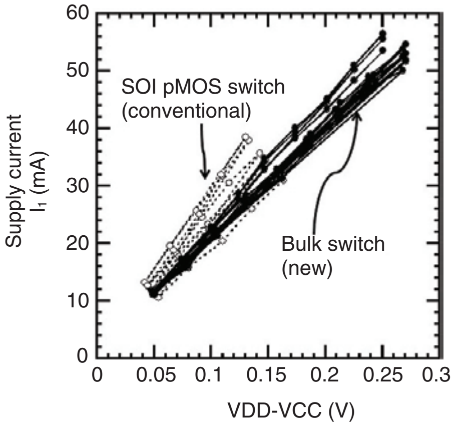


Fig. 16. Measured supply current characteristics of the main path (I_1).

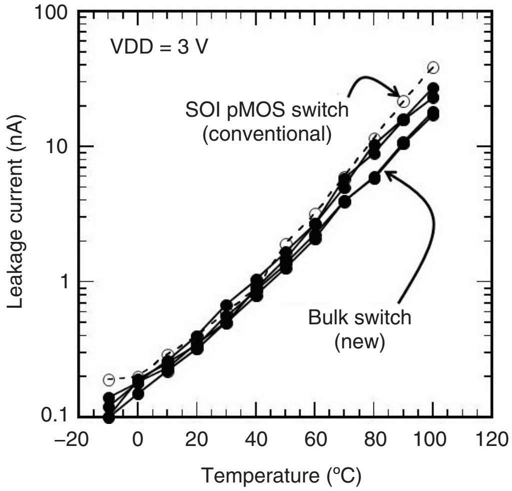


Fig. 18. Measured leakage current versus temperature characteristics. The total leakage current of I_1 and I_2 is plotted.

rent characteristics of conventional SOI pMOS switches are also shown. The gate widths of the pMOSFETs and nMOSFETs of the switches were 900 µm and 600 µm, respectively. The gate length for both was 1.0 µm. From the data, we see that the supply current of the new bulk switches is almost the same as that of the SOI switches and that the supply-current characteristics of the former are more stable than those of the latter with respect to process variation.

The leakage current versus VDD characteristics of the TEG measured at 100°C, 40°C, and -10ºC are shown in **Fig. 17**. The leakage current is the sum of the main path current I_1 and control current I_2. The maximum voltage for VDD is 4 V. The leakage current of the new bulk switch is always lower than 100 nA in the graph's temperature and voltage ranges.

The measured leakage current versus temperature characteristics of the TEG at VDD = 3 V are shown in **Fig. 18**. The open plots are measured data for the conventional SOI MOS switch and the closed ones are for the new bulk switches. The temperature dependence of the leakage current is almost the same for the new bulk switch and conventional SOI switch in this temperature range.

5. Conclusion

Power-management-circuit techniques for low-power intermittent LSI operation were described. Since they depend strongly on the wireless applications, three typical wireless applications were presented as circuit examples. These circuit techniques—an analog RC timer circuit with megaohm resistors, a two-stage power management circuit, and a new regulator circuit, individually or in combination, enable the supply of current ranging from 1 mA to 100 mA and can cut the leakage current to the sub-nanoampere level. Therefore, they are applicable to almost all low-power wireless terminals.

References

[1] H. Saito, M. Umehira, and T. Ito, "Proposal of the Wide Area Ubiquitous Network," Proc. of the World Telecom. Congress, Budapest, Hungary, 2006.

[2] H. Saito, O. Kagami, M. Umehira, and Y. Kado, "Wide Area Ubiquitous Network: The Network Operator's View of a Sensor Network," IEEE Communications Magazine, Vol. 46, No. 12, pp. 112–120, 2008.

[3] W. Weber, J. M. Rabaey, and E. Aarts, "Ambient Intelligence," Springer, 2005.

[4] H. De Man, "Ambient Intelligence: Gigascale Dreams and Nanoscale Realities," Proc. of IEEE International Solid-State Circuits Conference (ISSCC), Vol. 1, pp. 29–35, San Francisco, CA, USA, 2005.

[5] J. M. Kahn, R. H. Katz, and K. S. J. Pister, "Next Century Challenges: Mobile Networking for "Smart Dust"," Proc. of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking, pp. 271–278, Seattle, WA, USA, 1999.

[6] S. Roundy, P. K. Wright, and J. M. Rabaey, "Energy Scavenging for Wireless Sensor Networks," Kluwer Academic Publishers, p. 22, 2004.

[7] A. Shameli, A. Safarian, A. Rofougaran, M. Rofougaran, and F. De Flaviis, "An RFID System with Fully Integrated Transponder," Proc. of the 2007 IEEE Radio Frequency Integrated Circuits Symposium (RFIC), pp. 285–288, Honolulu, HI, USA, 2007.

[8] M. Polivka, M. Svanda, P. Hudec, and S. Zvanovec, "UHF RF Identification of People in Indoor and Open Areas," IEEE Transactions on Microwave Theory and Techniques, Vol. 57, No. 5, pp. 1341–1347, 2009.

[9] K. Mizuno, T. Tsubaki, H. Tsuboi, H. Nakada, A. Nakajima, A. Ikeda, and M. Shimizu, "Experimental Results of EPCglobal Phase 2 Pilot with Active RFID," Proc. of the 14th Asia-Pacific Conference on Communications (APCC), pp. 1–4, Tokyo, Japan, 2008.

[10] C.-S. Cheng, H. H. Chang, Y.-T. Chen, T. H. Lin, P. C. Chen, C. M. Huang, H. S. Yuan, and W. C. Chu, "Accurate Location Tracking Based on Active RFID for Health and Safety Monitoring," Proc. of the 3rd International Conference on Bioinformatics and Biomedical Engineering (iCBBE), Beijing, China, 2009.

[11] E. Iadanza and F. Dori, "Custom Active RFID Solution for Children Tracking and Identifying in a Resuscitation Ward," Proc. of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Minneapolis, MN, USA, pp. 5223–5236, 2009.

[12] S. Polito, D. Biondo, A. Iera, M. Mattei, and A. Molinaro, "Performance Evaluation of Active RFID Location Systems based on RF Power Measures," Proc. of the 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), pp. 1–5, Athens, Greece, 2007.

[13] M. Harada, A. Yamagishi, M. Ugajin, M. Nakamura, K. Suzuki, and Y. Kado, "Low-power Circuit Techniques for Wireless Terminals in Wide Area Ubiquitous Network," NTT Technical Review, Vol. 6, No. 3, 2008. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200803sp5.html

[14] M. Harada, A. Yamagishi, M. Ugajin, K. Fujii, and Y. Kado, "Low Power Wireless Terminals for Wide Area Ubiquitous Network," Proc. of the 2010 Asia-Pacific Radio Science Conference (AP-RASC), Paper #D1-1, Toyama, Japan, 2010.

[15] K. Suzuki, M. Ugajin, and M. Harada, "A 1-Mbps 1.6-μA Micropower Active-RFID CMOS LSI for the 300-MHz Frequency Band," Proc. of the IEEE International Microwave Symposium (MTT-S), pp. 571–574, Honolulu, HI, USA, 2007.

[16] K. Suzuki, M. Ugajin, J. Kodate, and M. Harada, "300-MHz-Frequency-band Impulse-radio Receive Architecture with All-digital Compensation for Clock Jitter and Frequency Variation," Proc. of the European Radar Conference (EuRAD), pp. 339–342, Rome, Italy, 2009.

[17] K. Suzuki, M. Ugajin, and M. Harada, "A 1-Mbps 1.6-μA Active-RFID CMOS LSI for the 300-MHz Frequency Band with an All-digital RF Transmitting Scheme," IEICE Trans. on Electronics, Vol. E94-C, No. 6, pp. 1084–1090, 2011.

[18] T. Shimamura, M. Ugajin, K. Kuwabara, K. Takagahara, K. Suzuki, H. Morimura, M. Harada, and S. Mutoh, "MEMS-switch-based Power Management with Zero-power Voltage Monitoring for Energy Accumulation Architecture on Dust-size Wireless Sensor Node," Proc. of VLSI Symposium on Circuits, pp. 276–277, Kyoto, Japan, 2011.

[19] K. Ono, N. Sato, T. Shimamura, M. Ugajin, T. Sakata, S. Mutoh, and Y. Sato, "A Millimeter-sized Electret-energy-harvester with Micro-fabricated Horizontal Arrays and Vertical Protrusions for Power Generation Enhancement," Proc. of the 16th International Conference on Solid-State Sensors, Actuators and Microsystems (IEEE Transducers'11), pp. 1863–1866, Beijing, China, 2011.

[20] T. Shimamura, M. Ugajin, K. Suzuki, K. Ono, N. Sato, K. Kuwabara, H. Morimura, and S. Mutoh, "Vibration Sensor Node Prototype with Nanowatt Circuit Techniques for Ambient Intelligence," Proc. of the 2nd Integrated MEMS Symposium 2010, Paper #IM-C-3, Shimane, Japan, 2010.

[21] T. Shimamura, M. Ugajin, K. Suzuki, K. Ono, N. Sato, K. Kuwabara, H. Morimura, and S. Mutoh, "Nano-watt Power Management and Vibration Sensing on a Dust-size Batteryless Sensor Node for Ambient Intelligence Application," Proc. of IEEE International Solid-State Circuits Conference (ISSCC), pp. 504–505, San Francisco, CA, USA, 2010.

[22] M. Ugajin, T. Shimamura, S. Mutoh, and M. Harada, "A Sub-nanoampere Two-stage Power Management Circuit in 0.35-μm CMOS for Dust-size Batteryless Sensor Nodes," Proc. of International Conference on Solid State Devices and Materials (SSDM), pp. 347–348, Tokyo, Japan, 2010.

[23] T. Shimamura, H. Morimura, M. Ugajin, and S. Mutoh, "A Zero-power Vibration-sensing Circuit for Dust-size Battery-less Sensor Nodes," Proc. of International Conference on Solid State Devices and Materials (SSDM), pp.1156-1157, Tokyo, Japan, 2010.

[24] M. Ugajin, T. Simamura, S. Mutoh, and M. Harada, "Design and Performance of a Sub-nano-ampere Two-stage Power Management Circuit in 0.35-μm CMOS for Dust-size Sensor Nodes," IEICE Trans. on Electronics, Vol. E94-C, No. 7, pp. 1206–1211, 2011.

[25] M. Ugajin, A. Yamagishi, M. Harada, and Y. Kado, "Ultra-low Leak Regulator Circuits with SOI and Bulk Technologies Controlling Intermittent LSI Operation for Wireless Terminals in Wide Area Ubiquitous Network," Proc. of the 2010 Asia-Pacific Radio Science Conference (AP-RASC), Paper #D1-2, Toyama, Japan, 2010.

[26] M. Ugajin, A. Yamagishi, K. Suzuki, and M. Harada, "Operation of Ultra-low Leakage Regulator Circuits for Controlling Wireless Transceivers," IEICE Trans. on Electronics, Vol. E94-C, No. 10, 2011 (accepted for publication).

[27] G. K. Balachandran and R. E. Barnett, "A 110 nA Voltage Regulator

System with Dynamic Bandwidth Boosting for RFID Systems," IEEE J. Solid-State Circuits, Vol. 41, No. 9, pp. 2019–2028, 2006.

[28] A. O. Adan and K. Higashi, "OFF-state Leakage Current Mechanisms in Bulk Si and SOI MOSFETs and Their Impact on CMOS ULSIs Standby Current," IEEE Trans. on Electron Devices, Vol. 48, No. 9, pp. 2050–2057, 2001.

