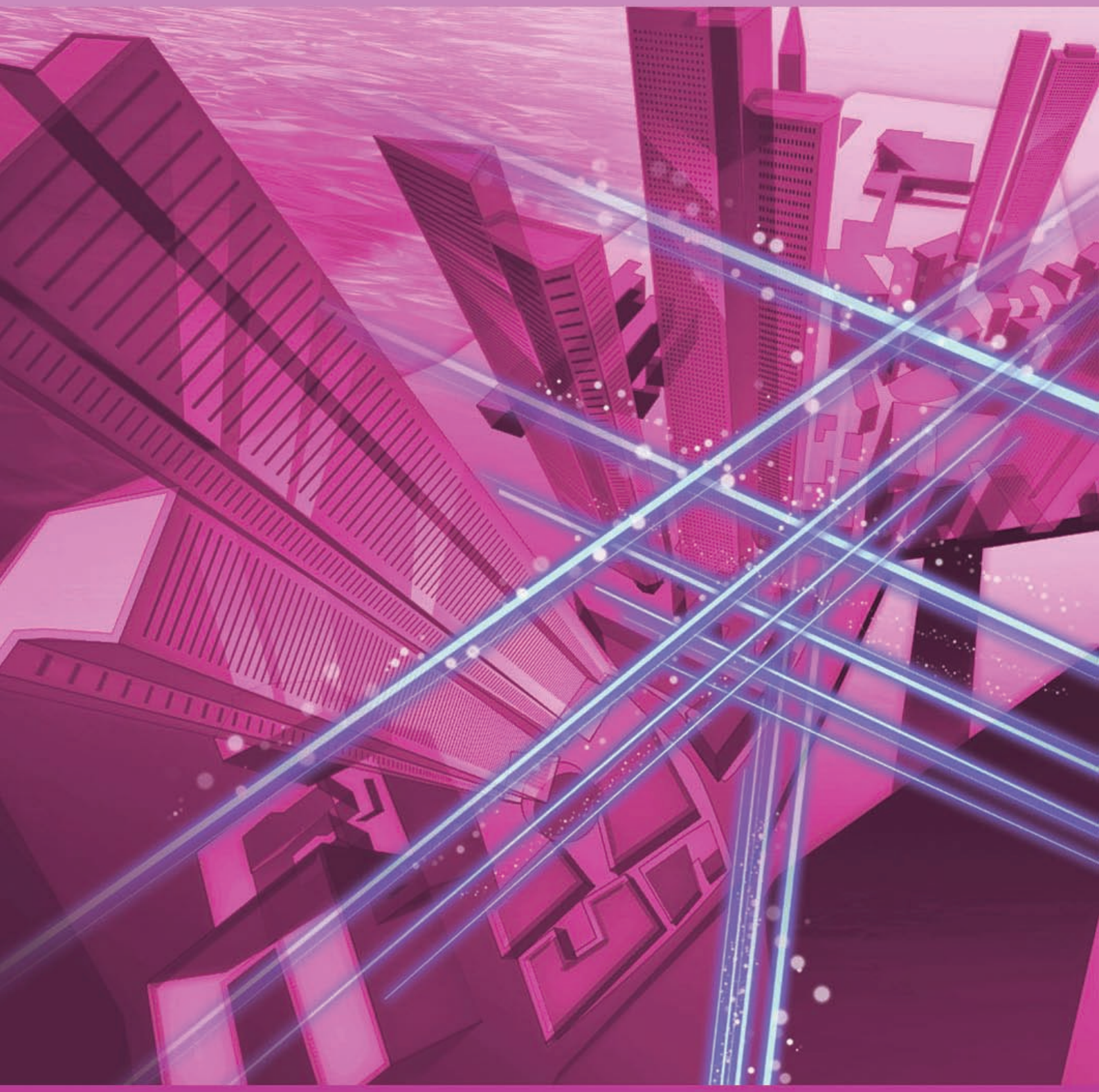
NTT Technical Review

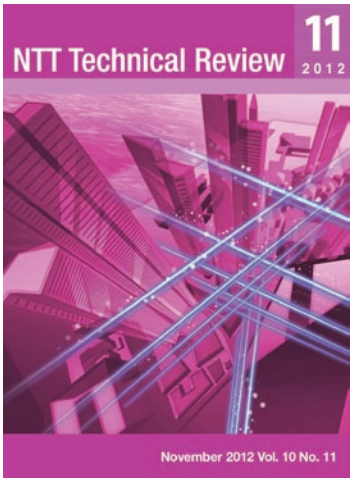
11
2012



November 2012 Vol. 10 No. 11

NTT Technical Review

November 2012 Vol. 10 No. 11



Global Standardization Activities

Revision of the Common Patent Guidelines for ITU/ISO/IEC

Information

Report on NTT Communication Science Laboratories Open House 2012

External Awards/Papers Published in Technical Journals and Conference Proceedings

External Awards/Papers Published in Technical Journals and Conference Proceedings

Achieving Dramatic Gains in Domestic and International Markets through Remarketing

Toshio Iwamoto
President and Chief Executive Officer, NTT DATA

Overview

To respond to the accelerating globalization of the information and communications technology (ICT) market, NTT DATA is enthusiastically working as a unified enterprise to establish a dominant presence in world markets through its Global One NTT DATA system. Using novel strategies and viewpoints, Toshio Iwamoto, the new President and Chief Executive Officer, has demonstrated outstanding ability as a project manager throughout his years at NTT DATA. After our previous talk with him two and a half years ago, we sat down with him again to discuss the company's new Medium-Term Management Policy.

Understanding society's needs and structural changes through fresh viewpoints and up-to-date information

—Congratulations, Mr. Iwamoto, on your appointment as President of NTT DATA. Could you tell us what you would like to accomplish going forward?

There are various things that I would like to accomplish in the years ahead, and in this regard, let me tell you about the two main pillars of our new Medium-Term Management Policy. One is becoming a Global Top 5 player, which is quite a challenge, and the other is improving our corporate value to an earnings-per-share (EPS) of 20,000 yen.

At present, we are ranked no. 6 in terms of market share in global information technology (IT) services. In other words, we are just one rank away from entering the top 5, but there is a large gulf between ranks 5 and 6. Growth is the root of corporate activities, and to provide our customers with top-notch, compelling services, we plan to execute an all-out growth strategy that integrates the abilities of both domestic and overseas companies in the NTT DATA Group.

To meet our EPS target, we are strongly committed to improving our corporate value as demanded of a global enterprise. We intend to do this by helping our customers raise their earnings and creating new value for society. To reach an EPS of 20,000 yen, we must raise our net profit by around 10% annually over the next four years, which is no small challenge.

In all sincerity, I would like to work toward this goal together with all 60,000 or so NTT DATA Group employees working in Japan and overseas.

A new strategy unfolding under the keyword "remarketing"

—Do you have a specific strategy for accomplishing these objectives?

We actually have three concrete strategies here:

remarketing; major reform of software production technologies; and expansion, enhancement, and fortification of global business.

Of course, our business activities are based on marketing. It may sometimes happen that we give up on past marketing results, viewing them as ineffective, and then lose faith in all marketing. However, the market environment is changing significantly over time and the technologies of our industry are advancing rapidly. In times past, there were cases in which we simply could not meet the needs of our customers, but nowadays, there are increasingly more cases in which we can satisfy them.

I call this the *afterimage* phenomenon. If we can completely eliminate the afterimage of results obtained by past marketing efforts and take another look at the market from a fresh perspective, then perhaps we can open up new markets in Japan and abroad. This is what I call remarketing, which is an approach that I would like to pursue wholeheartedly. This concept of remarketing also has another meaning: by obtaining a clear understanding of society's needs and the changes in social structures, we should be able to provide new and novel services and create new markets. I would like to pursue these activities on an ongoing basis.

For example, the removal of the ban on banks being involved in the insurance business allowed them to sell insurance from bank counters, and in response to this change in social structure, we developed a system for handling such transactions. We also created an e-money system making use of technical innovations. This required extensive work at the time of development, but we take pride in having created a system that pleased our customers and in being able to spread and establish it in society.

I feel that NTT DATA can play a useful social role by seizing these opportunities to enter new markets through social changes, the easing of regulations, and technical innovations. By working to cultivate new markets, we can support social progress and help bring about a feeling of well-being and happiness in many people.

—NTT DATA has undertaken the development of unique, large-scale systems in the past. On the basis of this experience, what do you think will be the issues confronting Japan in the future and the role of NTT DATA in solving them?

In the public sector, there are a number of issues that must be addressed. These include the creation of a new national identity (ID) system and comprehensive reform of social security and tax. Of course, we would like to play an active, helpful role in solving these social issues. Most of Japan's public systems are unique in the sense that each has only a single version. For example, only one customs system is needed in Japan, but I would like to see it applied in such a way that other countries could make use of it too.

The financial industry is likewise in a severe situation. If the market changes, then so do the methods used for settlement and associated IT processing. We would like to provide and support outstanding products in pace with those changes.

Let me now move on to major reform of software production technologies. In general, the software development process begins with the definition of requirements, which are then used as a basis for basic and detailed design, programming, testing, and shipping. Examining this sequence of tasks, we can see that there are some process flows in which human labor is concentrated. In today's software industry, collaboration with partner companies is the usual choice as a short-term remedy for such a problem. Moreover, we outsource those portions of work that are labor intensive to overseas companies to cut costs.

With these things in mind, let's take a quick look at the manufacturing industry. The trend toward automation in production plants is accelerating and it is becoming possible to stabilize product quality and ensure satisfactory levels of quality, cost, and delivery. I often think about whether such a system—which entails a transformation from labor-intensive to knowledge-intensive work—could also be applied to the software industry. What I would like to do here is

focus talented personnel and manpower on performing critical design work and extracting and satisfying customer needs.

In fact, we faced a similar challenge about 30 years ago, but were not successful in achieving such a laborsaving system. Compared with today, computers were severely lacking in power, and system implementation was costly. Now, however, such a system is possible. We would like to exploit the high performance of today's computers to achieve a major reform of software production technologies. We are already making progress in automating nearly 100% of software development in several large-scale projects, so I think we are getting better in this regard. We would like to become a global leader in such systems and propose a business model created in Japan.

Finally, let's talk about global business. At present, we are doing business from offices in 136 cities in 35 countries and regions. Although our expansion to date has centered on mergers and acquisitions, we have been integrating and reorganizing overseas group companies since January 2012, beginning in North America, and have nearly completed the integration of the entire group under the NTT DATA brand. I think that we are finally in the first stage of expanding our global business, and we have already built an environment in which we can quickly provide services to our customers throughout the world, many of whom are Japanese global enterprises. In the second stage of this process, the plan is to link all NTT DATA Group employees worldwide and expand, enhance, and fortify our global business.

At the same time, there's more to global business than just talking about ways of expanding business. There are also problems dealing with the merging of cultures and the integration of human resources. We intend to address and solve these problems.

Recognize one's role by zooming in and taking a historical viewpoint

—Our previous talk, two and a half years ago [1], left a deep impression on me because you talked about the need for a manager to have firm philosophical convictions and a strong desire to develop business based on those convictions. Now that you have become president, what do you see as your cornerstone in working to achieve NTT DATA's objectives?

That would be the way of looking at things. I tell employees that there are two important ways of looking at the world around you.

The first is looking by zooming in. Advances in the network are enabling us to understand things in the world around us as if we were taking them in our hands. When something major happens in the world, you should first examine it closely and then consider its effects on roles and activities from a sequence of viewpoints, starting with Japan as a whole and then moving on to NTT DATA, its various sections and projects, your family or the environment immediately surrounding you, and finally yourself. In other words, by first obtaining the big picture of a world trend and then gradually zooming in on yourself, you should be able to understand what you should do and what stance you should take in response to that trend.

The second is looking from a historical viewpoint. Treating the question of how far back to look as a separate issue, let's take a look at the history of the Internet as one example. If we consider 1995 to be the first year of the Internet in Japan, we can see that not even 20 years have passed since then. From a world perspective, the diffusion of web browsers is said to have begun in 1993 (although there are various opinions about this), and that too is only about 20 years ago. And what about computer history? Well, the first computer in the world is said to have been the ENIAC in 1946, but even if we include analog computers (as opposed to today's digital computers), the birth of the computer is still barely more than 65 years ago. Against this background, we are now entering an era in which many people are carrying around handheld

computers such as smartphones that are having a major effect on personal lifestyle and on government, economic, and business activities. Looking at technology developments in a historical manner does not mean concentrating on just certain points but rather examining such development processes comprehensively and consistently. This approach enables us to foresee how another process may unfold. Applying it to smartphones and looking five years into the future, I can imagine that some kind of device may arise to replace it even without changing the basic smartphone concept, or perhaps the basic form of the smartphone will change.

Zooming in or adopting a historical viewpoint in this way lets you understand the present and focus on things to come. I would like us to work toward achieving our objectives on the basis of this way of looking at the world around us.

—Mr. Iwamoto, could you leave us with a few words of encouragement for researchers and project managers?

To NTT DATA researchers, I would say, "Let's clarify what we want to accomplish through technology development." While asking ourselves what kind of information society needs to be created, let's improve our ability to foresee the near future. To this end, we need to collaborate with other NTT Group companies and form alliances whenever appropriate or feasible. There are various keywords that can be used to describe the work of NTT DATA, such as *big data*, machine-to-machine (M2M), and sensor networks, but in any case, we take great pride in being a technology company. The tools that technology-development personnel go on to create lie at the heart of what we are. Furthermore, I would like to see us develop Japan-original technologies and services not only for Japan but also for the whole world, and I would like to make investments to that end.

To project managers, I would say, "Be philosophers!" Without inspiration and empathy, people will not move. So on the basis of this principle, I would like them to motivate people while having firm beliefs about life. This is a mindset that I want all project managers to have at all times.

Reference

[1] T. Iwamoto, "'Trust' and 'Spread of New Technologies' Toward an Era where IT Surpasses our Dreams," NTT Technical Review, Vol. 8, No. 5, 2010.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201005tp1.html

Interviewee profile

■ Career highlights

Toshio Iwamoto joined Nippon Telegraph and Telephone Public Corporation (now NTT) in 1976. After the spinoff of the NTT Data Communications division into NTT DATA Corporation in 1988, he served as Senior Executive Manager of the Settlement Solutions Sector from 2004, Director and Executive Vice President in charge of the Financial Business Sector from 2007, and Director and Executive Vice President in charge of the Financial Sector from 2008. He was appointed Representative Director and Senior Executive Vice President in 2009. He assumed his present position in June 2012.

Communication Science for the Big Data Era

Naonori Ueda

Abstract

This set of Feature Articles, on the theme of communication science that connects information and humans, introduces the research efforts and achievements of NTT Communication Science Laboratories for the *big data* era. Around the world, digital data is being created and stored in enormous quantities and the era of big data is now upon us. For research and business applications, it is vital to have some way of extracting new value from these huge and diverse data resources.

1. Introduction

The spread of social media and mobile information devices has accelerated in recent years, and huge quantities of digital data are now being produced and stored throughout the world. It reached 1.8 ZB (zettabytes: 10^{21} bytes) in 2011 and is expected to increase another fiftyfold over the next 20 years. We can safely say that the era of *big data* is now upon us. The use of big data as a buzzword can be traced back to a May 2011 report "*Big data: The next frontier for innovation, competition, and productivity*" by MGI (McKinsey Global Institute, the research division of McKinsey & Company) [1]. This report suggested that big data can lead to the creation of huge monetary value in diverse fields including healthcare, economics, and manufacturing. For example, it is estimated that the utilization of big data helps the US healthcare sector save $300 billion each year. There is also a growing trend towards the spread of cloud environments and open-source large-scale parallel distributed computing environments such as *Hadoop*, so big data is clearly set to become a driving force behind information and communications technology (ICT) innovation in this century [2].

The term big data encompasses not only large and highly diverse (atypical) data that is difficult to accumulate, edit, and store in conventional databases, but also the substantial benefits to industry and society that are gained by deriving new value from it. Or to put is another way, big data is not just about how much data you have, but also about the scenarios you use to gain new value from it and the techniques you use to implement these scenarios. For example, Indiana University, USA, has devised an analysis method that can make stock market forecasts with 87% accuracy by analyzing 9.8 million tweets from 2.7 million Twitter users. Its ability to predict the stock market from tweets can be attributed to the easy availability of vast quantities of data. This example indicates that the big data era is characterized by the profitable application of completely new analysis methods. In other words, businesses must somehow derive valuable information from the serendipitous effects of combining large quantities of seemingly unrelated information.

On the other hand, researchers need core technologies that can extract and merge latent information from large and diverse data sources in order to make predictions. In addition to these core technologies, it is also important to construct a next-generation ICT infrastructure to provide a foundation for big data. This infrastructure should support a cyber-physical system that allows computing resources to be used in cyberspace to perform advanced analysis of highly diverse data, and it should use the results to promote a real-world system in order to construct a highly efficient social system. In this set of Feature Articles, as some of the research currently under way at NTT Communication Science Laboratories (NTT CS Labs), we introduce the core technology and infrastructure technology that is essential for the big data era.

2. Cyber-physical system core technology and infrastructure technology

Machine learning techniques have been attracting attention for big data analysis applications. Simple statistical analysis is insufficient for dealing with a huge and diverse set of data, so to support a cyber-physical system we need techniques that can learn the underlying models of data (data generation mechanisms) from previous data and can use these models for learning in order to predict future events. However, the majority of conventional machine learning techniques are targeted at predefined tasks such as regression analysis and classification problems. In the big data era, what we really need is a technique that can take in a diverse variety of data and extract the latent data structures (correlation relationships and causality relationships) that exist within it. In other words, we need hypothesis discovery techniques rather than hypothesis testing techniques. At NTT CS Labs, with this in mind, we are researching relational data mining techniques based on a machine learning approach, and we are also conducting empirical studies using real data from sources such as Twitter. This is described in detail in the Feature Article "Extracting Essential Structure from Data" [3] in this issue.

As more information is made public in the big data era, issues of security and privacy become more apparent. At present, encryption and authentication techniques are used to prevent the disclosure of or tampering with data transmitted across networks. These security techniques are based on random numbers, but since most random numbers are currently pseudorandom numbers generated by numerical series calculations, it is easy to predict the random numbers that will be output if the initial value (seed) and generating formula are leaked. Thus, from this viewpoint, absolute security is not assured. An alternative method called physical random number generation produces random numbers from physical phenomena. For some years now, we have been working on a method for generating physical random numbers from the chaotic behavior of semiconductor lasers. In 2008, we achieved the world's fastest physical random number generation rate of 1.7 Gbit/s. In 2010, we implemented this technique in a miniaturized device by taking advantage of optical integrated circuit technology. Since this method is theoretically guaranteed to provide unpredictable results, it can be said to be a core security technology that provides the ultimate in security. Details can be found in the Feature Article "Fast Physical Random Number Generation Using Semiconductor Laser Chaos" [4].

Attention is also being drawn to machine-to-machine (M2M) systems in which machines connected to computers in a cyber-physical system can implement intelligent control through the mutual exchange of information without any human intervention. Sensor network technology is therefore an important part of the technical infrastructure for such systems. In the future, sensor network technology will not only make use of information derived from the sensor values themselves, such as temperature readings and chemical concentrations, but also require advanced information processing capabilities such as the ability to infer the occurrence of events from the information obtained from multiple sensors. At the CS Labs, we have developed a technique for recognizing human behavior from multiple sensors such as accelerometers, cameras, and microphones. We are also researching a method for gathering information efficiently from a large number of wireless sensor nodes. Details can be found in the Feature Article "Information Processing of Sensor Networks" [5].

3. Basic research exploring the essential nature of communication

Needless to say, conversations between humans form the basis of communication. However, a conversation is not just an exchange of linguistic information; other factors such as feelings of sympathy and antipathy (nonverbal information) also play important roles. We are researching a system and conversation scene analysis technique that utilizes techniques for extracting verbal information (speech recognition) and nonverbal information (emotion recognition) in a conversation to ascertain the circumstances under which messages are sent (when, where, who, to whom, what, how, and why (known as 6W1H)). We are working on a system called MM-Space that reproduces human head movements as physical movements of a display screen in order to allow a remote conversation participant to be represented at a different location. This system is described in detail in the Feature Article "MM-Space: Recreating Multiparty Conversation Space by Using Dynamic Displays" [6].

In addition to simple enhancements of telecommunications technology in a cyber-physical system, it is also essential to implement a communications-based society that lets anyone enjoy the convenience of this

advanced technology in a safe, secure, and enriching way. To this end, we feel that it is important to promote not only information science but also research on its human science and social science aspects. Ever since its establishment, the CS Labs has been conducting research aimed at clarifying the nature of communication. ICT must adapt to human society. To create a richly featured communication environment, we cannot simply confine our studies to information per se; we must also study the information processing mechanisms that humans use to send and receive information. Recently, we have been trying to clarify how the human brain assimilates information gathered by the senses, and we have made new discoveries related to the interaction between sight and hearing. Details can be found in the Feature Article "Hearing Sound Alters Seeing Light" [7].

References

[1] MGI (McKinsey Global Institute, the research division of McKinsey & Company), "Big data: The next frontier for innovation, competition, and productivity," 2011.

[2] H. Shinohara, "R&D to Create the Future of ICT," NTT Technical Review, Vol. 10, No. 4, 2012. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201204fa2_s.html

[3] K. Ishiguro and K. Takeuchi, "Extracting Essential Structure from Data," NTT Technical Review, Vol. 10, No. 11, 2012. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201211fa2.html

[4] K. Yoshimura, S. Shinohara, and K. Arai, "Fast Physical Random Number Generation Using Semiconductor Laser Chaos," NTT Technical Review, Vol. 10, No. 11, 2012. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201211fa3.html

[5] T. Suyama and Y. Yanagisawa, "Information Processing of Sensor Networks," NTT Technical Review, Vol. 10, No. 11, 2012. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201211fa4.html

[6] K. Otsuka, "MM-Space: Recreating Multiparty Conversation Space by Using Dynamic Displays," NTT Technical Review, Vol. 10, No. 11, 2012. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201211fa5.html

[7] T. Kawabe, "Hearing Sound Alters Seeing Light," NTT Technical Review, Vol. 10, No. 11, 2012. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201211fa6.html

Naonori Ueda

Director, NTT Communication Science Laboratories.

He received the B.S., M.S., and Ph.D. degrees in communication engineering from Osaka University in 1982, 1984, and 1992, respectively. He joined the Yokosuka Electrical Communication Laboratories of Nippon Telegraph and Telephone Public Corporation (now NTT) in 1984. In 1994, he moved to NTT CS Labs in Kyoto, where he has been researching statistical machine learning, Bayesian statistics, and their applications to web data mining. From 1993 to 1994, he was a visiting scholar at Purdue University, Indiana, USA. He is a guest professor at the National Institute of Informatics and the Nara Advanced Institute of Science and Technology. He is a Fellow of the Institute of Electronics, Information and Communication Engineers and a member of the Information Processing Society of Japan and IEEE.

Extracting Essential Structure from Data

Katsuhiko Ishiguro and Koh Takeuchi

Abstract

In this article, we introduce part of our technique for analyzing and mining data matrices by using statistical machine learning approaches. The analysis of big data is becoming a hot trend in the information and communications technology (ICT) field, and automated computational methods are indispensable because the data volumes exceed manual capabilities. Statistical machine learning provides good solutions for this purpose, as we show in this article.

1. Introduction

The analysis of *big data* is becoming a hot trend in the information and communications technology (ICT) business. However, there is no concrete definition of big data. Most of the data treated in big data analyses is characterized by its large amount, which greatly exceeds the amount that can be treated manually.

For example, gigantic purchase records of an online commerce service are organized and managed by computers, and this data helps to generate product recommendations for each customer. And the servers used by Twitter, a well-known online social networking service (SNS) and microbloging service, handle more than 4.5 million tweets (messages) per day [1]. Analyzing trends in Twitter obviously requires computers to deal with this amount of digital data.

We need an automated computational process and data mining technique to understand the characteristics of such big data and extract useful knowledge from it. However, they will not come into existence by themselves. We must choose a principle, or criterion, that defines the actual computation process and controls the data usage. NTT Communication Science Laboratories (CS Labs) is researching data mining techniques based on statistical machine learning that seek the best answers in the statistical and probabilistic senses.