[29] T. Ohno, Y. Kado, M. Harada, and T. Tsuchiya, "Experimental 0.25-μm-gate Fully Depleted CMOS/SIMOX Process Using a New Two-step LOCOS Isolation Technique," IEEE Trans. on Electron Devices, Vol. 42, No. 8, pp. 1481–1486, 1995.

[30] J.-W. Park, Y.-G. Kim, I.-K. Kim, K.-C. Park, K.-C. Lee, and T.-S. Jung, "Performance Characteristics of SOI DRAM for Low-power Application," IEEE Journal of Solid-State Circuits, Vol. 34, No. 11, pp. 1446–1453, 1999.

Mamoru Ugajin

Senior Research Engineer, Smart Device Laboratory, NTT Microsystem Integration Laboratories.

He received the B.S., M.S., and Ph.D. degrees in applied physics from the University of Tokyo in 1983, 1985, and 1996, respectively. He joined NTT in 1985. From 1985 to 1997, he worked on silicon-BJT and SiGe-HBT device technologies for high-speed digital applications at NTT LSI Laboratories, Atsugi. During 1992–1993, he was a visiting researcher at the University of Florida, Gainesville, where he worked on modeling and analysis of SiGe HBTs. From 1997 to 1999, he was at NTT headquarters supporting NTT's telecommunications standardization activities. Since 1999, he has been engaged in circuit design for CMOS wireless transceiver ICs. He served as a Program Committee Member of the Symposium on VLSI Circuits, an Associate Editor of the IEICE Transactions on Electronics, and an Editor of the IEICE Electronics Express. He is a member of IEEE and the Institute of Electronics, Information and Communication Engineers (IEICE).

Toshishige Shimamura

Senior Research Engineer, Smart Device Laboratory, NTT Microsystem Integration Laboratories.

He received the B.E. degree in physical electronics and the M.E. degree in advanced applied electronics from Tokyo Institute of Technology in 1995 and 1997, respectively. He joined NTT in 1997. He has been engaged in research on sensing techniques for MEMS devices stacked on CMOS LSIs (CMOS-MEMS). He is currently researching nanowatt wireless sensor nodes for ambient intelligence. He received the 2006 IEICE Best Paper Award and the Best Paper Award of the Symposium on Integrated MEMS Technology at the 26th Sensor Symposium. He is currently serving as an Associate Editor of the IEICE Transactions on Electronics. He is a member of IEICE, the Japan Society of Applied Physics (JSAP), and IEEE.

Akihiro Yamagishi

Senior Research Engineer, Third Promotion Project, NTT Access Network Service Systems Laboratories.

He received the B.E. and M.E. degrees in electrical engineering from Toyama University and the Ph.D. degree in electrical engineering from Tohoku University, Miyagi, in 1986, 1988, and 2005, respectively. In 1988, he joined NTT LSI Laboratories, where he engaged in R&D of frequency synthesizers ICs. From 2000 to 2004, he studied 1-V-operation CMOS RF circuits for wireless systems. He is currently studying wireless terminal for the WAUN. He is a member of the Institute of Image Information and Television Engineers of Japan.

Kenji Suzuki

Research Engineer, Third Promotion Project, NTT Access Network Service Systems Laboratories.

He received the B.E. and M.E. degrees in electrical and electronic engineering from Tokyo Institute of Technology in 1999 and 2001, respectively. In 2001, he joined NTT Telecommunications Energy Laboratories, where he worked on a wireless transceiver LSI architecture for low power dissipation. His interests include analog and RF IC design for wireless communications. In 2009, he moved to the Wireless Systems Innovation Laboratory, Yokosuka. He worked on the wireless terminal units of WAUN systems. In 2010, he moved to NTT Access Network Service Systems Laboratories. He is currently developing WAUN systems. He is a member of IEEE and IEICE.

Mitsuru Harada

Leader, Wireless Communication Circuits Research Group, NTT Microsystem Integration Laboratories.

In 1990, he joined NTT Atsugi Electrical Communication Laboratories, where he engaged in research on thin-film SOI devices. Since 1997, he has been researching low-power CMOS circuits for wireless terminals. At the time the research reported in this article was done, he was a Senior Research Engineer and Supervisor in the Smart Device Laboratory, NTT Microsystem Integration Laboratories.

Fast Algorithm for Monitoring Data Streams by Using Hidden Markov Models

Yasuhiro Fujiwara[†] *and Yasushi Sakurai*

Abstract

We describe a fast algorithm for exact and efficient monitoring of streaming data sequences. Our algorithm, SPIRAL-Stream, is a fast search method for finding the best model among a set of candidate hidden Markov models (HMMs) for given data streams. It is based on three ideas: (1) it clusters model states to compute approximate likelihoods, (2) it uses several granularities of clustering and approximation level of likelihood values in search processing, and (3) it focuses on the efficient computation of only promising likelihoods by pruning out low-likelihood state sequences. Experiments verified its effectiveness and showed that it was more than 490 times faster than the naive method.

1. Introduction

Significant applications that use hidden Markov models (HMMs), including traffic monitoring and traffic anomaly detection, have emerged. The goal of this study is efficient monitoring of streaming data sequences by finding the best model in an exact way. Although numerous studies have been published in various research areas, this is, to the best of our knowledge, the first study to address the HMM search problem in a way that guarantees the exactness of the answer.

1.1 Problem definition

Increasing the speed of computing HMM likelihoods remains a major goal for the speech recognition community. This is because most of the total processing time (30–70%) in speech recognition is used to compute the likelihoods of continuous density HMMs. Replacing continuous density HMMs by discrete HMMs is a useful approach to reducing the computation cost [1]. Unfortunately, the central processing unit cost still remains excessive, especially for large datasets, since all possible likelihoods are computed.

Recently, the focus of data engineering has shifted toward data stream applications [2]. These applications handle continuous streams of input from external sources such as a sensor. Therefore, we address the following problem in this article:

Problem *Given an HMM set and a subsequence of data stream* $X = (x_1, x_2, \cdots, x_n)$*, where* x_n *is the most recent value, identify the model whose state sequences have the highest likelihood, estimated with respect to X, among the set of HMMs* by monitoring the incoming data stream.

Key examples of this problem are traffic monitoring [3], [4] and anomaly detection [5], [6].

1.2 New contribution

We have previously proposed a method called SPIRAL that offers fast likelihood searches for static sequences [7]. In this article[*], it is extended to data streams; this extended approach, called SPIRAL-Stream, finds the best model for data streams [8]. To reduce the search cost, we (1) reduce the number of states by clustering multiple states into clusters to compute the approximate likelihood, (2) compute the

[†] NTT Cyber Space Laboratories
Yokosuka-shi, 239-0847 Japan

[*] The present article is basically a shortened version of ref. [8]. Readers who would like more details, should refer to ref. [8].

approximate likelihood with several levels of granularity, and (3) prune low-likelihood state sequences that will not yield the best model. SPIRAL-Stream has the following attractive characteristics based on the above ideas:

- High-speed searching: Solutions based on the Viterbi algorithm are prohibitively expensive for large HMM datasets. SPIRAL-Stream uses carefully designed approximations to efficiently identify the most likely model.
- Exactness: SPIRAL-Stream does not sacrifice accuracy; it returns the highest likelihood model without any omissions.

To achieve high performance and find the exact answer, SPIRAL-Stream first prunes many models by using approximate likelihoods at a low computation cost. The exact likelihood computations are performed only if absolutely necessary, which yields a drastic reduction in the total search cost.

The remainder of this article is organized as follows. Section 2 overviews some background of HMMs. Section 3 introduces SPIRAL-Stream and shows how it identifies the best model for data streams. Section 4 presents the results of our experiments. Section 5 is a brief conclusion.

2. HMMs

In this section, we explain the basic theory of HMMs.

2.1 Definitions

Unlike the regular Markov model, in which each state corresponds to an observable event, an HMM is used when there is a set of unobserved, thus hidden, states and the observation is a probabilistic function of the state. Let $\{u_i\}$ (i = 1, ···, m) be a set of states. An HMM is composed of the following probabilities:

- Initial state probability $\pi = \{\pi_i\}$: The probability of the state being u_i (i = 1, ···, m) at time t.
- State transition probability a = $\{a_{ij}\}$: The probability of the state transiting from state u_i to u_j.
- Symbol probability b(v) = $\{b(v)\}$: The probability of symbol v being output from state u_i.

HMMs are classified by the structure of the transition probability. Ergodic HMMs, or fully connected HMMs, have the property that every state can be reached from every other state. Another type of HMM is the left-right HMM; its state transitions have the property that, as time increases, the state number increases or stays the same.

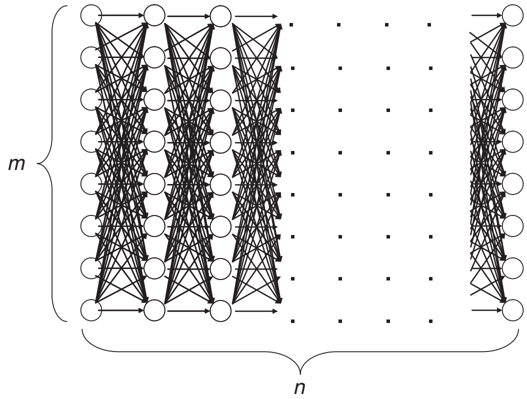


Fig. 1. Trellis structure

2.2 Viterbi algorithm

The well-known Viterbi algorithm is a dynamic programming algorithm for HMMs that identifies the most-likely state sequence with the maximum probability and estimates its likelihood given an observed sequence. The state sequence, which gives the likelihood, is called the Viterbi path. For a given model, the likelihood P of X is computed as follows:

$$P = max_{1 \leq i \leq m} (p_{in}), \text{ where}$$

$$p_{it} = \begin{cases} max_{1 \leq i \leq m} (p_{j(t-1)} a_{ji}) b_i(x_t) & (2 \leq t \leq n) \\ \pi_i b_i(x_1) & (t = 1) \end{cases},$$

where m is the length of the sequence, n is the number of states, and p_{it} is the maximum probability of state u_i at time t. The likelihood is computed on the basis of the trellis structure shown in **Fig. 1**, where states lie on the vertical axis and sequences are aligned along the horizontal axis. The likelihood is computed using the dynamic programming approach that maximizes the probabilities from previous states (i.e., each state probability is computed using all previous state probabilities, associated transition probabilities, and symbol probabilities).

The Viterbi algorithm generally needs O(nm^2) time since it compares m transitions to obtain the maximum probability for every state; that is, it requires O(m^2) in each time tick. The naive approach to monitoring data streams is to perform this procedure each time a sequence value arrives. However, considering the high frequency with which new values will arrive, more efficient algorithms are needed.

3. Finding the best model for data streams: SPIRAL-Stream

In this section, we discuss how to handle data streams.

3.1 Ideas behind SPIRAL-Stream

Our solution is based on three ideas: likelihood approximation, multiple granularities, and transition pruning. These are outlined below in this subsection and explained in more detail in subsections 3.2–3.4.

(1) Likelihood approximation

In a naive approach to finding the best model among a set of candidate HMMs, the Viterbi algorithm would have to be applied to all the models, but the algorithm's cost would be too high when it is applied to the entire set of HMMs. Therefore, we introduce approximations to reduce the high cost of the Viterbi algorithm solution. Instead of computing the exact likelihood of a model, we approximate the likelihood; thus, low-likelihood models are efficiently pruned.

The first idea is to reduce the model size. For given m states and granularity g, we create m/g states by merging *similar* states in the model (**Fig. 2(a)**), which requires $O(nm^2/g^2)$ time to obtain the approximate likelihoods instead of the $O(nm^2)$ time demanded by the Viterbi algorithm solution. We use a clustering approach to find groups of similar states and then create a compact model that covers the groups. We refer to it as the *degenerate* model.