Statistical machine learning primarily handles numbers, i.e., numerical data. In this article, we assume that the data can be converted into a data matrix like the cells of a spreadsheet (**Fig. 1**). For example, we can convert the purchase records of an online shopping service by placing the customer's identity (ID) on the vertical axis and the product ID on the horizontal axis. In the same manner, an SNS friend network can be converted into a data matrix. A friend relation, or a follower relation, between two users is defined by the source (user who follows) and the destination (user who is followed). Setting the source as the vertical axis and the destination as the horizontal axis, we obtain the data matrix of an SNS friend network. These are just examples: many kinds of data can be converted into data matrices.

In this article, we introduce a few data mining techniques that are being studied and developed at NTT CS Labs. We use statistical machine learning techniques to model such data as structural relations between a small number of essential factors such as product purchase patterns or communities and hubs in networks. These methods automatically decompose complicated data and find essential factors without careful manual tuning thanks to statistical criteria.

In sections 2 and 3, we introduce two remarkable methods: nonnegative matrix factorization (NMF) [3] and infinite relational models (IRMs).

2. Pattern extraction from real-valued data matrix by NMF

NMF is applicable for a data matrix whose elements are nonnegative. Its goal is to decompose the

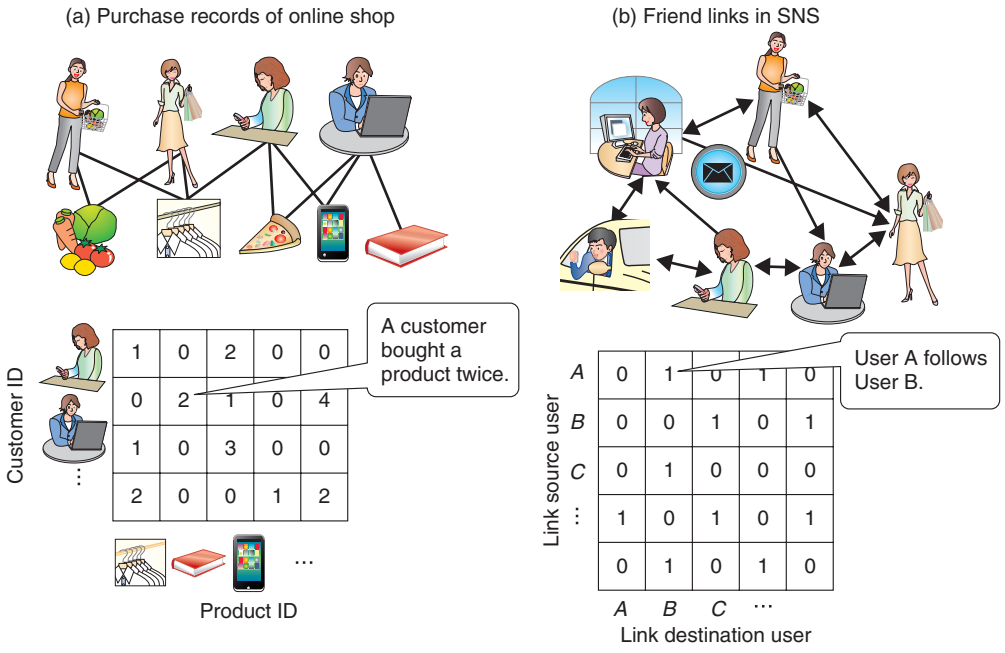


Fig. 1. Data matrix format.

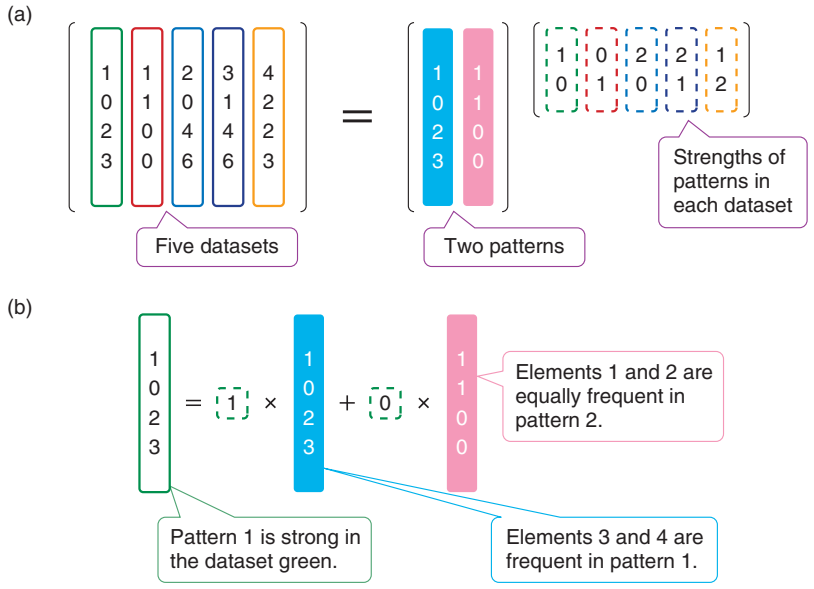
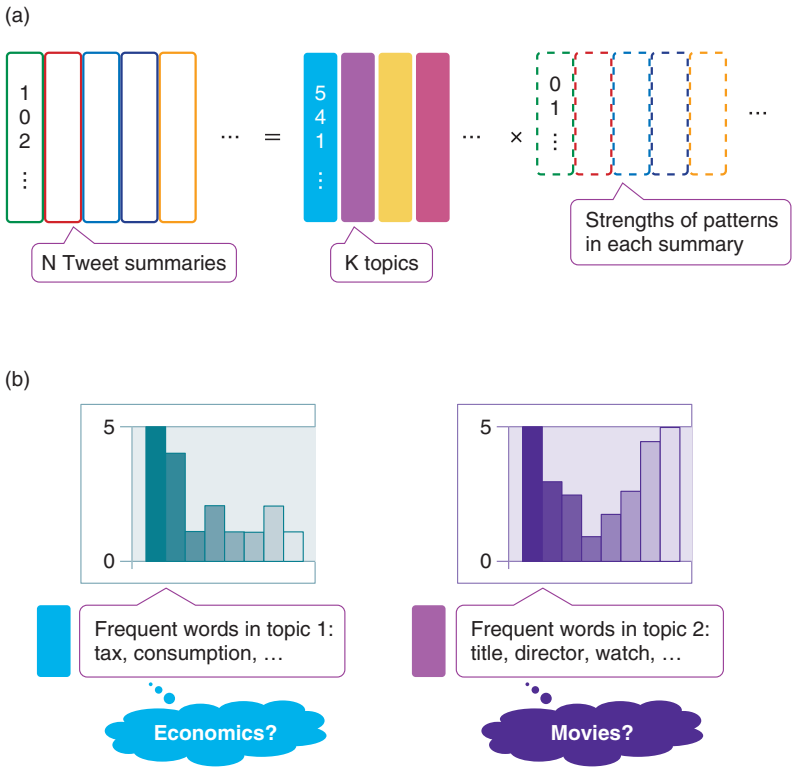


Fig. 2. Overview of NMF method.

data matrix into two smaller matrices: one representing patterns and the other representing the strengths of the patterns in each dataset. NMF finds two optimal nonnegative matrices in terms of the reconstruction errors with respect to the original data matrix (**Fig. 2(a)**). An example of how to interpret the NMF decomposition is shown in **Fig. 2(b)**.

Let us consider the analysis of a newspaper data matrix. The data matrix consists of the word counts of the newspaper articles. Colored column vectors



(c)

Topics	Top words in terms of frequency
1	economics, consumption, tax, policy, yen, deflation, finance, ratio, type, crisis, heavier-tax, inflation, financial, %, Japan central bank,
2	movie, watch, title, director, see, roadshow, class, enter, ghost, smell, product, interesting, version,
3	mouth, foot, infection, dispose, cow, bird, Miyazaki, kill, prefecture, outbreak, former, head, information, vaccine, Korea, pig, animal husbandry, animal, virus, day, inspection,
4	town, evacuation, information, place, prefecture, human, Sanriku, Iwate, tsunami, city, intact, Oozuchi, contact, Sendai, Miyagi, south,
5	earthquake, day, prefecture, Fukushima, hour, magnitude, minute, occurrence, breaking news, shake, information, emergency, offshore, tsunami, epicenter, city, JMA

Fig. 3. Application of NMF to Twitter data.

correspond to articles, and a number in a column vector represents the number of times a word appears in an article. In this case, the pattern matrix consists of column vectors, where each vector corresponds to a topic discussed in articles, such as sports, economics, and politics. Each pattern has its own word frequency distribution: thus, we can identify the contents of a pattern. Moreover, each article (dataset) has its own pattern strengths, as summarized in a strength matrix. Using this strength, we can easily summarize the contents of many articles. In summary, NMF extracts patterns and pattern strengths at the same time by decomposing the original data matrix.

Here, we introduce an application of NMF to Twitter topic extraction. It is not an easy task to find and grasp topics existing in Twitter because of the number and variety of tweets. We tackle this problem by decomposing summary articles (summaries) in a content curation site, which presents collections of social media content voluntarily collected and reordered.

In this task, the data matrix consists of a number of column vectors where each column vector corresponds to a summary. The elements in a column vector indicate the appearance frequencies of words in an summary (**Fig. 3(a)**). NMF gives patterns of word distributions corresponding to topics in Twitter (**Fig. 3(b)**) and the strengths of patterns within each summary at the same time.

The results of analyzing approximately 100,000 summaries (summarizing approx. 10,000,000 tweets) from Feb. 2010 to Apr. 2011 are presented in **Fig. 3(c)**. Extracted topics are very clear and easy to interpret. For example, the first topic is related to economics, the third topic corresponds to foot-and-mouth disease outbreaks in Miyazaki, the fourth topic concerns tsunamis at Sanriku in the previous Tohoku Earthquake, and the fifth topic is related to emergency earthquake warnings.

3. Relation cluster extraction from binary data matrix by IRMs

In this section, we introduce IRMs [4] that are especially applicable for relational data, which represents existences of relations between multiple objects. Examples of relational data are shown in **Fig. 4**. In

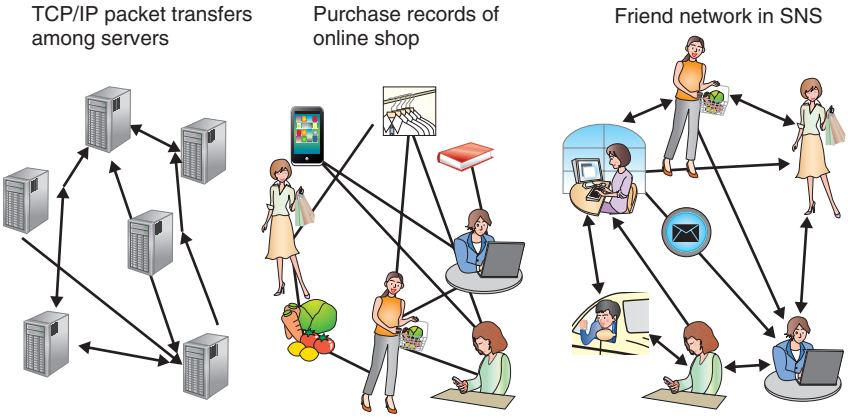


Fig. 4. Examples of relational data.

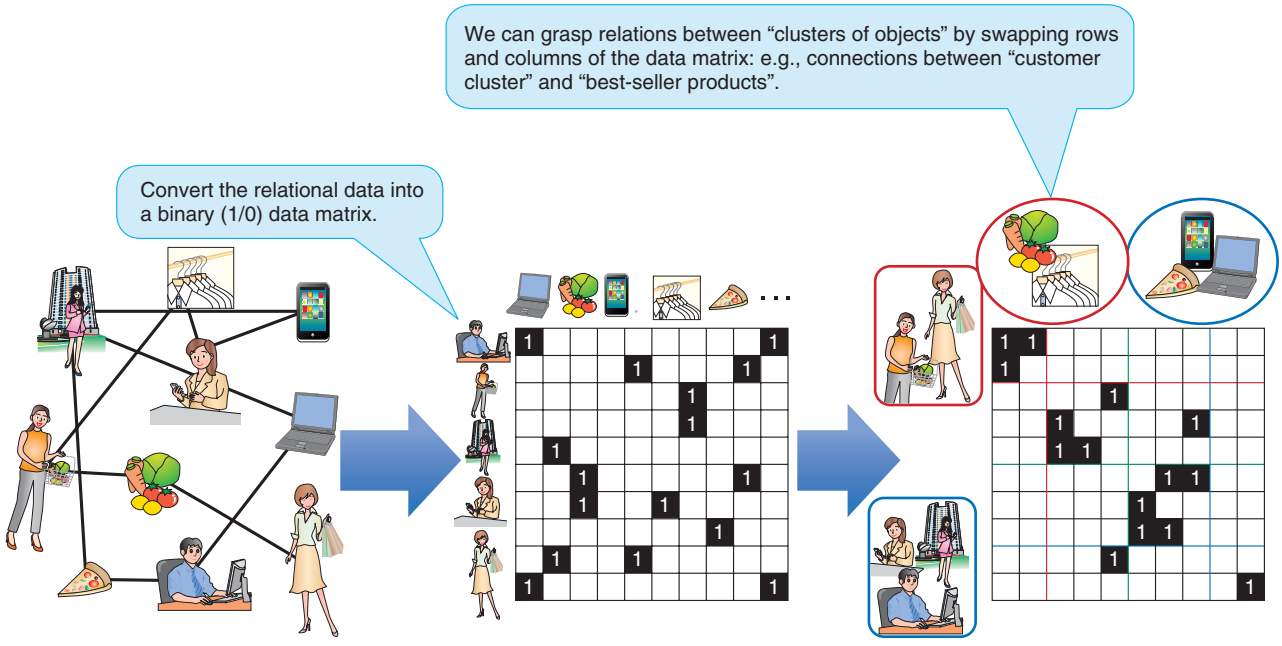


Fig. 5. Overview of IRM method.

general, relational data focuses on the structure of links between several objects (servers, customers, products, etc.).

We can convert such relational data into a binary data matrix, as in **Fig. 5**. The value 1 indicates the existence of a link, and the value 0 indicates the nonexistence of a link. Given the data matrix, IRM extracts some good groupings (colored partitions in the figure) by swapping and re-ordering row indices and column indices. By good, we mean that the resultant data matrix consists of partitions that are very white or black and not mixtures of white and black. In the case of Fig. 5, IRM extracts pairs of a “specific customer group” and “specific products frequently purchased by the members of the group”: in other words, relations between (customer) groups and

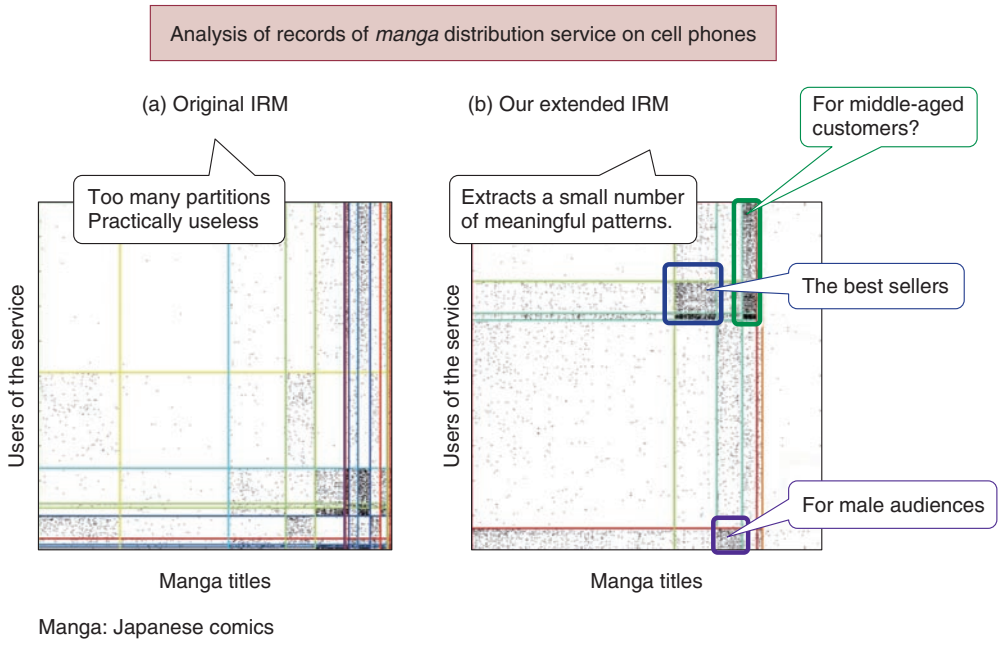


Fig. 6. Our extended IRM for sparse relational data.

(product) groups. If we apply IRM to SNS friend link data, we can find communities in the network links and interactions between communities.

One benefit of using IRM is that it automatically decides the number of groups, or the cardinality of the partitions, in a statistically optimal manner. Thus, once a relational data matrix has been input, IRM takes care of the remaining tasks and enables analyses like the one described above.

In this article, we would like to present two novel extensions of IRM developed by CS Labs. The first extension enables users to analyze time-varying relational data. Many relations change over time. For example, human connections in an organization gradually change day by day, and sometimes change drastically with a reorganization.

Our IRM extension can represent such mixed changes in relations. In experiments, we analyzed email histories at Enron Corporation. Our method successfully extracted and tracked changes in a community of employees related to its financial and monetary sections, and a few key persons who were very influential throughout the company.

Another of our recent achievements can handle problems induced by data sparsity, which is often observed in real-world purchase data. Most purchase history records in e-commerce are characterized by the very small number of actual purchases compared with the huge number of customer-product pair entries. When such data is converted into a data matrix, most of the matrix cells have the value 0; namely, the data matrix is sparse. The naïve IRM does not perform well on the sparse data matrix shown in the left panel of **Fig. 6**. It extracts so many groups that a lot of human effort is required to interpret and extract knowledge from the analysis results.

We extended IRM to extract only important purchase patterns from a sparse data matrix. Our key hypothesis is that products that everyone buys and that no one buys do not provide useful information for data mining. Similarly, users who buy nothing are also uninteresting. Our extended IRM automatically excludes objects that do not have specific patterns of relations, and the remaining interesting part of the data matrix is analyzed by IRM. An example of extended IRM is shown in the right panel of Fig. 6: the model successfully extracts a limited number of groups, and these groups are easy for the human eye to interpret.

4. Conclusion

In this article, we introduced two data mining techniques, NMF and IRM, for data matrices. The scope of statistical machine learning studies is limited to data matrices. CS Labs is committed to developing

state-of-the-art machine learning technologies and is contributing to the development of data mining techniques for even larger and more complicated datasets.

References

[1] Twitter. https://twitter.com

[2] C. M. Bishop, "Pattern Recognition and Machine Learning," Springer, 2006.

[3] M. Mørup, "Applications of Tensor (multiway array) Factorizations and Decompositions in Data Mining," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol. 1, No. 1, pp. 24–40, 2011.

[4] C. Kemp, J. B. Tenenbaum, T. L. Griffiths, T. Yamada, and N. Ueda, "Learning Systems of Concepts with an Infinite Relational Model," Proc. of the 21st National Conference on Artificial Intelligence (AAAI'06), Vol. 1, pp. 381–388, Boston, MA, USA, 2006.

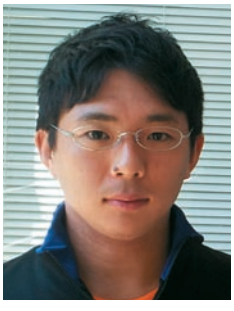
[5] K. Ishiguro, T. Iwata, N. Ueda, and J. Tenenbaum, "Dynamic Infinite Relational Model for Time-varying Relational Data Analysis," Proc. of the 24th Annual Conference on Neural Information Processing Systems (NIPS 2010), Vancouver, B.C., Canada.

[6] K. Ishiguro, N. Ueda, and H. Sawada, "Subset Infinite Relational Models," Proc. of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS 2012), La Palma, Canary Islands, Spain.

Katsuhiko Ishiguro

Researcher, Innovative Communication Laboratory, NTT Communication Science Laboratories.

He received the B.E. and M.Informatics degrees from the University of Tokyo in 2004 and 2006, respectively, and the Ph.D. degree from the University of Tsukuba, Ibaraki, in 2010. Since joining NTT CS Labs in 2006, he has been working on various research projects including multimedia data modeling with Bayesian approaches, probabilistic models for structured data mining, and time series analysis. He is a member of the Institute of Electronics, Information and Communication Engineers, the Information Processing Society of Japan, and IEEE.



Koh Takeuchi

Researcher, Innovative Communication Laboratory, NTT Communication Science Laboratories.

He received the B.E. and M.E. degrees from Waseda University, Tokyo, in 2009 and 2011, respectively. His research interests include statistical modeling of the brain-computer interface and social media analysis.

Fast Physical Random Number Generation Using Semiconductor Laser Chaos

Kazuyuki Yoshimura, Susumu Shinohara, and Kenichi Arai

Abstract

We review recent developments in research on fast physical random number generation using a random phenomenon in semiconductor lasers. Random numbers are widely used in information security technology, and there is a need for high-quality, i.e., unpredictable and uniformly distributed, random numbers.

1. Introduction

A number (or a sequence of numbers) generated with no apparent rule is called a *random number* (*random numbers*). Random numbers are widely used for various purposes. Familiar examples are the number showing on a thrown dice or selected in roulette for gaming and a winning lottery number. Random numbers are also often used for computer simulation in both science and technology. In addition, the random number is an essential component of security technologies such as cryptographic systems and authentication systems.

There are two fundamental properties that should be possessed by random numbers. One is *statistical uniformity*, which is the property that the values of random numbers are uniformly distributed over their range. The other is *unpredictability*, which is the property that there is no means of fully or partially predicting the next value in a random number sequence. In particular, for security purposes, a random number must have unpredictability because this ensures a high security level. Therefore, it is crucial to develop a method of generating random sequences with these properties.

Some security technologies require a large quantity of random numbers in their operation. Typical examples are secret sharing schemes, which are used for securely storing data in storage devices by encrypting and dividing it, and quantum key distribution, which is expected to achieve the ultimate security. Therefore, there is a need for methods capable of generating random numbers at a high generation rate.

This article deals with a random number generation technique based on a random physical phenomenon, focusing on the method using fast random oscillation exhibited by semiconductor lasers. This method has the potential to generate high-quality unpredictable random numbers and achieve a high generation rate, and there has been great interest in it recently. We review recent developments in the research of fast physical random number generation using semiconductor lasers. More detailed technical reviews are given in Refs. [1] and [2].

2. Types of random number generation methods

Roughly speaking, there are two types of random number generation methods.

The first method produces *pseudorandom numbers*. It is a computational method. Given an initial number a_0 called a seed, the method computes the next value a_1 from a_0 via a prescribed mathematical formula and then computes a_2 from a_1, and so on. The sequence a_0, a_1, a_2 ... seems to be a random sequence, provided that an appropriate formula is used. Because of its

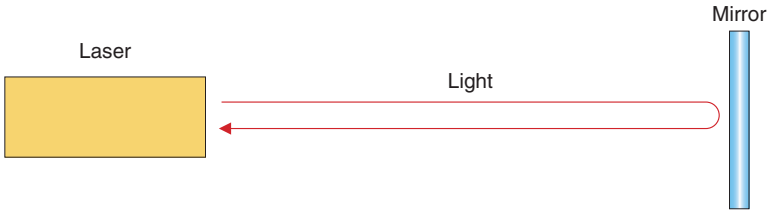


Fig. 1. Laser with optical feedback.

computational nature, it is in principle possible to predict the sequence if both the formula and the seed become known. Pseudorandom number generators for security purposes are devised in such a manner that it will be difficult to determine the seed from the generated sequence. However, it is still impossible for pseudorandom number generators to achieve complete unpredictability.

The other method produces *physical random numbers* by utilizing randomness in physical phenomena. Typical and well known examples are dice and roulette. In addition, there are technologically more useful methods, which are based on the measurement of thermal noise in an electrical circuit or measurement of a quantum optical phenomenon. Compared with pseudorandom numbers, physical random numbers have the great advantage that can achieve complete unpredictability if an appropriate random physical phenomenon is used. However, the existing methods have the disadvantage of a low generation rate.