(2) Multiple granularities

Instead of creating degenerate models at just one granularity, we use multiple granularities to optimize the tradeoff between accuracy and comparison speed. As the size of a model increases, its accuracy improves (i.e., the upper bounding likelihood decreases), but the likelihood computation time increases. Therefore, we generate models at granularity levels that form a geometric progression: $g = 1,2,4,\cdots, m$, where $g = 1$ gives the exact likelihood while $g = m$ means the coarsest approximation. We then start from the coarsest model and gradually increase the size of the models to prune unlikely models; this improves the accuracy of the approximate likelihood as the search progresses (**Fig. 2(b)**).

(3) Transition pruning

Although our approximation technique can discard unlikely models, we still rely on exact likelihood computation to guarantee the correctness of the search results. Here, we focus on reducing the cost of this computation.

The Viterbi path shows the state sequence from which the likelihood is computed. Even though the Viterbi algorithm does not compute the complete set of paths, the trellis structure includes an exponential number of paths. Clearly, exhaustive exploration of all paths is not computationally feasible, especially for a large number of states. Therefore, we ask the question: Which paths in the structure are not promising to explore? This can be answered by using a threshold (say θ).

Our search algorithm that identifies the best model maintains the candidate (i.e., best-so-far) likelihood before reporting the final likelihood. Here, we use θ as the best-so-far highest likelihood. θ is updated, i.e., increased, when a more promising model is found during search processing. Note that we assume that no two models have exactly the same likelihoods.

We exclude the unlikely paths in the trellis structure by using θ, since θ never decreases during search processing. If the upper bounding likelihood of paths that pass through a state is less than θ, that state cannot be contained in the Viterbi path, and we can safely discard these paths, as shown in **Fig. 2(c)**.

3.2 Likelihood approximation

Our first idea involves clustering states of the original models and computing upper bounding likelihoods to achieve reliable model pruning.

3.2.1 State clustering

We reduce the size of the trellis structure by merging similar states in order to compute likelihoods at low computation cost. To achieve this, we use a clustering approach. Given granularity g, we try to find m/g clusters from among the m original states. First, we describe how to compute the probabilities of a new degenerate model; then, we describe our clustering method.

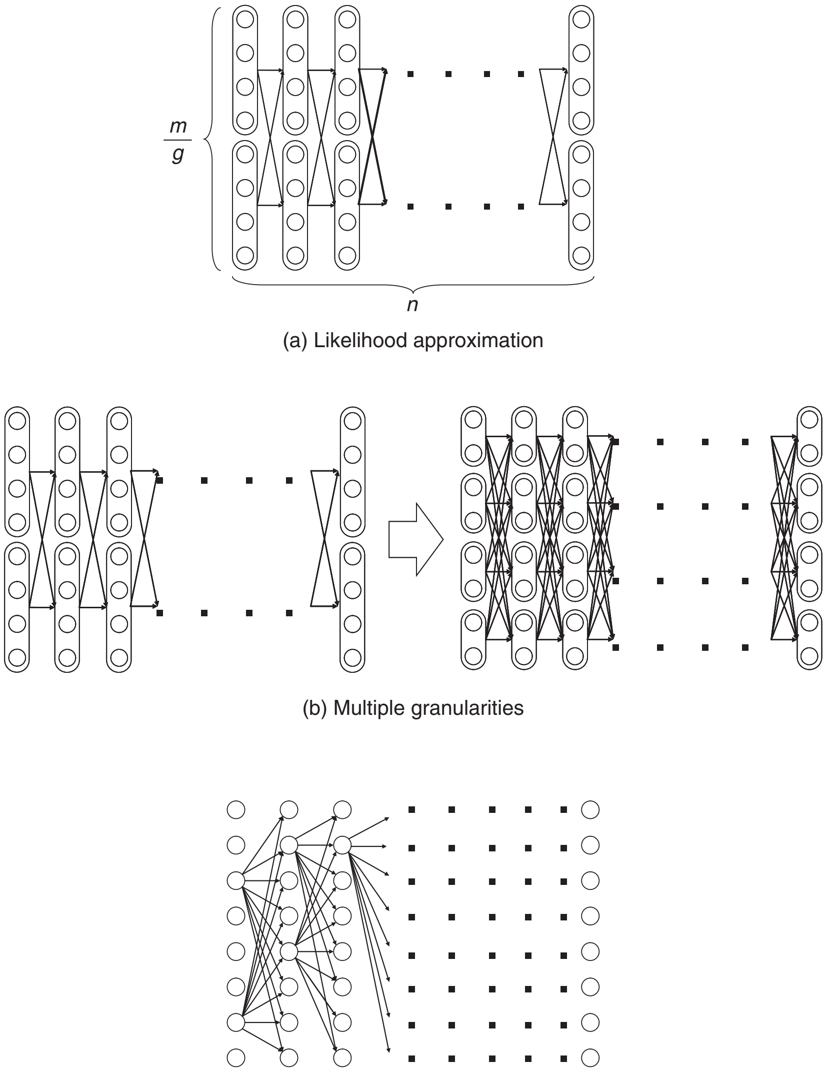
We merge all the states in a cluster and create a new state. For the new state, we choose the highest probability among the probabilities of the states to compute the upper bounding likelihood (described in subsection 3.2.2). We obtain the probabilities of new state u_c by merging all the states in cluster C as follows.

$$\hat{\pi}_C = \max_{u_i \in C}(\pi_i),\ \hat{a}_{CD} = \max_{u_i \in C, u_k \in D}(a_{ik}),\ \hat{b}_C(v) = \max_{u_i \in C}(b_i(v))$$

We use the following vector of features F_i to cluster state u_i.

$F_i = (\pi_i; a_{i1},\cdots, a_{im};\ a_{1i}, \cdots, a_{mi}; b_i(v_1),\cdots, b_i(v_s))$,

where s is the number of symbols. We choose this vector to reduce the approximation error. The highest probabilities are the probabilities of a new state.



(a) Likelihood approximation

(b) Multiple granularities

(c) Transition pruning

Fig. 2. Basic ideas behind SPIRAL.

Therefore, the greater the difference in probabilities possessed by the two states, the greater the difference in the vectors becomes. Thus, a good clustering arrangement can be found by using this vector.

In our experiments, we used the well-known k-means method to cluster states where the Euclidean distance is used as a distance measure. However, we could exploit BIRCH [9] instead of the k-means method, the L1 distance as a distance measure, or singular value decomposition to reduce the dimensionality of the vector of features. The clustering method is completely independent of SPIRAL-Stream and is beyond the scope of this article.

3.2.2 Upper bounding likelihood

We compute approximate likelihood $\hat{P}$ from degenerate models that have $\hat{m}(=m/g)$states. Given a degenerate model, we compute its approximate likelihood as follows:

$\hat{P} = max_{1\leq c\leq \hat{m}}\ (\hat{p}_{cn})$, where

$$\hat{p}_{cn} = \begin{cases} max_{1\leq j\leq \hat{m}}\ (\hat{p}_{j(t-1)}\hat{a}_{jc})\hat{b}_c(x_t) & (2\leq t\leq n) \\ \hat{\pi}_c\hat{b}_c(x_1) & (t=1) \end{cases},$$

where $\hat{p}$ is the maximum probability of states.

Theorem 1 *For any HMM model, $P \leq \hat{p}$ holds.*

Proof *Omitted owing to space limitations.*

Theorem 1 provides SPIRAL-Stream with the property of finding the exact answer [8].

3.3 Multiple granularities

The algorithm presented in subsection 3.2 uses a single level of granularity to compute the approximate likelihood of a degenerate model. However, we can also exploit multiple granularities. Here, we describe the gradual refinement of the likelihood approximation with multiple granularities. In this subsection, we first describe the definition of data streams and then our approach for data streams.

In data stream processing, the time interval of interest is generally called the *window* and there are three temporal spans for which the values of data streams need to be calculated [10]:

- **Landmark window model**: In this temporal span, data streams are computed on the basis of the values between a specific time point, called the landmark, and the present.
- **Sliding window model**: Given sliding window length n and the current time point, the sliding window model computes the subsequence from the prior $n-1$ time to the current time.
- **Damped window model**: In this model, recent data values are more important than earlier ones. That is, in a damped window model, the weights applied to data decrease exponentially into the past.

This article focuses on the sliding window model because it is used most often and is the most general model. We consider a data stream as a time-ordered series of tuples (time point, value). Each stream has a new value available at each time interval, e.g., every second. We assume that the most recent sample is always taken at time n. Hence, a streaming sequence takes the form $(\cdots, x_1, x_2, \cdots, x_n)$. Likelihoods are computed only with n values from the streaming sequence, so we are only interested in subsequences of the streaming sequence from x_1 to x_n.

For data streams, we use $h+1$ distinct granularities that form a geometric progression $g_i = 2^i$ $(i = 0,1,2, \cdots, h)$. Therefore, we generate trellis structures of models that have $\lfloor m/g_i \rfloor$ states. Here, g_h represents the smallest (coarsest) model while g_0 corresponds to the original model, which gives the exact likelihood. In the previous our study for static sequences [7], we first compute the coarsest structure for all models. We then obtain the candidate and the exact likelihood θ. If a model has an approximate likelihood smaller than θ, that model is pruned with no further computation. Otherwise, we compute a finer-grained structure for that model and check whether the approximate likelihood is smaller than θ. We iterate this check until we reach g_0.

For data streams, model granularity can be more efficiently decided by referring to the immediately prior granularity for data streams. It is reasonable to expect that the likelihood of the model examined for the subsequence will change little and that we can prune the models efficiently by continuing to use the prior granularities. That is, in the present time tick, the initial granularity is set relative to the finest granularity in the previous time tick at which the model likelihood was computed. If model pruning was conducted at the coarsest granularity, we use this granularity in the next time tick; otherwise, we use the granularity level that is one step down (coarser) as the initial granularity. If the model is not pruned at the initial granularity, the approximate likelihood of a finer-grained structure is computed to check for model pruning against the given θ.

Example *If the original HMM has 16 states and the model was pruned with the 1-state model (granularity g_4, coarsest), we choose to use the 1-state model (granularity g_4) as the initial model in the next time tick; if the model is pruned using the approximate likelihood of 16 states (granularity g_0), we select the 8-state model (granularity g_1) as the initial model.*

3.4 Transition pruning

We introduce an algorithm for computing likelihoods efficiently on the basis of the following theorem:

Lemma 1 *Likelihoods of a state sequence are monotonically nonincreasing with respect to X in the trellis structure.*

Proof *Omitted owing to space limitations.*

We exploit the above lemma in pruning paths in the trellis structure. We introduce e_{it}, which indicates a conservative estimate of likelihood p_{it}, to prune unlikely paths as follows:

$$e_{it} = \begin{cases} p_{it} = (a_{max})^{n-t} \prod_{j=t+1}^{n} b_{max}(x_j) & (1 \leq t \leq n-1) \\ p_{in} & (t = n) \end{cases},$$

where a_{max} and $b_{max}(v)$ are the maximum values of the state transition probability and symbol probability, respectively:

$a_{max} = max_{i,j}(a_{ij})$, $b_{max}(v) = max_i b_i(v)$, $(i = 1, \cdots, m; j = 1, \cdots, m)$.

The estimate is exactly the same as the maximum probability of u_i when $t = n$. Estimate e_{it}, the product of the series of the maximum values of the state transition probability and symbol probability, has the upper bounding property assuming that the Viterbi path passes through u_i at time t.

Theorem 2 *For paths that pass through state $u_i (i = 1, \cdots, m)$ at time $t (1 \leq t \leq n)$, $p_{jn} \leq e_{it}$ holds for any state*

$u_j (j = 1, \cdots, m)$ at time n.

Proof *Omitted owing to space limitations.*

This property enables SPIRAL-Stream to search for models exactly.