From the viewpoint of security applications, it is important to develop a random number generator that is guaranteed to have unpredictability and can achieve a high generation rate.

3. Chaos in semiconductor lasers

In a variety of systems in nature and technology, the time evolution of the system is described by a rule such that its future state is uniquely determined from its present and past states. Such a rule is called a deterministic rule. It has been known that systems governed by deterministic rules often exhibit irregular and very complex behaviors. Such a phenomenon with irregular and complex behavior is called *chaos*. The remarkable feature of chaos is a sensitive dependence on the initial state; i.e., when there are two identical systems with slightly different states at the initial, the difference rapidly becomes large with time.

Nowadays, it is recognized that chaos is not a rare anomaly but a ubiquitous phenomenon. Chaos can be easily observed in semiconductor lasers. In a semiconductor laser with optical feedback, the laser's output light is reflected by a mirror and injected back into the laser (**Fig. 1**). An ordinary semiconductor laser, which has no feedback, emits light with constant intensity; by contrast, a semiconductor laser with optical feedback is intrinsically unstable, and the intensity of its output light exhibits irregular and complex oscillation. An example of the waveform of light output from a semiconductor laser with optical feedback is shown in **Fig. 2**. This experimental result clearly shows that chaos, which is characterized by irregular and complex oscillation, occurs in a semiconductor laser with optical feedback.

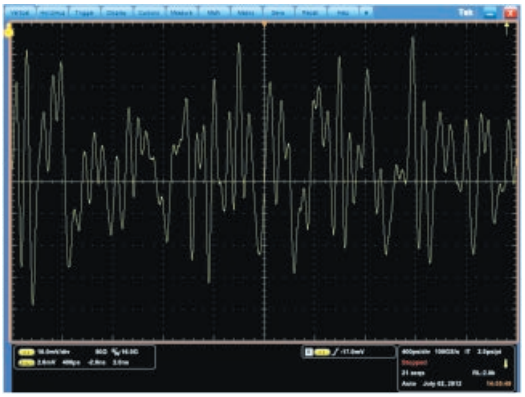


Fig. 2. Intensity of output light of laser with optical feedback.

The chaos in a semiconductor laser with optical feedback has been extensively studied over the last three decades from the viewpoint of basic research. This subject has been being studied in collaborative research between Dr. Uchida's group at Saitama University and NTT Communication Science Laboratories. Recently, we focused attention on the fast irregular oscillation in a semiconductor laser with optical

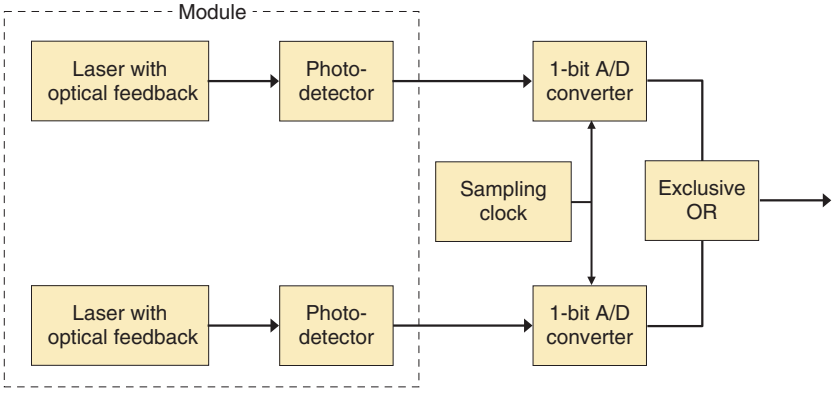


Fig. 3. Configuration of random number generator using laser with optical feedback.

feedback, which has broad bandwidth of the order of several gigahertz, and we proposed its application to fast random number generation.

4. Random number generation using semiconductor lasers

A simple method for obtaining binary random bits from the analog output waveform of a semiconductor laser with optical feedback is to periodically sample the waveform and digitize the sampled values by comparing them with a prescribed threshold: if the sampled value is smaller than the threshold, it is digitized to 0, otherwise to 1. Our research group has been developing a scheme for fast random number generation based on this simple method.

4.1 System configuration

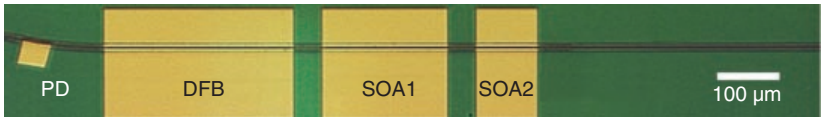
The configuration of our experimental laser random number generator system, which has been developed in research collaboration with Dr. Uchida's group (at Takushoku University at that time) is shown in **Fig. 3**. Although it is in principle possible to generate random bits with only one laser, our system in Fig. 3 consists of two independent lasers. This system configuration was devised to remove effects due to the weak periodicity of a single laser output. The use of two lasers improves the quality of random bits generated at a high generation rate.

In Fig. 3, each semiconductor laser with optical feedback emits output light, and the output light intensity is measured by a photodetector. Each measured intensity value is converted into a binary value by an analog-to-digital (A/D) converter. The final output bit is obtained by applying an exclusive OR operation between the two binary values. Using this experimental system, we succeeded in achieving a random bit generation rate of 1.7 Gbit/s in November 2008; this was a world record for physical random bit generation rate [3]. Since this pioneering work, fast random bit generation using a chaotic laser has become an active research area studied at many institutes around the world.

4.2 Challenges

The most advantageous feature of using a semiconductor laser in random bit generation is its potential high-speed performance, which can be much faster than the rate achievable with electrical circuits. It is of course important to increase the random bit generation rate by fully utilizing the high-speed performance. On the other hand, from the viewpoint of security applications, it is essential to guarantee the unpredictability of generated random bits so that users can use the bits without anxiety. In addition, from a practical viewpoint, it is necessary to make a system that is compact. We have been working on these important issues.

Our first experimental system developed in 2008 was composed of a number of optical components, and its size was about 1 m × 1 m, which is not small enough for practical use. Therefore, we applied optical integrated circuit technology developed by NTT Photonics Laboratories to our system and succeeded in drastically reducing the system size in 2010. The optical integrated circuit, which is about 10 mm × 0.3 mm, is shown in **Fig. 4**. It incorporates a semiconductor laser with optical feedback. A distributed feedback semiconductor laser (DFB) emits light, which propagates along the waveguide (black line in Fig. 4)



SOA: semiconductor optical amplifier
PD: photodetector

Fig. 4. Optical integrated circuit of laser with optical feedback.

and is reflected by a mirror placed at the right end to be reinjected into the DFB laser. This integrated circuit also has a photodetector at the left end.

A module containing two of the optical integrated circuits shown in Fig. 4 is shown in **Fig. 5**. This module can perform the same function as the components enclosed by dashed lines in Fig. 3. It generates two irregularly fluctuating signals, which are the intensities of light output from the two built-in optical integrated circuits. Random bits can be obtained by digitizing these signals and applying an exclusive OR operation, as shown in Fig. 3. We have shown that it is possible to generate random bits at the rate 2.08 Gbit/s by using this module [4].

The other important issue is the unpredictability of generated random bits. We have developed a theory for ensuring unpredictability [4], which we outline below. It is known that there is some unavoidable small quantum noise called spontaneous emission noise in any laser. This spontaneous emission noise is the origin of the unpredictability. Because of this noise, the state of a laser at any given moment cannot be determined uniquely, so it is necessary to consider that the laser state is distributed according to a certain probability distribution. The small uncertainty arising from this probability distribution rapidly becomes large within a short time because of the dynamical instability due to the chaos in a laser with optical feedback. This enlarged uncertainty leads to the uncertainty in the output light intensity waveform and hence to that in the random bits. Our computer simulation indicates that even if you know the state of a laser with optical feedback exactly at a certain moment, you cannot predict the intensity of its output light after 1 ns at all [4].

An experimental result confirming the rapid increase in uncertainty in a single semiconductor laser with optical feedback is shown in **Fig. 6**. In this experiment, we used the optical integrated circuit in Fig. 4, reset the system to the same initial state for each trial, and allowed the system to evolve. Figure 6 shows the time evolution of the output light intensity obtained for three trials. In the initial stage, the three curves are close to each other. By contrast, after 0.5 ns, the three curves are significantly different. This

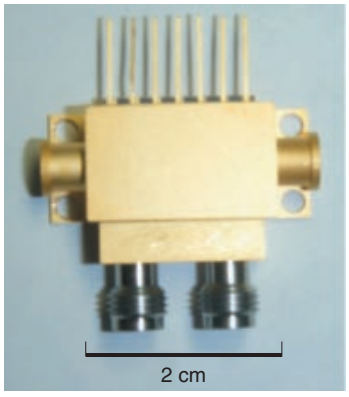


Fig. 5. Module with two built-in optical integrated circuits.

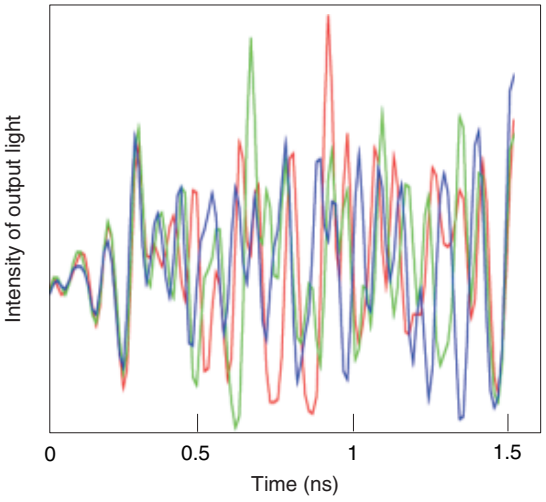


Fig. 6. Experimental result for amplification of uncertainty.

demonstrates that a small initial uncertainty rapidly increases to the macroscopic output level within only 0.5 ns.

5. Future perspective

The goal of our research is to establish the theoretical and experimental basis for physical random number generation technology using chaos in lasers. For this purpose, important issues are to generalize the theory guaranteeing the unpredictability of random bits and to develop a technique for experimentally confirming the theory. We are conducting research on these issues.

References

[1] A. Uchida, "Review on ultra-fast physical random number generators based on optical random phenomena," The Review of Laser Engineering, Vol. 39, No. 7, pp. 508–514, 2011 (in Japanese).

[2] T. Harayama, S. Sunada, and K. Tsuzuki, "Optical integrated circuits for laser chaos and fast physical random number generation," The Review of Laser Engineering, Vol. 39, No. 7, pp. 515–519, 2011 (in Japanese).

[3] A. Uchida, K. Amano, M. Inoue, K. Hirano, S. Naito, H. Someya, I. Oowada, T. Karashige, M. Shiki, S. Yoshimori, K. Yoshimura, and P. Davis, "Fast physical bit generation with chaotic semiconductor lasers," Nature Photonics, Vol. 2, No. 12, pp. 728–732, 2008.

[4] T. Harayama, S. Sunada, K. Yoshimura, P. Davis, K. Tsuzuki, and A. Uchida, "Fast nondeterministic random-bit generation using on-chip chaos lasers," Physical Review A, Vol. 83, No. 3, 031803, 2011.

Kazuyuki Yoshimura

Distinguished Researcher, Media Information Laboratory, NTT Communication Science Laboratories.

He received the B.E. degree in engineering physics from Kyoto University and the M.E. degree in aeronautics and astronautics from the University of Tokyo in 1992 and 1994, respectively. He received the Ph.D. degree in applied mathematics and physics from Kyoto University in 1997. He joined NTT in 1997. He was a visiting scholar at the University of California, San Diego, USA, during 2001–2002. His research interests are in nonlinear dynamics and its applications to communications. He is a member of the Physical Society of Japan, the Japan Society for Industrial and Applied Mathematics, the Institute of Electronics, Information and Communication Engineers (IEICE), and the Japan Society for Aeronautical and Space Sciences.

Susumu Shinohara

Research Specialist, Media Information Laboratory, NTT Communication Science Laboratories.

He received the Ph.D. degree in physics from Waseda University, Tokyo, in 1999. He joined NTT as a research specialist in 2012 after working at Waseda University, Ritsumeikan University, Advanced Telecommunications Research Institute International, and Max Planck Institute for the Physics of Complex Systems. His research interests are in nonlinear physics, classical and quantum chaos, and their applications. He is a member of the Physical Society of Japan (JPS).

Kenichi Arai

Senior Research Scientist, Media Information Laboratory, NTT Communication Science Laboratories.

He received the B.S. and M.S. degrees both in pure and applied physics from Waseda University, Tokyo, in 1991 and 1993, respectively, and the Ph.D. degree from Waseda University in 2003. He joined NTT Communication Science Laboratories in 1993. His research interests are in nonlinear dynamics, stochastic systems, neural networks, and complex networks. He is a member of IEICE and JPS.

Information Processing of Sensor Networks

Takayuki Suyama and Yutaka Yanagisawa

Abstract

In this article, we introduce technologies for collecting information from many sensor nodes deployed in the environment and/or attached to people, interpreting the sensor data as meaningful information, and presenting it to people appropriately.

1. Introduction

In our daily lives, we are surrounded by many sensors. For example, a video game machine uses an accelerometer as one of its input devices, and a mobile phone uses GPS (global positioning system) to determine the user's location. In the field of weather forecasting, pressure, temperature, and other information about various places is collected in a network and utilized in creating a weather forecast. Although individual sensors are used in various situations, sensor networks are becoming easier to use as a result of sensor evolution and the spread of sensor network technology driven by progress in semiconductor technology. Sensor networks and their information processing technology are expected to be used for various purposes, such as collecting information to enable the provision of appropriate services that suit a particular situation and for recognizing the real world.

2. Research on sensor networks

Research on sensor networks is progressing with a focus on a sensor node equipped with various sensors, a calculation function (central processing unit (CPU), memory, etc.), wireless-communications functions, a battery, and networking capability. The sensor nodes that we have developed are shown in **Fig. 1**. Such research was triggered by the Smartdust project announced in 1999. In the Smartdust concept, the user deploys many sensor nodes, each of which consists of sensors, a battery, a digital signal processor, and a transmitter assembled into a cuboid with sides of about 1 mm to 2 mm, and observes environmental information collected from the sensor nodes. However, the need for miniaturization, which was initially advocated and dictated the sensor node's size, is no longer considered to be such a high priority.

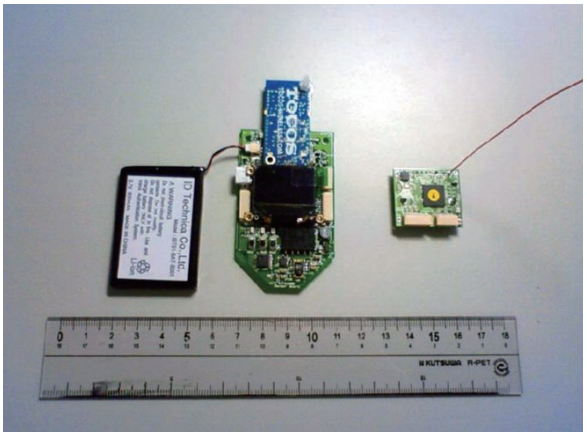


Fig. 1. Sensor nodes.

The flow of information on a sensor network is shown in **Fig. 2**. We assume that sensor nodes are deployed in the environment and/or attached to people. First of all, it is necessary to collect information efficiently from a lot of sensor nodes. Because the collected sensor data is the waveform of a time series, it is difficult for humans to understand the waveform's meaning. So the next step is to interpret such time series information. When suitably interpreted meaning is presented to the user, the collected sensor data is finally being used effectively.

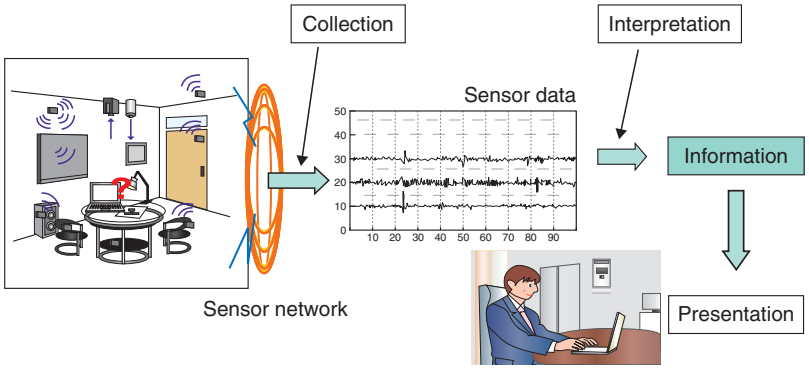


Fig. 2. Flow of sensor data.

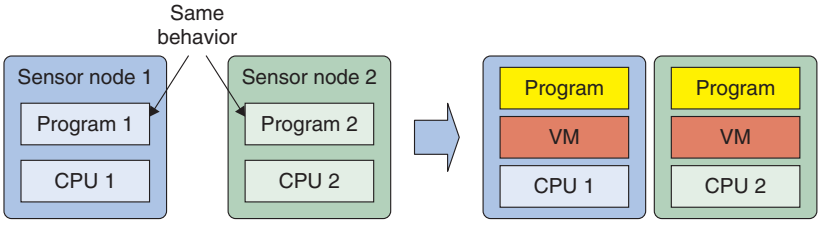


Fig. 3. Virtual machine for wireless sensor node.

3. Virtual machine for wireless sensor node

When many sensor nodes are used, it is important to reduce the cost of using them. In particular, in a network where two or more kinds of sensor (e.g., ones from different makers or with different CPUs) nodes are intermingled, the usage cost is high. To enable easy use in such situations as well, we have developed a virtual machine (VM) for a wireless sensor node. Usually, when users use various different sensor nodes, as shown on the left in **Fig. 3**, even when the same operation is carried out, another program needs to be prepared. But with the VM, even if the CPUs differ, the same program can be run on them.

Our VM is based on CIL (Common Intermediate Language). Since CIL originated in Windows .NET, executable code can be created using the development environment of MS-Windows. Moreover, a sensor network's manner of operation can be changed, even after sensor nodes have been deployed in the environment, because this VM has a function for wirelessly distributing created programs.

4. Sensor data collection from many sensor nodes

In order to collect information from many sensor nodes, analyze the situation in the real world from the sensor data, or create a sensor network that provides services according to a situation, it is necessary to collect data efficiently. Data compression technology based on the similarity of sensor data is shown in **Fig. 4**. Sensor data measuring the situation in the real world has the characteristics that there are periodic changes in the same sequence or changes in similar values obtained from neighboring sensors. Figure 4(a) shows the compression when sensor data from a single sensor node shows a similar value repeatedly. First, the sequence of sensor data repeated periodically is cut out. Then, the sensor data is decomposed into typical data sequences and weight variables by using singular value decomposition (SVD).

Weight variables express the importance for every typical data sequence, and the original sensor data can be restored by using these typical data sequences and weight variables. When information is transmitted from a sensor node, it is possible to compress the

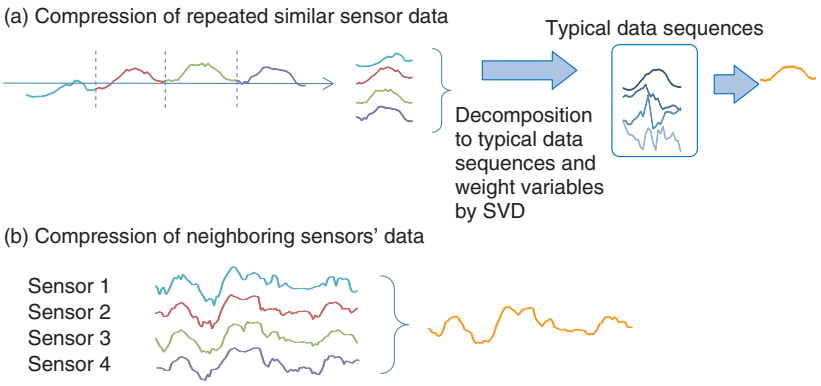


Fig. 4. Compression of sensor data.

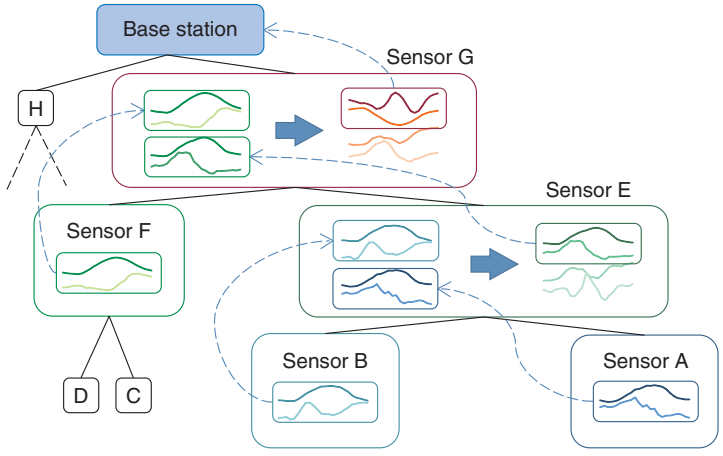


Fig. 5. Collection of sensor data in a hierarchical network.

information by not transmitting all of the typical data sequences but transmitting only typical data sequences that make a high contribution. However, because not all of the information is sent, a certain amount of error occurs. An example of transmitting the data of neighboring sensors is shown in **Fig. 4(b)**. Temperature may be data that shows similar values among neighboring sensors. This sensor data can be compressed by using the abovementioned method. The sensor data is transmitted to the base station while the sensor nodes are used as a hierarchical network (**Fig. 5**). After the data of sensor A and sensor B has been compressed and sent to sensor E, it is further sent to sensor G. Finally, the data is sent to a base station. Thus, it becomes possible to reduce the amount of information even further by applying the compression technique shown in Fig. 4 hierarchically. By reducing the data amount in the wireless communications, we can reduce the battery consumption, enabling the sensors to measure efficiently for a long time and collect more sensor information [1].

5. Interpretation of collected data

If the information collected from the sensors can be interpreted and human activity can be inferred, this kind of technique lets us make applications such as ones for automatic lifelog creation, elderly care support, and home automation. We assume that almost all human activities are performed using the hands, so we developed technology to infer the kind of action being performed by using only a sensor device attached to a person's wrist. Moreover, whereas previous methods can infer only simple human activities

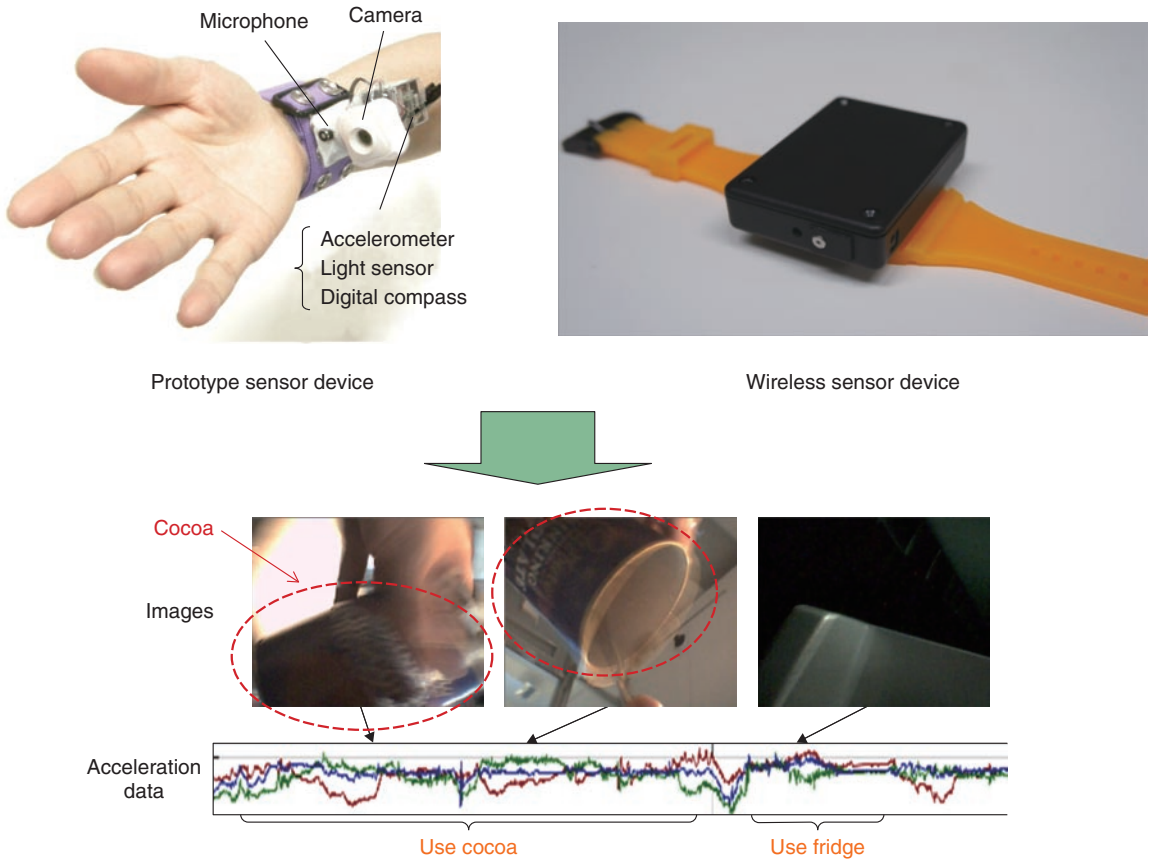


Fig. 6. Wrist-worn sensor devices and example of sensor data.

such as running and walking by using an accelerometer, our method can infer more complex activities with a comparatively high degree of abstraction, such as vacuuming and making cocoa. The wrist-worn sensor device consists of two or more kinds of sensors, for example, a camera, accelerometer, and microphone. The cameras acquire visual information about an object being manipulated, the accelerometer captures hand movements, and the microphones captures ambient sounds [2]. Prototype sensor devices and the information obtained from them are shown in **Fig. 6**.