In search processing, if e_{it} gives a value smaller than θ (i.e., the best-so-far highest likelihood in the search processing for the best model), state u_i at time t for the model cannot be contained in the Viterbi path. Accordingly, unlikely paths can be pruned with safety.

3.5 Search algorithm

Our approach to data stream processing is shown in **Fig. 3**. Here, M_i represents the set of models for which we compute the likelihood of granularity g_i, and M'_i represents the set of models computed with the finest granularity in the previous time tick, g_i. SPIRAL-Stream first computes the initial value of θ on the basis of the best model at the last time; it then sets the initial granularity. If a model is pruned at the coarsest granularity at the last time tick, SPIRAL-Stream uses this granularity as the initial granularity. Therefore, we add M'_h to M_h in this algorithm. The one-step-lower granularity is used as the initial granularity if the model was not pruned at the coarsest granularity; this procedure is expressed by "add M'_{i-1} to M'_i".

```
Algorithm       Monitoring
Input: subsequence $X$ of stream, set of models $M'$, the
    previous best model $M'_{best}$.
Output: the best model $M_{best}$.
compute $P_0$ for $M'_{best}$;
$\theta := P_0$;
$M_{best} := M'_{best}$;
add $M'_h$ to $M_h$;
for $i := h$ to 1 do
    add $M'_{i-1}$ to $M_i$;
end for
for $i := h$ to 0 do
    $\theta' := 0$;
    for each model $M \in M_i$ do
        compute $P_i$ for $M$:
        if $P_i \geq \theta'$ then
            $M_{max} := M$;
            $\theta' := P_i$
        end if
    end for
    compute $P_0$ for $M_{max}$;
    if $P_0 \geq \theta$ then
        $M_{best} := M_{max}$;
        $\theta = P_0$;
    end if
    for each model $M \in M_i$ do
        if $P_i \geq$ then
            add $M$ to $M_{i-1}$;
            subtract $M$ from $M_i$;
        end if
    end for
    $M'_i := M_i$;
end for
$M'_{best} := M_{best}$ ;
return $M_{best}$;
```

Fig. 3. Search algorithm.

3.6 Theoretical analysis

In this subsection, we provide a theoretical analysis that shows the accuracy and complexity of SPIRAL-Stream.

Theorem 2 *SPIRAL-Stream guarantees the exact answer when identifying the model whose state sequence has the highest likelihood.*

Proof *Let M_{best} be the best model in the dataset and θ_{max} be the exact likelihood of M_{best} (i.e., θ_{max} is the highest likelihood). Moreover, let P_i be the likelihood of model M for granularity g_i and θ be the best-so-far (highest) likelihood in the search process. From Theorems 1 and 2, we obtain $P_0 \leq P_i$, for any granularity g_i, for any M. For M_{best}, $\theta_{max} \leq P_i$ holds. In the search process, since θ is monotonically nondecreasing and $\theta_{max} \geq \theta$, the approximate likelihood of M_{best} is never lower than θ, where θ is monotonically nondecreasing. The algorithm discards M if (and only if) $\theta > P_i$. Therefore, the best model M_{best} cannot be pruned erroneously during the search process.*

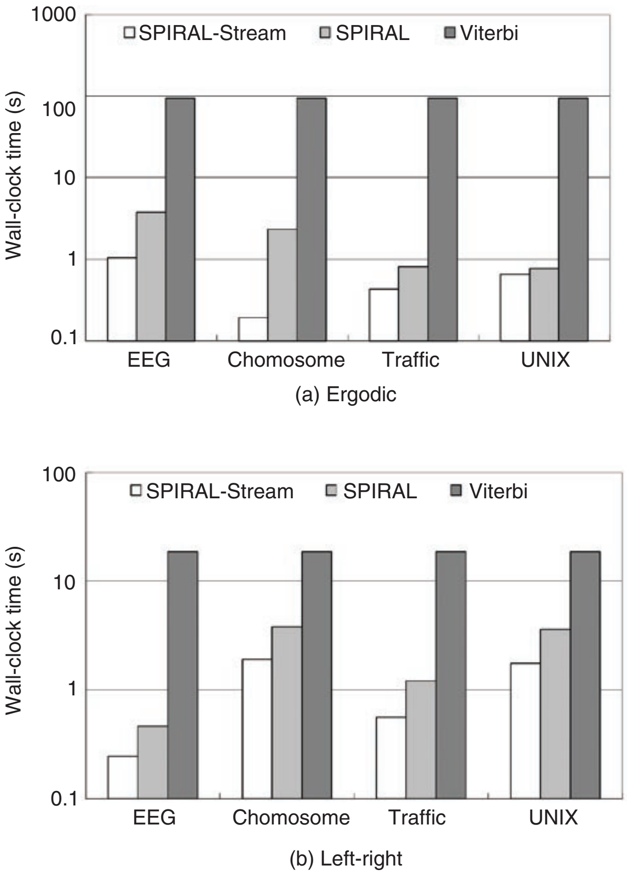
4. Experimental evaluation

We performed experiments to test SPIRAL-Stream's effectiveness. We compared SPIRAL-Stream [8] with the Viterbi algorithm, which we refer to as *Viterbi* hereinafter, and SPIRAL [7], which is our previous approach for static sequences.

4.1 Experimental data and environment

We used four standard datasets in the experiments.

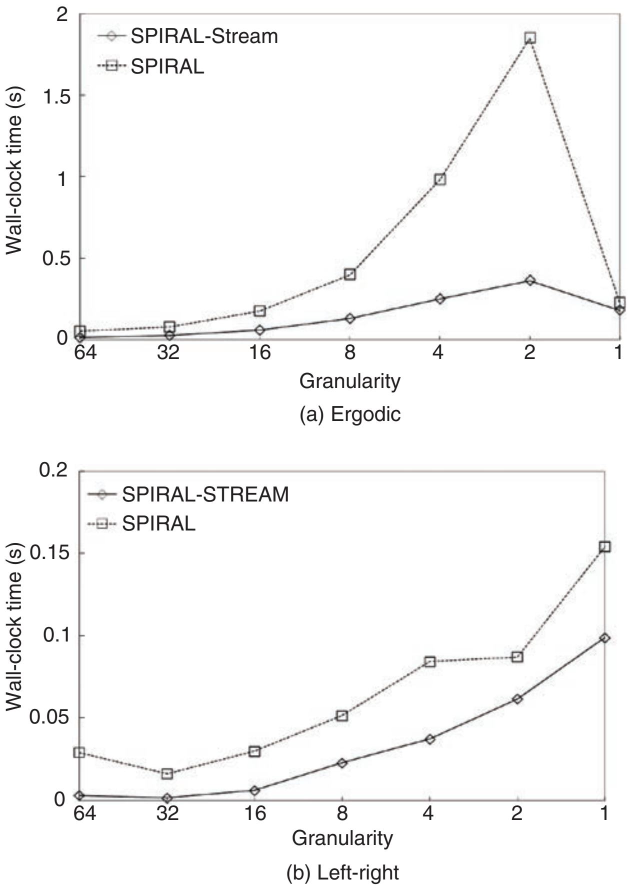
- EEG: This dataset was taken from a large electroencephalography (EEG) study that examined the EEG correlates of genetic predisposition to alcoholism. In our experiments, we quantized EEG values in 1-μV steps, which resulted in 506 elements.
- Chromosome: We used DNA (deoxyribonucleic acid) strings of human chromosomes 2, 18, 21, and 22. DNA strings are composed of the four letters of the genetic code: A, C, G, and T; however, here we use an additional letter N to denote an unknown letter. Thus, the number of symbols (symbol size) is 5.
- Traffic: This dataset contains loop sensor



(a) Ergodic

(b) Left-right

Fig. 4. Wall-clock time for monitoring data stream.



(a) Ergodic

(b) Left-right

Fig. 5. Breakdown of search cost.

measurements of the Freeway Performance Measurement System. The symbol size is 91.

- UNIX: We exploited the command histories of 8 UNIX computer users at a university over a two-year period. The symbol size is 2360.

The models were trained by the Baum-Welch algorithm [11]. In our experiments, the sequence length was 256 and possible transitions of a left-right model were restricted to only two states, which is typical in many applications.

We evaluated the search performance mainly by measuring the duration using a wall clock. All experiments were conducted on a Linux quad 3.33 GHz Intel Xeon server with 32 GB of main memory. We implemented our algorithms using the GCC compiler. Each result reported here is the average of 100 trials.

4.2 Results of data stream monitoring

We conducted several experiments to test the effectiveness of our approach for monitoring data streams.

4.2.1 Search cost

In **Figs. 4(a)** and **(b)**, SPIRAL-Stream is compared with Viterbi and with the state-of-the-art approach for data sequences, SPIRAL, which finds the best model for static sequences, in terms of the wall-clock time for various datasets where the number of states and number of models are 100 and 10,000, respectively. As expected, SPIRAL-Stream outperformed the other two algorithms: In particular, SPIRAL-Stream could find the best model up to 490 times faster than the Viterbi algorithm.

4.2.2 Effectiveness of the data stream algorithm

Our stream algorithm (SPIRAL-Stream) automatically changes the granularity and effectively sets the initial candidate to find the best model. To determine the effectiveness of these ideas, we plotted the time at each granularity for SPIRAL-Stream and SPIRAL. The results of time versus granularity in the model search cost for 10,000 models for EEG, where each model has 100 states, are shown in **Figs. 5(a)** and **(b)**.

SPIRAL-Stream required less computation time at each granularity. Instead of using g_h (the coarsest) as the initial granularity for all models, this algorithm sets the initial granularity with the finest granularity at the prior time tick, which ensures that the algorithm reduces the number of models at each granularity. Furthermore, the stream algorithm sets the best model at the prior time tick as the initial candidate, which is expected to remain the answer. As a result, it can find the best model for data streams much more efficiently.

5. Conclusion

This article addressed the problem of conducting a likelihood search on a large set of HMMs with the goal of finding the best model for a given query sequence and for data streams. Our algorithm, SPIRAL-Stream, is based on three ideas: (1) it prunes low-likelihood models in the HMM dataset by their approximate likelihoods, which yields promising candidates in an efficient manner; (2) it varies the approximation granularity for each model to maintain a balance between computation time and approximation quality; and (3) it focuses on the efficient computation of only promising likelihoods by pruning out low-likelihood state sequences. Our experiments confirmed that SPIRAL-STREAM worked as expected and quickly found high-likelihood HMMs. Specifically, it was significantly faster (more than 490 times) than the naive implementation.

References

[1] S. Sagayama, K. M. Knill, and S. Takahashi, "On the Use of Scalar Quantization for Fast HMM Computation," Proc. of ICASSP, Vol. 1, pp. 213–216, Detroit, MI, USA, 1995.

[2] D. J. Abadi, D. Carney, U. Cetintemel, M. Cherniack, C. Convey, S. Lee, M. Stonebraker, N. Tatbul, and S. B. Zdonik, "Aurora: A New Model and Architecture for Data Stream Management," Journal of VLDB, Vol. 12, No. 2, pp. 120–139, 2003.

[3] P. Bickel, C. Chen, J. Kwon, J. Rice, P. Varaiya, J. R. Pravin, and E. V. Zwet, "Traffic Flow on a Freeway Network," In Workshop on Nonlinear Estimation and Classification, 2001.

[4] J. Kwon and K. Murphy, "Modeling Freeway Traffic with Coupled HMMs," Tech. Rep., University of California at Berkeley, 2000.

[5] T. Lane, "Hidden Markov Models for Human/Computer Interface Modeling," Proc. of the IJCAI-99 Workshop on Learning About Users, pp. 35–44, 1999.

[6] C. Warrender, S. Forrest, and B. A. Pearlmutter, "Detecting Intrusions Using System Calls: Alternative Data Models," Proc. of the 1999 IEEE Symposium on Security and Privacy, pp. 133–145, Oakland, CA, USA.

[7] Y. Fujiwara, Y. Sakurai, and M. Yamamuro, "SPIRAL: Efficient and Exact Model Identification for Hidden Markov Models," Proc. of KDD'08, pp. 247–255, Las Vegas, NV, USA, 2008.

[8] Y. Fujiwara, Y. Sakurai, and M. Kitsuregawa, "Fast Likelihood Search for Hidden Markov Models," ACM Trans. on Knowledge Discovery from Data (TKDD), Vol. 3, No. 4, 2009.

[9] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: An Efficient Data Clustering Method for Very Large Databases," Proc. of SIGMOD Conference, pp. 103–114, Montreal, Quebec, Canada, 1996.

[10] V. Ganti, J. Gehrke, and R. Ramakrishnan, "DEMON: Mining and Monitoring Evolving Data," IEEE Trans. Knowl. Data Eng., Vol. 13, No. 1, pp. 50–63, 2001.

[11] S. E. Levinson, L. R. Rabiner, and M. M. Sondhi, "An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition," Bell Syst. Tech. J, Vol. 62, No. 4, pp. 1035–1074, 1982.

Yasuhiro Fujiwara

Researcher, NTT Cyber Space Laboratories.

He received the B.E. and M.E. degrees from Waseda University, Tokyo, in 2001 and 2003, respectively, and the Ph.D. degree from the University of Tokyo in 2012. He joined NTT Cyber Solutions Laboratories in 2003. His research interests include data mining, databases, natural language processing, and artificial intelligence. He received two KDD best paper awards in 2008, IPSJ Best Paper Awards in 2008, IEICE Best Paper Award in 2008, and DASFAA Best Paper Award in 2012. He is a member of the Institute of Physical Society of Japan (IPSJ), Institute of Electronics, Information and Communication Engineers (IEICE), and Database Society of Japan (DBSJ).

Yasushi Sakurai

Senior Research Scientist, NTT Communication Science Laboratories.

He received the B.E. degree from Doshisha University, Kyoto, in 1991 and the M.E. and Ph. D. degrees from Nara Institute of Science and Technology in 1996 and 1999, respectively. He joined NTT Cyber Space Laboratories in 1998. He was a visiting researcher at Carnegie Mellon University, Pittsburgh, PA, USA, during 2004–2005. Since 2007, he has been a senior researcher at NTT Communication Science Laboratories. He received the IPSJ Nagao Special Researcher Award (2007), DBSJ Kambayashi Incentive Award (Young Scientist Award, 2007), and twelve best paper awards, including two KDD best research paper awards (2008 and 2010), IPSJ best paper awards (2004 and 2008), and IEICE Best Paper Award (2008). His research interests include indexing, data mining, and data stream processing.

IEC TC86 WG4 Standardization Activities Aimed at Developing International Standards from Domestic Standards

Noriyuki Araki[†]

Abstract

This article describes the progress in the International Electrotechnical Commission (IEC) on global standardization of a method for calibrating optical measurement equipment and its relation to a Japanese domestic standard.

1. Introduction

With the rapid spread of fiber to the home (FTTH) in recent years, the importance of optical testing technology for the construction and maintenance of optical fiber cable networks has been increasing. In terms of evaluating the characteristics of optical fiber cable networks and optical devices, the standardization of the calibration method plays an important role in guaranteeing the performance of optical measurement equipment.

2. IEC

Standardization activities related to fiber optic measurement equipment calibration and its procedures are the concern of the International Electrotechnical Commission (IEC) [1]. The IEC is a leading global organization that prepares and publishes international standards for all electrical, electronic, and related technologies. These serve as a basis for national standardization and as references when international tenders and contracts are drafted. The IEC was founded in 1906. As of March 2011, 81 countries were registered as members or associate members of the IEC.

Standards are discussed by Technical Committees (TCs). Subcommittees (SCs) and Working Groups (WGs), which work under the supervision of a TC, are established as necessary. IEC TC86 WG4 (Fiber optic test equipment calibration) was set up in 1985 by TC86 (Fiber optics) to discuss standards related to calibration methods and procedures for optical measurement equipment.

The Japanese National Committee of TC86 is organized by the Institute of Electronics, Information and Communication Engineers (IEICE), which has a strong relationship with the Standardization Technical Committee of Japanese Industrial Standards (JIS), which is organized by the Optoelectronic Industry and Technology Development Association (OITDA). After policy on how to deal with issues has been approved by the national committee in Japan, members of the Japanese National Committee of TC86 WG4 participate in discussions at IEC TC86 WG4 meetings. The relationships among these committees is shown in **Fig. 1**.

3. Standards development process in IEC

An International Standard (IS) results from agreement among the IEC's National Committees. The technical work is developed through project stages. The sequence of project stages and the name and

[†] NTT Access Network Service Systems Laboratories
Tsukuba-shi, 305-0805 Japan

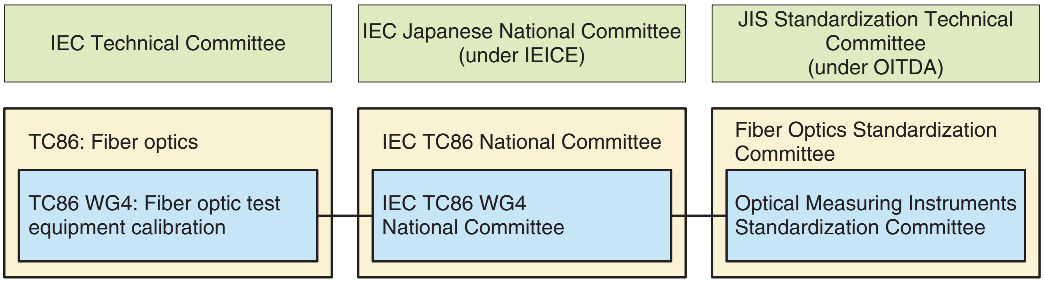

IEC Technical Committee
IEC Japanese National Committee (under IEICE)
JIS Standardization Technical Committee (under OITDA)
TC86: Fiber optics
TC86 WG4: Fiber optic test equipment calibration
IEC TC86 National Committee
IEC TC86 WG4 National Committee
Fiber Optics Standardization Committee
Optical Measuring Instruments Standardization Committee

Fig. 1. Relationships among IEC Technical Committee, IEC Japanese National Committee, and JIS Standardization Technical Committee related to IEC TC86 WG4.

Table 1. Project stages and associated documents.

Project stage	Associated document	
	Name	Abbreviation
Preliminary stage	Preliminary Work Item	PWI
Proposal stage	New Work Item Proposal	NP
Preparatory stage	Working Draft(s)	WD
Committee stage	Committee Draft(s)	CD
Enquiry stage	Committee Draft for Vote	CDV
Approval stage	Final Draft International Standard	FDIS
Publication stage	International Standard	IS

abbreviation of the document associated with each stage are listed in **Table 1**. The associated documents are prepared during the corresponding stages and submitted as comments and voted on by the participating member countries (P-members) within the TC or SC. They then proceed to the publication stage. For each stage, there is a commenting or voting period of two to five months, and it might be several years before an IS is issued.

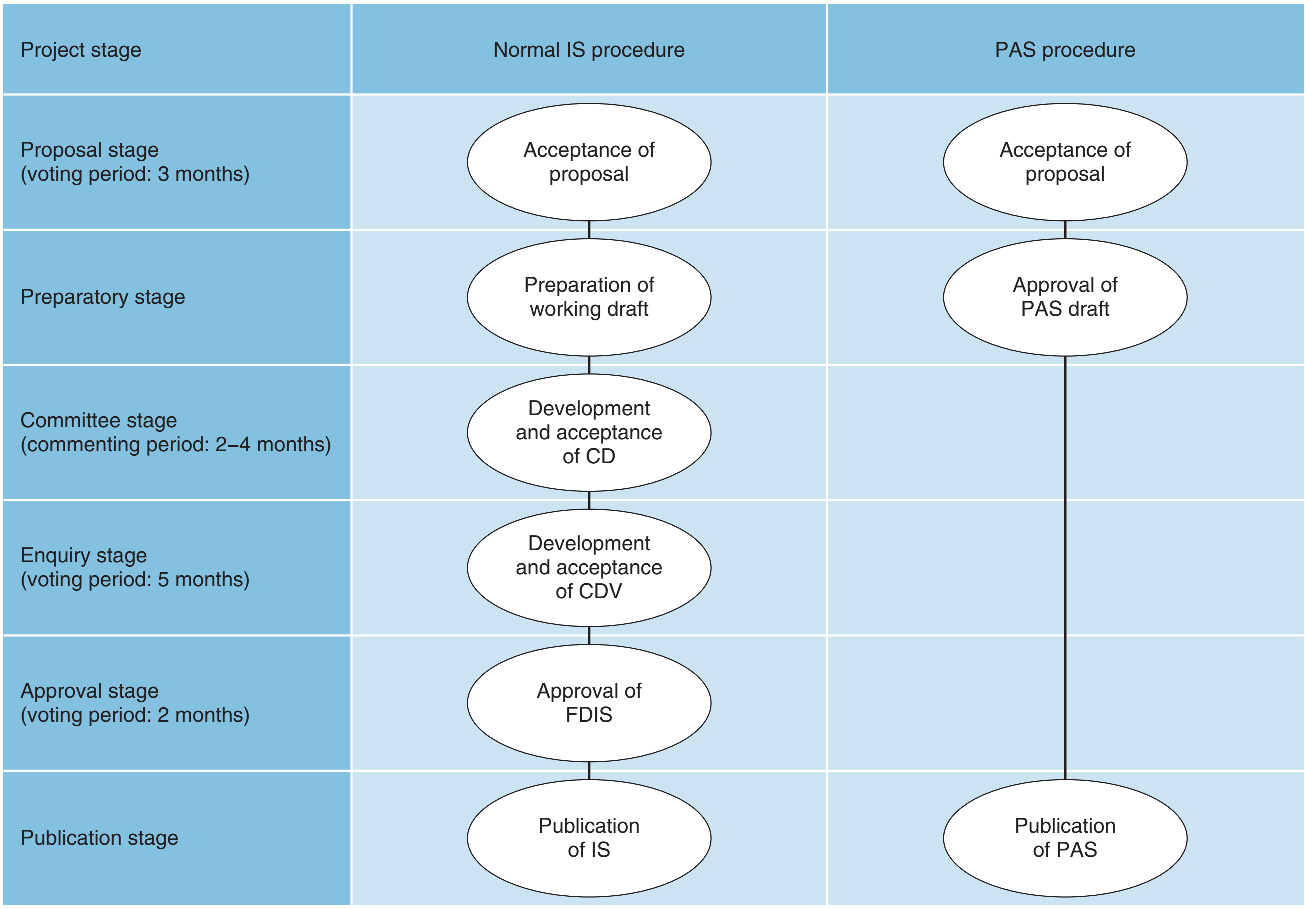
For example, in the proposal stage, a proposal for new work generally originates from industry via a National Committee. It is then communicated to the members of the appropriate TC or SC. A New Work Item Proposal (NP) is approved after a commenting and voting period of three months if:

- the committee's P-members approve it by a simple majority and
- if enough experts nominated by the P-members approve it. For committees with 16 or fewer P-members, the minimum number of experts is four; for committees with 17 or more P-members, it is five.

The proposal is rejected if it does not satisfy the above requirements.