Human activity is inferred from sensor data by applying the machine learning technique. A system that can infer actions from sensor data is shown in **Fig. 7**. First, features are extracted from the sensor data. For example, in the case of a camera image, the raw image might be heavily blurred and privacy issues might arise depending on what other details are captured, including ones in reflections. In our system, these problems are avoided by using the picture's color information for recognition. A histogram of the color contained in a picture is used as the feature. Other sensor data is also converted into corresponding features. Next, the extracted features are input to a classifier.

First, we prepare several binary classifiers—one for each activity that we want to identify—and the probabilities for all actions are output. For example, the data input from one binary classifier to the operation "making cocoa" judges the degree of agreement with the cocoa-making operation. Since 15 kinds of actions were identified in the experiment, 15 classifiers were used. Features were input in parallel with these classifiers. Next, they were input to a hidden Markov model (HMM) classifier, the time direction was inferred, and the output result of the binary classifier produced the final classification result. In the experiment, when all the sensors were used, activities were recognized with probabilities ranging from 80% to 90%.

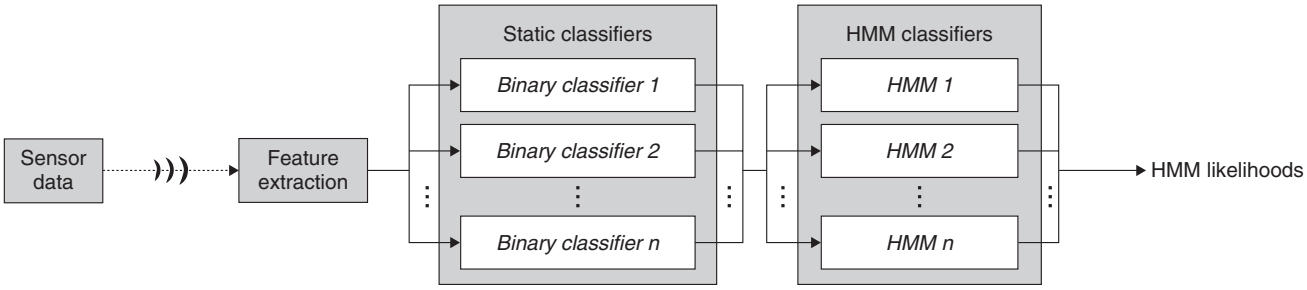


Fig. 7. Method of recognizing human activity.

6. Presentation of sensor data

Collected information or interpreted sensor information is eventually presented to a user. Research is being conducted on presenting information by automatically contributing information from sensors attached to objects to a weblog (blog). The presentation of such automatically contributed information can be made clearer and more natural by personifying the objects. Moreover, nowadays, microblogs and social networks like Twitter and Facebook are recognized as a basis for information distribution. For example, the sensor nodes described in section 4 will be able to tweet the current temperature.

7. Conclusion

People can now be provided with various types of information by a system that collects information from many sensors and various kinds of sensor data and interprets it. However, at present, sensor information is not unified, so while helpful information can be offered to a user, many elements requiring further research and development still remain. Moreover, there is also the issue of personal information privacy and security in connection with information collection. From now on, research will move toward easy-to-use sensor network technology that is intelligible to people.

References

[1] Y. Kishino, Y. Sakurai, K. Kamei, T. Maekawa, Y. Yanagisawa, and T. Okadome, "Efficient Data Gathering for Hierarchical Sensor Networks," IPSJ Trans. on Databases (TOD), Vol. 3, No. 4, pp. 82–93, 2010.

[2] T. Maekawa, Y. Kishino, Y. Sakurai, and T. Suyama, "Recognizing the Use of Portable Electrical Devices with Hand-worn Magnetic Sensors," Proc. of the 9th International Conference on Pervasive Computing (Pervasive 2011), pp. 276–293, San Francisco, CA, USA.

Takayuki Suyama
Senior Research Scientist, Supervisor, Learning and Intelligent Systems Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.
He received the B.E. and M.E. degrees in mechanical engineering from Osaka University in 1990 and 1992, respectively, and the Ph.D. degree in informatics from Kyoto University in 2007. He joined NTT Communication Science Laboratories in 1992 and studied a high-level synthesis system for hardware. During 1999–2003, he worked for the Research and Development Center of NTT WEST. He returned to NTT Communication Science Laboratories in 2003. His research interests are sensor network systems and distributed systems. He is a member of IEEE and the Information Processing Society of Japan (IPSJ).

Yutaka Yanagisawa
Senior Research Scientist, Learning and Intelligent Systems Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.
He received the B.E., M.E., and Ph.D. degrees in information system engineering from Osaka University in 1994, 1996, and 1998, respectively. He joined NTT Basic Research Laboratories in 1998 and studied a moving object database system. During 2009–2010, he was active in developing a thin-client service and SaaS service in NTT WEST. He moved to NTT Communication Science Laboratories and studied sensor network systems until 2010. He is a member of IEEE and IPSJ.

MM-Space: Recreating Multiparty Conversation Space by Using Dynamic Displays

Kazuhiro Otsuka

Abstract

This article introduces our system, called MM-Space, which can recreate a multiparty conversation space at a remote place. This system features a novel visualization scheme that represents human head motions by controlling the poses of the screens displaying human facial images. This physical augmentation helps viewers understand more clearly the behaviors of remote conversation participants such as their gaze directions and head gestures. The viewer is expected to experience a more realistic feeling of the presence of the remote people.

1. Introduction

Face-to-face conversation is one of the most basic forms of communication in daily life, and group meetings are used for conveying and sharing information, understanding others' intentions and emotions, and making decisions. To support communication among remote places, videoconferencing systems have been developed and are widely used. However, they still feel unnatural and uncomfortable. To resolve the problems that have arisen in communications between remote places, which may be not only spatially but also temporally separated, NTT Communication Science Laboratories believes that it is important to deeply understand the mechanism of human-to-human communication and answer questions such as "How do we communicate with each other and what kinds of messages are exchanged by what types of behaviors?" On the basis of this concept, my colleagues and I have been conducting conversation scene analysis for multiparty face-to-face communications [1]. We are trying to extend it toward designing better future communication systems, and we have begun representation/visualization research on multimodal telecommunication and telepresence environments. As the first step, we targeted the problem of reconstructing/recreating the conversation space of a conversation that was originally held at a different location and different time and enabling viewers to visualize the conversation scene as if the people were talking in front them. This article overviews our novel representation scheme and a prototype system after reviewing some of the background.

2. Research progress in conversation scene analysis

In face-to-face conversations, people exchange not only verbal information, but also nonverbal information expressed by eye gaze, facial expressions, head motion, hand gestures, body posture, prosody, etc. Psychologists have suggested that such nonverbal information and behaviors play important roles in human communications [2]. Conversation scene analysis aims to understand human communication though these types of nonverbal information, which is captured by multimodal sensing devices such as cameras and microphones. The goal is to provide an automatic description of conversation scenes in terms of 6W1H, namely who, when, where, whom, what, why, and how. By combining some 6W1H information, we can define a number of problems from low-level (close to physical behavior) ones to high-level (contextual and social level) ones.

Let us consider some examples that NTT Communication Science Laboratories (CS Labs) has targeted.

"Who is speaking when?" is the most essential question: it is called speaker diarization [3]. The estimation of "Who looks at whom and when?" is also called the problem of the visual focus of attention [4], [5]. "Who is talking to whom and when?" is a question about the conversation structure [4]. "Who responds to whom and how?" is related to the problem of interaction structure estimation [6]. As a higher-level problem, "Who feels empathy/antipathy with whom?" is a question about inter-personal emotion [7]. "Who speaks what?" is known as the speech recognition problem [8]. For each of these problems, NTT CS Labs has devised automatic detection, recognition, or estimation methods.

In addition, NTT CS Labs developed the first real-time system for multimodal conversation analysis, which can automatically analyze multiparty face-to-face conversations in real time [9]. This system targets small-scale round-table meetings with up to eight people and uses a compact omnidirectional camera-microphone device, which captures audio-visual data around the table. From the audio-visual data, this system can estimate "who is talking" and "who is looking at whom" and display them on screen. The latest system has added speech recognition and displays "who speaks what" in semi-realtime [8].

3. MM-Space: Reconstructing conversation space by using physical representation of head motions

Beyond the analysis research toward future communication systems, we have recently devised a novel representation scheme and made a prototype system, called MM-Space [10], [11]. It aims to reconstruct multiparty face-to-face conversation scenes in the real world. The goal is to display and playback recorded conversations as if the remote people were talking in front of the viewers. The key idea is a novel representation scheme that physically represents human head motions by movements of the screens showing facial images of the conversation participants. This system consists of multiple projectors and transparent screens attached to actuators. Each screen displays the life-sized face of a different meeting participant, and the screens are spatially arranged to recreate the actual scene. The main feature of this system is *dynamic projection*: the screen pose is dynamically controlled in synchronization with the actual head motions of the participants to emulate their head motions, including head turning, which typically accompanies shifts in visual attention, and head nodding. We expect physical screen movement with image motion to increase viewers' understanding of people's behaviors. In addition, we expect background-free human images, which are projected onto the transparent screens, to be able to enhance the presence of the remote people. Experiments suggest that viewers can more clearly discern the visual focus of attention of meeting participants and more accurately identify who is being addressed.

The idea of this system comes from the importance of nonverbal behavior and nonverbal information in human conversations. This nonverbal information, which is exchanged in a face-to-face setting, cannot be fully delivered in a remote communication setting, e.g., videoconferencing. This insufficiency causes the unnaturalness of telecommunication environments. On the basis of this perspective, we introduced the key idea that physical representation of such nonverbal behaviors, especially head motions, can enhance the viewers' understanding of remote conversations. The nonverbal information expressed with the head motions includes the visual focus of attention, i.e., "who is looking at whom". Its function includes monitoring others, expressing one's attitude or intention, and regulating the conversation flow by taking and yielding turns [12]. A human tends to seize upon a target of interest at the center of his or her visual field: you orient your head toward the target, and various head poses appear according to the relative spatial position to the target. Our previous work indicated that interpersonal gaze directions can be correctly estimated with an accuracy of about 60–70% from head pose information and the presence or absence of utterances without actual eye-gaze directions. In addition, head gestures such as nodding, shaking, and tilting are important nonverbal behaviors. The speaker's head gesture is considered to be a sign of addressing or questioning, and the hearer's head gestures indicate listening, acknowledgement, agreement or disagreement, and level of understanding. Our system can replicate head gestures as physical motions of a screen, which gives viewers a stronger sensation of the presence of the meeting participants.