4. Development procedure of IEC-PAS

To enable a Publicly Available Specification (PAS) [2] originally developed by a consortium or forum with a common interest to be accepted by the IEC, the IEC has a structure called the IEC-PAS [3]. An IEC-PAS is either published by the IEC after organizations outside the IEC reach a consensus or published by the experts in an IEC WG. It may also be published to meet urgent market needs. It is published after the committee concerned has confirmed its content and checked that there are no conflicts with existing ISs and after simple majority approval in a vote by the P-members on the committee. An IEC-PAS can be developed as an IS in parallel with the above procedure. Although some WGs have utilized the PAS development procedure actively since 2000, the PAS publication process was revised in 2007 because the PAS procedure was being used to avoid the need for NP approval, and the handling of comments on the draft IEC-PAS was causing confusion. Since 2007, an IEC-PAS must be approved in the proposal stage if it is to be published as a Technical Specification (TS) or as an IS. A simplified diagram of the normal



Project stage	Normal IS procedure	PAS procedure
Proposal stage (voting period: 3 months)	Acceptance of proposal	Acceptance of proposal
Preparatory stage	Preparation of working draft	Approval of PAS draft
Committee stage (commenting period: 2–4 months)	Development and acceptance of CD	
Enquiry stage (voting period: 5 months)	Development and acceptance of CDV	
Approval stage (voting period: 2 months)	Approval of FDIS	
Publication stage	Publication of IS	Publication of PAS

Stages enclosed by dotted circles may be omitted.

Fig. 2. Simplified diagram of IEC publication procedure.

procedure for developing an IS and the PAS development procedure is shown in **Fig. 2**. The IEC-PAS is approved by a simple majority of the committee's P-members after a three-month commenting and voting period.

5. Reflecting domestic standards in international standards

5.1 Trend of standardization activity of optical measurement equipment calibration

IEC TC86 WG4 has established a Sub-Working Group (SWG) for each study item, and each SWG discusses standards related to optical measurement equipment. The Japanese National Committee of IEC TC86 has proposed the JIS draft (which includes the JIS that has already been standardized) being prepared by the IEC TC86 WG4 National Committee as a new international standard. Therefore, the Japanese National Committee is providing project leaders for some SWGs. The standardization activities in IEC TC86 WG4 are listed in **Table 2**.

5.2 Calibration of optical spectrum analyzers

IEC 62129 "Calibration of optical spectrum analyzers", which was published in 2006, was approved as an IS after first being positioned as a PAS drawn up by a regional standardization organization and then proposed and approved as an IEC-PAS in IEC TC86 WG4. This is an example where the JIS was developed as an IS in parallel with being positioned as an IEC-PAS. However, since the scope of IEC TC86 WG4 is limited to calibration methods and procedures for optical measurement equipment, some parts of the description related to test methods were revised and removed from the draft IEC-PAS. The maintenance

Table 2. Standardization trend in IEC TC86 WG4.

Sub-Working Group	Project number	Established	Title	Project leader
SWG1	IEC 61315	Oct. 2005	Calibration of fibre optic power meters	Canada
SWG2	IEC 61746-1	Dec. 2009	Calibration of optical time-domain reflectometers (OTDR) - Part 1: OTDR for single mode fibres	France
	IEC 61746-2	Jun. 2010	Calibration of optical time-domain reflectometers (OTDR) - Part 2: OTDR for multimode fibres	
SWG3	IEC 61744	Sep. 2005	Calibration of fibre optic chromatic dispersion test sets	Canada (interim)
SWG4	IEC 61745	Aug. 1998	End-face image analysis procedure for the calibration of optical fibre geometry test sets	Canada (interim)
SWG5	EC 62129	Jan. 2006	Calibration of optical spectrum analyzers	Japan
SWG7	IEC 62129-2	May 2011	Calibration of wavelength/optical frequency measurement instruments - Part 2: Michelson interferometer single wavelength meters	UK
SWG8	IEC 62522	CD	Calibration of tuneable laser sources	Japan
SWG9	IEC 62129-3	PWI	Calibration of optical frequency meters using optical frequency comb	Japan

team in TC86 WG4 is examining whether there is a need to revise this standard because the time for its revision is approaching.

5.3 Calibration of tunable laser sources

IEC 62522 "Calibration of tuneable laser sources", which is currently at the committee stage, is also a document that was proposed for new IEC international standardization from Japan based on JIS C6191 "Test methods of tunable laser sources". It was initially considered that the international standardization of this document would be processed after it was published as an IEC-PAS like the calibration of optical spectrum analyzers, but it will be processed via the normal IS procedure because the PAS procedure has been changed.

However, since a document regarding tunable laser source (TLS) test methods has already been prepared as a JIS in Japan, and the necessity for TLS calibration was fully recognized in TC86 WG4, we easily obtained cooperation aimed at NP approval. Since there was a domestic standard (JIS) in Japan and the market demand had increased, it was easy to obtain the agreement of members to make a new IS in IEC TC86 WG4, and it seems that this became the motivation behind the promotion of IEC international standardization.

However, since the JIS, which is a domestic standard, prescribes a TLS test method, namely not only a calibration method but also a performance guarantee as an industrial product, some revisions were made that moved some of the test method description to the Annex, as in the case of optical spectrum analyzers.

6. Concluding remarks

This article introduced standardization activities related to calibration methods and procedures for optical measurement equipment in IEC TC86 WG4 and the relationships to domestic standards in Japan. The Japanese national standards, JIS, must harmonize with IEC specifications because of the World Trade Organization Technical Barriers to Trade (WTO/TBT) scheme [4]. The organization deliberating the standardization of optical measurement equipment in Japan is proceeding with the transfer of important IEC international standards in a related field to JIS. Meanwhile, that organization is also proceeding with an approach that reflects Japanese advanced technologies in international standards by proposing a JIS as a new international specification.

Some JISs related to optical measurement equipment have been specified as measurement equipment test methods. These JISs include descriptions that go beyond the category of calibration method specification in IEC international standards. This highlights the issue of JIS and IEC international standards having different standardization structures.

Members of the Japanese standardization committee are investigating the differences between JIS and IEC standards including the above issues and will try to resolve these problems. They have also made a timely proposal regarding a standard in the field of optical measurement equipment calibration, for which technical studies are proceeding, and they continue to take an approach aimed at producing international standards derived from domestic standards.

References

[1] IEC. http://www.iec.ch/

[2] PAS: http://en.wikipedia.org/wiki/Publicly_Available_Specification

[3] IEC-PAS. http://www.iec.ch/standardsdev/publications/pas.htm

[4] WTO/TBT. http://www.wto.org/english/tratop_e/tbt_e/tbt_e.htm

Noriyuki Araki

Senior Research Engineer, Access Media Project, NTT Access Network Service Systems Laboratories.

He received the B.E. and M.E. degrees in electrical and electronic engineering from Sophia University, Tokyo, in 1993 and 1995, respectively. He joined NTT Access Network Systems Laboratories, Ibaraki, in 1995. He has been engaged in R&D of operation and maintenance systems for optical fiber cable networks. He has been contributing to the activities of ITU-T SG15 and IEC TC86 WG4 since 2008 and 2007, respectively. He is a member of the Institute of Electronics, Information and Communication Engineers.

Visualization of Problems with Wireless Local Area Networks

Abstract

This article introduces a tool for visualizing problems related to wireless local area network (WLAN). NTT EAST has developed the wireless trouble shooting tools in order to solve Electromagnetic compatibility (EMC) problems in the field. It is the eighth in a bimonthly series on the theme of practical field information about telecommunication technologies. This month's contribution is from the EMC Engineering Group, Technical Assistance and Support Center, Maintenance and Service Operations Department, Network Business Headquarters.

1. Introduction

A wireless local area network (WLAN) provides a variety of convenient features, not the least of which is easy deployment without the need for wiring or cabling in customer premises or office buildings and so on. WLANs are also making progress in terms of higher transmission speeds, improved interconnectivity, and lower fees. WLAN usage has gone beyond personal computers to include game consoles, personal digital assistants (PDAs), mobile phones, music players, and other devices.

However, it tends to be difficult to isolate the causes of problems like failed connections, dropped connections, and drops in throughput because WLAN signals are invisible to the human eye. Moreover, spectrum analyzers for measuring signals and early WLAN analysis tools require advanced, specialized skills to operate, and these tools are so expensive that they are not widely used in the field.

Therefore, the Technical Assistance and Support Center has developed a WLAN troubleshooting tool (**Fig. 1**) designed from the perspective of maintenance personnel working in the field. By adding a WLAN monitoring function to its previous WLAN troubleshooting tool [1], [2], NTT EAST expects to make WLAN troubleshooting faster and even more efficient. The new tool visualizes WLAN problems, making them easier to understand and solve. It is now being introduced into the maintenance department.

Fig. 1. Wireless LAN troubleshooting tool.

This article presents examples of the use of this tool to troubleshoot and solve WLAN problems that commonly occur in the field.

† NTT EAST
Ota-ku, 144-0053 Japan

2. Main functions of WLAN troubleshooting tool

The WLAN troubleshooting tool conforms to the IEEE802.11a/b/g WLAN standards and incorporates the following five functions.

(1) Access point detection function
(2) Signal strength measurement function
(3) Throughput measurement function
(4) Signal strength distribution display function
(5) Electromagnetic environment monitoring function

These five functions enable even personnel without specialized skills in WLAN and radio signals to measure and analyze WLAN conditions and the peripheral environment.

In more detail, the access point (AP) detection function can be used to display nearby APs by channel and service set identifier (SSID) and to check channel usage conditions so that a channel that is not easily affected by other APs can be selected.

The signal strength and throughput measurement functions enable receive signal strength and throughput to be measured at any point in real time. This enables the troubleshooting of problems such as low throughput due to a drop in signal strength.

The signal strength distribution display function displays measured signal strength and throughput on a two-dimensional floor map so that areas with insufficient signal strength or those with poor throughput due, for example, to interference can be identified.

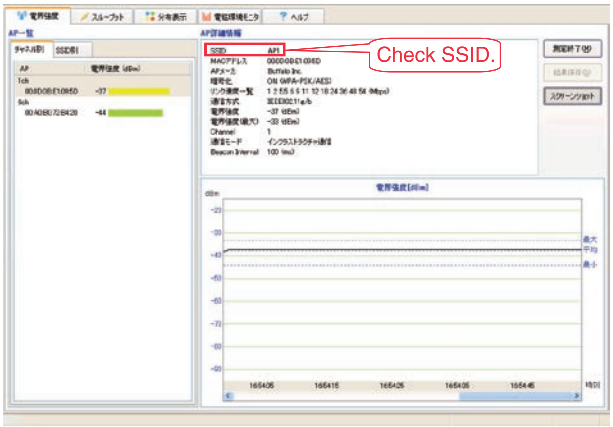
Finally, the electromagnetic environment monitoring function can be used to detect interference from industrial, scientific, and medical (ISM) devices using the same 2.4-GHz frequency band as specified by the IEEE802.11b/g standards.

3. Examples of WLAN troubleshooting

Some examples of using the WLAN troubleshooting tool to identify and solve actual WLAN problems are given below.

3.1 Erroneous SSID for WLAN AP

The SSID is an AP setting and an identifier that separates groups of WLAN connections. The AP and WLAN terminal must agree on this symbol in order to establish a connection. Since the SSID is case sensitive, the AP detection function can be used as shown in **Fig. 2** to quickly discover a SSID setting mistake by checking the actual SSID assigned to the AP in question.



SSID AP1
MACアドレス 00:0D:0B:E1:09:5D
APメーカ Buffalo Inc.
暗号化 ON (WPA-PSK/AES)
リンク速度一覧 1 2 5.5 6 9 11 12 18 24 36 48 54 (Mbps)
通信方式 IEEE802.11g/b
電界強度 -37 (dBm)
電界強度(最大) -33 (dBm)
Channel 1
通信モード インフラストラクチャ通信
Beacon Interval 100 (ms)

Check SSID.

Fig. 2. SSID checking.

3.2 WLAN channel interference

The channels specified by the IEEE802.11b/g standards have an overlapping configuration, as shown in **Fig. 3**. To avoid interference between WLAN signals, they need to be separated by five or more channels as in a ch1/ch6/ch11 arrangement (ch: channel). The AP detection function can be used to display all measured APs by channel to enable maintenance personnel to examine the overall state of channel use.

The effects of channel interference on throughput are shown in **Fig. 4**. If two APs—one set to ch1 and the other to ch3—begin to transmit at the same time, the effects of channel interference will cause throughput to fluctuate greatly in the range of 2–15 Mbit/s. However, if these two APs are set to ch1 and ch6, they will be separated by at least five channels, which prevents mutual interference. As a result, throughput becomes stable at about 15 Mbit/s even if ch1 and ch6 are transmitting simultaneously.

3.3 Interference from ISM devices

WLANs specified by the IEEE802.11b/g standards make use of the ISM 2.4-GHz band, which is also used by microwave ovens, wireless appliances, and medical equipment. Interfering signals from ISM devices can degrade throughput and impair communications. First, we examine the effects of wireless speakers using the 2.4-GHz band.

A wireless-speaker system consists of transmission equipment connected to a television set or piece of audio equipment and the wireless speakers themselves. The transmission equipment transmits audio signals by using the 2.4-GHz band so that the audio can be reproduced at the wireless speakers.

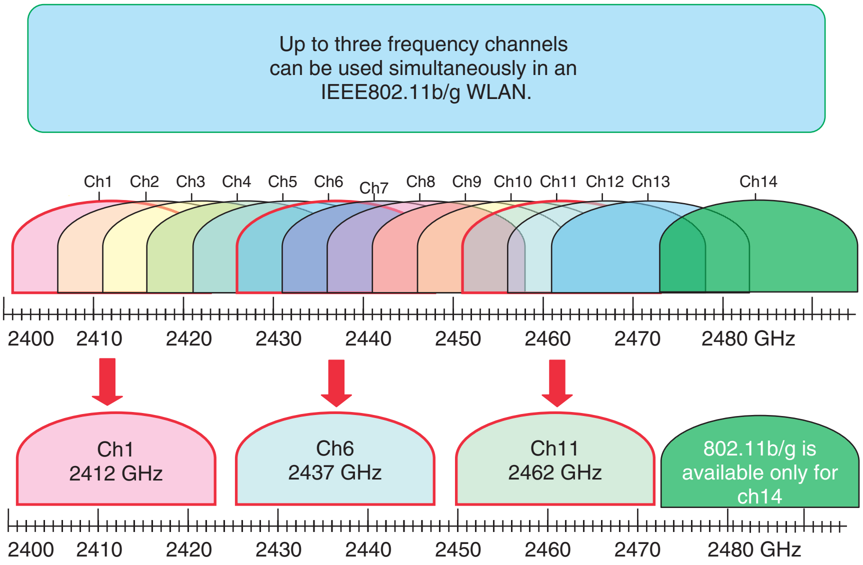
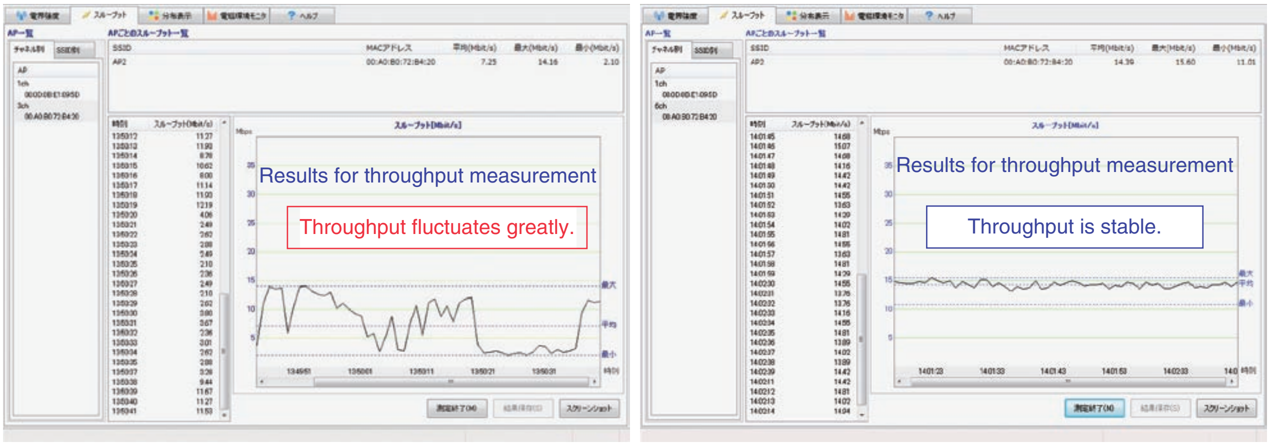


Fig. 3. Channel arrangement in IEEE802.11b/g.


Results for throughput measurement
Throughput fluctuates greatly.
Results for throughput measurement
Throughput is stable.

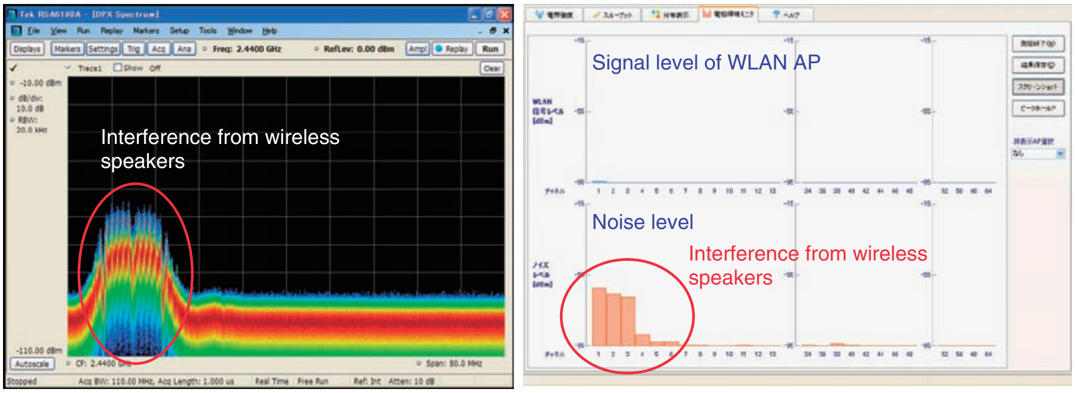
With channel interference (ch1–ch3)

Without any channel interference (ch1–ch6)

Fig. 4. Throughput measurements (effects of channel interference).

A key characteristic of a wireless-speaker system is that it continues to transmit signals on its transmission channel even if there is a WLAN in the vicinity. By contrast, the WLAN begins to transmit signals after performing carrier sensing and checking channel-usage conditions. As a consequence, a wireless-speaker system using a certain transmission channel can prevent a WLAN from beginning to transmit, cause throughput to drop noticeably, or prevent a connection from being established between the AP and terminal.

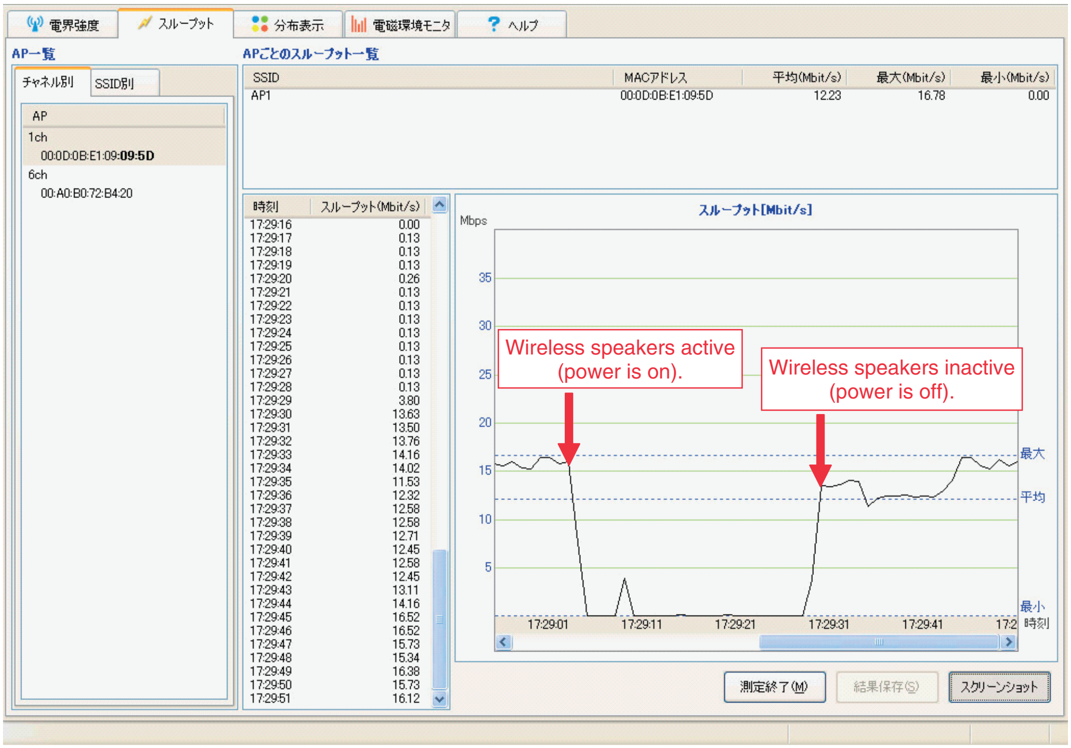
An example of measuring interference from wireless speakers by using a spectrum analyzer and the electromagnetic environment monitoring function is shown in **Fig. 5**. The spectrum analyzer displays signal level (vertical axis) versus frequency (horizontal axis). These measurement results show that the interference from the wireless speakers had a frequency component in the 2400–2430 MHz range, which corresponds to the ch1–ch5 range for WLANs. The electromagnetic environment monitoring function likewise showed a high noise level in the ch1–ch5


Interference from wireless speakers
Signal level of WLAN AP
Noise level
Interference from wireless speakers

Spectrum analyzer

Electromagnetic environment monitoring function

Fig. 5. Measurement of interference from wireless speakers.



SSID	MACアドレス	平均(Mbit/s)	最大(Mbit/s)	最小(Mbit/s)
AP1	00:0D:0B:E1:09:5D	12.23	16.78	0.00

時刻	スループット(Mbit/s)
17:29:16	0.00
17:29:17	0.13
17:29:18	0.13
17:29:19	0.13
17:29:20	0.26
17:29:21	0.13
17:29:22	0.13
17:29:23	0.13
17:29:24	0.13
17:29:25	0.13
17:29:26	0.13
17:29:27	0.13
17:29:28	0.13
17:29:29	3.80
17:29:30	13.63
17:29:31	13.50
17:29:32	13.76
17:29:33	14.16
17:29:34	14.02
17:29:35	11.53
17:29:36	12.32
17:29:37	12.58
17:29:38	12.58
17:29:39	12.71
17:29:40	12.45
17:29:41	12.58
17:29:42	12.45
17:29:43	13.11
17:29:44	14.16
17:29:45	16.52
17:29:46	16.52
17:29:47	15.73
17:29:48	15.34
17:29:49	16.38
17:29:50	15.73
17:29:51	16.12

Fig. 6. Throughput measurement (during interference from wireless speakers).

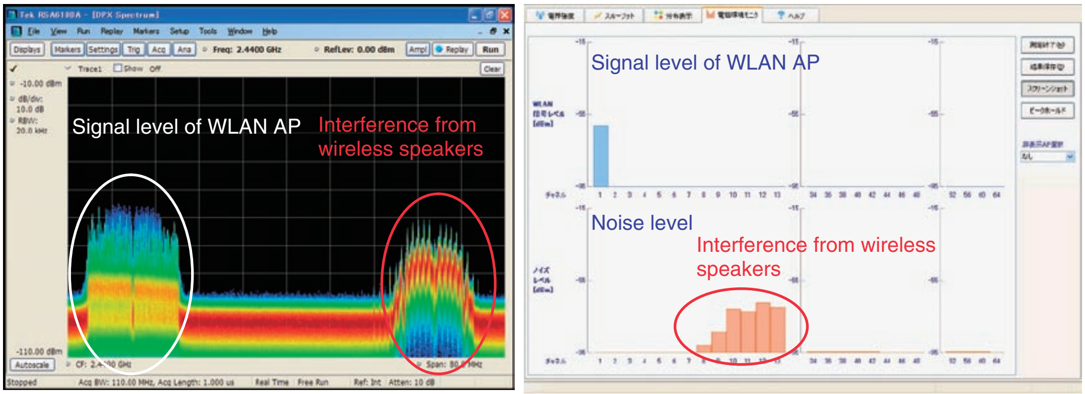
range.

Throughput affected by interference from wireless speakers is shown in **Fig. 6**. The throughput dropped from about 15 Mbit/s to 0 Mbit/s as soon as the power to the wireless speakers was turned on and returned to about 15 Mbit/s when the power was turned off.

As shown in **Fig. 7**, one measure for preventing wireless speaker signals from interfering with a WLAN is to change the channel used by either the AP or the wireless-speaker system.

3.4 Drop in receive signal strength

The field strength (measured electromagnetic power) distribution display function can be used to



Spectrum analyzer

Electromagnetic environment monitoring function

Fig. 7. Countermeasure to interference from wireless speakers.

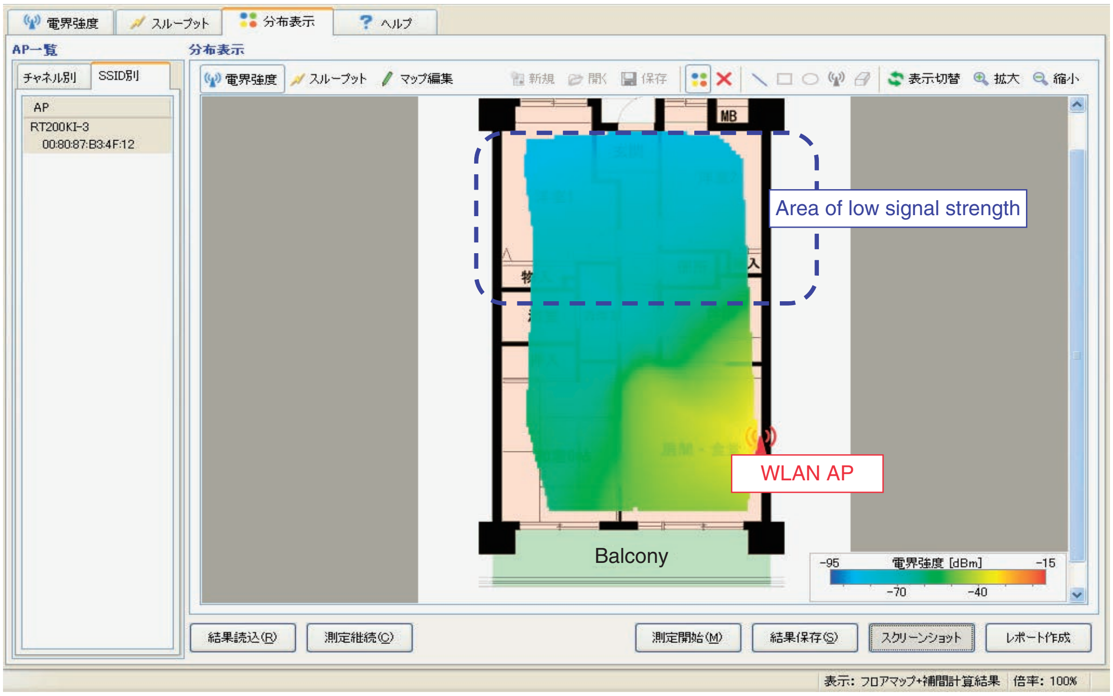


Fig. 8. Signal strength distribution.

identify areas of low signal strength for a certain AP so that relocation of the AP or the addition of another AP can be considered. An example of measured signal strength in an actual residence is shown in **Fig. 8**. This measurement reveals that AP signals did not sufficiently reach the blue area relatively far from the AP. A countermeasure such as relocation of the AP or the addition of another AP can therefore be considered.

4. Conclusion

This month's report presented examples of troubleshooting for actual WLAN problems using a tool developed by the Technical Assistance and Support Center. It would give us great pleasure if this tool finds widespread use in the field.

References

[1] “Wireless Local Area Networking Troubleshooting Technology,” NTT Technical Review, Vol. 7. No. 9, 2009. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200909sf1.html

[2] “Wireless Local Area Network Monitoring Tool,” NTT Technical Review, Vol. 8, No. 9, 2010. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201009sf1.html

External Awards

Awaya Prize Young Researcher Award

Winner: Hiroaki Itou, NTT Cyber Space Laboratories
Date: Sep. 21, 2011
Organization: Acoustical Society of Japan

For "A Study of Acoustic Evanescent Wave Reproduction Using Linear Loudspeaker Array" (in Japanese).

Awaya Prize Young Researcher Award

Winner: Shoichi Koyama, NTT Cyber Space Laboratories
Date: Sep. 21, 2011
Organization: Acoustical Society of Japan

For "Evaluation in Real Environments of Wave Field Synthesis Using Angular Spectrum Differentiation" (in Japanese).

Papers Published in Technical Journals and Conference Proceedings

Design and Implementation of New uTupleSpace Enabling Storage and Retrieval of Large Amount of Schema-less Sensor Data

T. Nakamura, K. Kashiwagi, Y. Arakawa, and M. Nakamura

Proc. of the IEEE/IPSJ 11th International Symposium on Applications and the Internet (SAINT 2011), pp. 414–420, Munich, Germany, 2011.

This paper proposes the design and implementation of a new uTupleSpace to meet increases in the variety and quantity of sensor data. We introduce two extensions to our storage and delivery system for sensor data in order to share a lot of sensor data and enable many applications to utilize them. One stores schema-less sensor data and searches among them. This enables the introduction of various types of applications for a sensor network one after another and the sharing of data stored through such applications. The other creates chunks of sensor data. This method fundamentally improves processing overheads caused by the existence of a huge amount of small sensor data. We implemented the uTupleSpace with the proposed enhancements and experimentally investigated the performance improvement achieved by creating chunks of data.

Large Array of Sub-10-nm Single-Grain Au Nanodots for use in Nanotechnology

N. Clément, G. Patriarche, K. Smaali, F. Vaurette, K. Nishiguchi, D. Troadec, A. Fujiwara, and D. Vuillaume

Small, Wiley-VCH, Vol. 7, No. 18, pp. 2607–2613, 2011.

A uniform array of single-grain Au nanodots, as small as 5–8 nm, can be formed on silicon using e-beam lithography. The as-fabricated nanodots are amorphous, and thermal annealing converts them to pure Au single crystals covered with a thin SiO_2 layer. These findings are based on physical measurements, such as atomic force microscopy (AFM), atomic-resolution scanning transmission electron microscopy, and chemical techniques using energy dispersive X-ray spectroscopy. A self-assembled organic monolayer is grafted onto the nanodots and characterized chemically with nanometric lateral resolution. The extended uniform array of nanodots is used as a new testbed for molecular electronic devices.

Population Relaxation Induced by the Boson Peak Mode Observed in Optical Hyperfine Spectroscopy of $^{167}Er^{3+}$ Ions Doped in a Silicate Glass Fiber

D. Hashimoto and K. Shimizu

J. Opt. Soc. Am. B, Vol. 28, No. 9, pp. 2227–2235, 2011.

We demonstrate transient saturation spectroscopy for $^{167}Er^{3+}$ ions doped in a silicate glass fiber cooled at 2.5–30 K to measure the population relaxation time t_1 of the hyperfine sublevels. The observed t_1 value is 3.1 ms at 4 K and we observe anomalous temperature dependence whereby t_1 becomes rather longer with heating from 4 to 30 K. We can regard the population relaxation as being a result of the Raman scattering of the Boson peak mode (BPM) peculiar to a silicate glass by the 4f-electrons of the $^{167}Er^{3+}$ ions. We can attribute the anomalous temperature dependence to the suppression of the Raman scattering by thermal hopping of the localized BPM.

Usability Evaluation of Pointing Using Self-image from Arbitrary Viewpoint

E. Hosoya, I. Harada, A. Onozawa, and H. Murase

Human Interface, The Transactions of Human Interface Society, Vol. 13, No. 3, pp. 221–233, 2011 (in Japanese).

The pointing operation in video communication with a "shared space" approach is evaluated. The shared space is a constructed image in which a remote image is overlaid with the self-image. It is known that such a manner of communication can promote natural conversation including easy gaze recognition and a feeling of space sharing. However, the performance of the pointing operation in the shared space has not been examined fully. Experiments have been done to evaluate it with various parameters such as viewing angle and mirroring of the shared space image. As a result of these experiments, guidelines for shared space design using pointing have been created.

CENSREC-4: An Evaluation Framework for Distant-talking Speech Recognition under Reverberant Environments

T. Fukumori, T. Nishiura, M. Nakayama, Y. Denda, N. Kitaoka, T. Yamada, K. Yamamoto, S. Tsuge, M. Fujimoto, T. Takiguchi, C. Miyajima, S. Tamura, T. Ogawa, S. Matsuda, S. Kuroiwa, K. Takeda, and S. Nakamura

Acoust. Sci. & Tech., Vol. 32, No. 5, p. 201–210, 2011.

We have been distributing a new collection of databases and evaluation tools called CENSREC-4, which is a framework for evaluating distant-talking speech in reverberant environments. The data contained in CENSREC-4 are connected-digit utterances as in CENSREC-1. Two subsets are included in the data: "basic data sets" and "extra data sets." The basic data sets are used for evaluating the room-impulse-response-convolved speech data to simulate various reverberations. The extra data sets consist of simulated data and corresponding real recorded data. Evaluation tools are currently provided only for the basic data sets and ones for the extra data sets will be delivered in the future. The task of CENSREC-4 with a basic data set appears simple; however, the results of experiments prove that CENSREC-4 provides a challenging reverberation speech-recognition task, in the sense that a traditional technique to improve recognition and a widely used criterion to represent the difficulty of recognition deliver poor performance. Within this context, this common framework can be an important step toward the future evolution of reverberant speech-recognition methodologies.

Present and Future of Terahertz Communications

H. J. Song and T. Nagatsuma

IEEE Trans. on Terahertz Science and Technology, Vol. 1, No. 1, pp. 256–263, 2011.

Recent changes in how people consume multimedia services are causing an explosive increase in mobile traffic. With more and more people using wireless networks, the demand for ultra-fast wireless communications systems is increasing. To date, this demand has been accommodated with advanced modulation schemes and signal-processing technologies at microwave frequencies. However, without increasing the carrier frequencies to create more spectral resources, it may be difficult to keep up with the needs of users. Although there are several alternative bands, recent advances in terahertz-wave (THz-wave) technologies have attracted attention owing to the huge bandwidth of THz waves and their potential for use in wireless communications. The frequency band of 275–3000 GHz , which has not been allocated for specific uses yet, is especially of interest for future wireless systems with data rates of 10 Gbit/s or higher. Although THz communications is still in a very early stage of development, there have been lots of reports that show its potential. In this review, we examine the current progress of THz-wave technologies related to communications applications and discuss some issues that need to be considered for the future of THz communications.