The physical representation of head motions is also related to human visual perception called *biological motion* [13]. Humans tend to anthropomorphize lifeless objects by assigning social context to movements, even if the objects are simple geometric shapes such as points and triangles. On the basis of the above findings, we hypothesize that realistic

Fig. 1. Overview of our system (MM-Space). (a) Reconstructed space and (b) actual meeting scene.

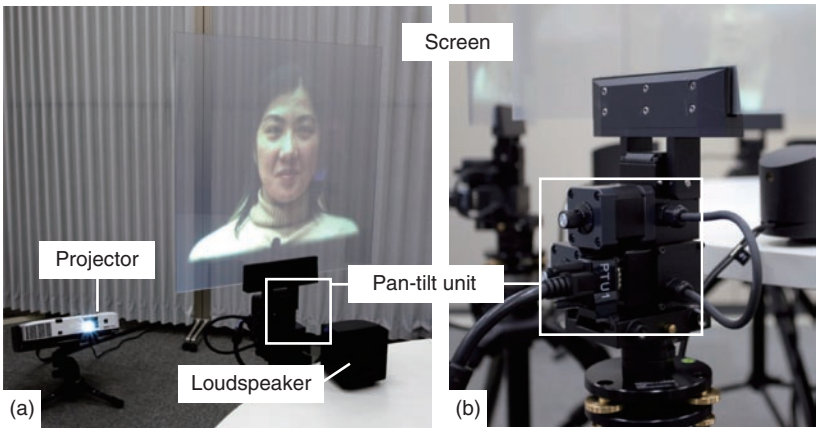


Fig. 2. Devices used in MM-Space.

kinematics offers strong cues for better understanding human communications, regardless of how it is implemented. Therefore, rather than pursuing realistic three-dimensional shape reproduction, our approach uses simple square screens and reproduces physical head motions to produce an augmented expression modality. We expect that combining this physical motion with image motion will boost the viewer's understanding. Moreover, it is known that human vision is highly sensitive to motion in the peripheral field. Therefore, viewers can perceive the entire conversation space including not only a person in front of them, but also the behaviors and presence of people located on their left or right.

3.1 System configuration

An overview of the MM-Space system is shown in **Fig. 1**. The reconstructed conversation space is shown in Fig. 1(a) and the actual conversation scene is shown in Fig. 1(b). In an actual conversation, multiple cameras and microphones capture the participants' faces and voices, respectively. In the reconstructed scene, multiple screens, projectors, actuators, and loudspeakers are placed to recreate the actual seating arrangement. Each participant's face is displayed on a flat transparent screen whose pose is dynamically changed in sync with his or her head motion. Each person's voice is play backed from the loudspeaker located in front of the screen displaying that person, so viewers can identify the speaker's position not only visually, but also aurally. The system described here was created to verify the effectiveness of head motion representation, so here we focus only on offline playback, but we plan to extend it to realtime telecommunication. In addition, our system provides a novel research platform for conversation analysis.

The screen, projector, and actuators are shown in **Fig. 2**. Each screen is highly transparent but includes

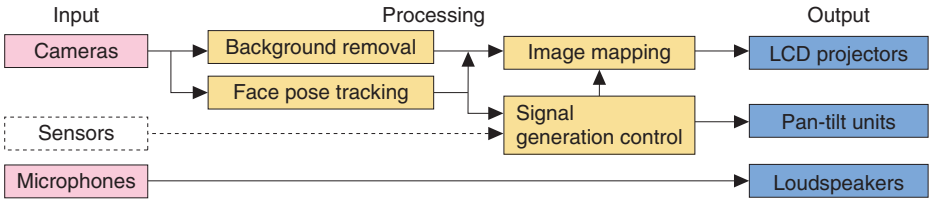


Fig. 3. Block diagram of MM-Space.

a diffusive material that catches the projector's output and makes it visible to the viewer. Each screen has its own liquid crystal display (LCD) projector behind it. Each screen is supported by an actuator, called the pan-tilt unit, that provides rotational motion in both the horizontal and vertical directions. We call this display device a *dynamic display*.

A block diagram of MM-Space is shown in **Fig. 3**. The processing parts provide visual face tracking, background removal, image mapping, and control signal generation. Visual face tracking measures the head position and pose of the meeting participants. Background removal creates images that emphasis the participants. The control signals drive the actuators holding the screens to reflect the participants' face poses, which are measured with visual face tracking and/or motion capture devices. Image mapping generates projected images that are skew-free.

3.2 Effectiveness

As the effect of the motion representation by dynamic screens, we hypothesized that a viewer can more clearly understand the gaze directions of meeting participants as well as who they are addressing. To verify this hypothesis, we compared two different conditions—with and without the motion representation—in terms of identification accuracy. Experimental results indicate that viewers could indeed more clearly recognize the gaze direction of meeting participants when the screens moved. In addition, the results statistically support the hypothesis that viewer can more accurately identify who is being addressed when the screens moved.

4. Conclusions and future perspective

This article introduced our system MM-Space for recreating a conversation space at different times and places. Its key feature is the physical representation of head motion as an additional expression modality. The synergy of physical screen motion and image motion on the screen is expected to boost our perception of social interactions involving the visual focus of attention. MM-Space is expected to be extended to realtime communication systems that can connect people located at different places. For that purpose, it will be necessary to explore in more detail the characteristics of the motion representation and evaluate how it can contribute to better expression and perception of addressing others and being addressed by others. In addition, other problems include the optimum camera configuration and the latency of telecommunications and physical systems. Finally, we believe that MM-Space will be a useful research platform for designing better communication systems and analyzing and understanding the mechanism of human communications.

References

[1] K. Otsuka, "Conversation Scene Analysis," IEEE Signal Processing Magazine, Vol. 28, No. 4, pp. 127–131, 2011.
[2] M. Argyle, "Bodily Communication—2nd ed. Routledge, London and New York, 1988.
[3] S. Araki, M. Fujimoto, K. Ishizuka, H. Sawada, and S. Makino, "A DOA based speaker diarization system for real meetings," HSCMA2008, pp. 29–32, 2008.
[4] K. Otsuka, Y. Takemae, J. Yamato, and H. Murase, "A Probabilistic Inference of Multiparty-Conversation Structure Based on Markov-Switching Models of Gaze Patterns, Head Directions, and Utterances," Proc. of the 7th International Conference on Multimodal Interfaces (ICMI'05), pp. 191–198, Trento, Italy, 2005.
[5] S. Gorga and K. Otsuka, "Conversation scene analysis based on dynamic Bayesian network and image-based gaze detection," Proc. of the ICMI-MLMI'10 International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction, Article No. 54, Beijing, China, 2010.
[6] K. Otsuka, H. Sawada, and J. Yamato, "Automatic inference of cross-modal nonverbal interactions in multiparty conversations: "Who responds to whom, when, and how?" from gaze, head gestures, and utterances," Proc. of the 9th International Conference on Multimodal Interfaces (ICMI'07), pp. 255–262, Nagoya, Japan, 2007.
[7] S. Kumano, K. Otsuka, D. Mikami, and J. Yamato, "Analyzing empathetic interactions based on the probabilistic modeling of the co-occurrence patterns of facial expressions in group meetings," Proc. of the 9th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2011), pp. 43–50, Santa Barbara, CA, USA, 2011.
[8] T. Hori, S. Araki, T. Yoshioka, M. Fujimoto, S. Watanabe, T. Oba, A.

Ogawa, K. Otsuka, D. Mikami, K. Kinoshita, T. Nakatani, A. Nakamura, and J. Yamato, "Low-Latency Real-Time Meeting Recognition and Understanding Using Distant Microphones and Omni-Directional Camera," IEEE Trans. on Audio, Speech & Language Processing, Vol. 20, No. 2, pp. 499–513, 2012.
[9] K. Otsuka, S. Araki, K. Ishizuka, M. Fujimoto, M. Heinrich, and J. Yamato, "A Realtime Multimodal System for Analyzing Group Meetings by Combining Face Pose Tracking and Speaker Diarization," Proc. of the 10th International Conference on Multimodal Interfaces (ICMI'08), pp. 257–264, Chania, Crete, Greece, 2008.
[10] K. Otsuka, S. Kumano, D. Mikami, M. Matsuda, and J. Yamato, "Reconstructing multiparty conversation field by augmenting human head motions via dynamic displays," Proc. of the 2012 ACM Annual Conference (CHI EA'12), pp. 2243–2248, Austin, TX, USA, 2012.
[11] http://www.brl.ntt.co.jp/people/otsuka/ACM_MM2011.html
[12] A. Kendon, "Some functions of gaze-direction in social interaction," Acta Psychol., Vol. 26, pp. 22–63, 1967.
[13] G. Johansson, "Visual perception of biological motion and a model for its analysis," Attention, Perception, & Psychophysics, Vol. 14, No. 2, pp. 201–211, 1973.

Kazuhiro Otsuka
Senior Research Scientist, Distinguished Researcher, Media Information Laboratory, NTT Communication Science Laboratories.

He received the B.E. and M.E. degrees in electrical and computer engineering from Yokohama National University, Kanagawa, in 1993 and 1995, respectively, and the Ph.D. degree in information science from Nagoya University, Aichi, in 2007. He joined NTT Human Interface Laboratories in 1995. He moved to NTT Communication Science Laboratories in 2001. He stayed at Idiap Research Institute, Switzerland, as a distinguished invited researcher in 2010. He has been a distinguished researcher at NTT since 2010. His current research interests include communication science, multimodal interactions, and computer vision. He received the Best Paper Award of the Information Processing Society of Japan (IPSJ) National Convention in 1998, the IAPR Int. Conf. on Image Analysis and Processing Best Paper Award in 1999, the ACM Int. Conf. on Multimodal Interfaces 2007 Outstanding Paper Award, the Meeting on Image Recognition and Understanding (MIRU) 2009 Excellent Paper Award, the Institute of Electronics, Information and Communication Engineers (IEICE) Best Paper Award 2010, the IEICE KIYASU-Zen'iti Award 2010, the MIRU2011 Interactive Session Award, and JSAI Incentive Award 2011. He is a member of IEEE, IPSJ, and IEICE.

Hearing Sound Alters Seeing Light

Takahiro Kawabe

Abstract

This article introduces three audiovisual illusions in which hearing sound alters how light is seen: the audiovisual tau effect, sound-induced visual motion, and attenuation of the sense of agency by sounds. For each of them, we discuss a specific psychophysical mechanism called perceptual grouping that commonly underlies these illusions. Finally, we describe how scientific understanding of the audiovisual mechanism is contributing to future information technologies.

1. Introduction

The brain tries to comprehend the external world by interpreting sensory signals from each of the five senses—sight, hearing, touch, smell, and taste—which are also known by the scientific names: vision, audition, tactition, olfaction, and taste. These senses do not always work independently of each other. There have been many reports of psychophysical studies finding that an interpretation of sensory signals in one modality is strongly altered by sensory signals in the other modality when two senses are simultaneously stimulated. How are these senses able to interact with each other? We still have no clear answer to this question.

This article focuses on the interaction between audition and vision, and it introduces several illusions that reflect this interaction. We also discuss the idea that perceptual grouping is a fundamental psychophysical mechanism underlying a broad range of audiovisual interaction. Finally, we describe how an understanding of the psychophysical mechanism for sensory integration can contribute to future information technologies in a novel and user-friendly manner.

2. Vision and audition in space and time

Older studies assumed that vision is superior to audition in space perception and that audition is superior to vision in time perception (modality appropriateness hypothesis [1]). However, recent studies suggest that this assumption is not always accurate. Instead, the brain probably weights auditory and visual signals according to the precision of each signal and judges an audiovisual event by relying on the weighted average of the sensory signals.

Recent studies have also reported an audiovisual illusion where vision and audition interact with each other beyond space and time. For example, when a sequence of flashes of light at constant space and time intervals (**Fig. 1(a)**) is accompanied by a sequence of tones at a non-constant time interval (**Fig. 1(b)**), the flash sequence is perceived to have non-constant space and time intervals (**Fig. 1(c)**). This illusion is called the audiovisual tau effect [2]. Auditory time can distort visual time judgments, and the distorted visual time judgments also alter the visual space judgments. Thus, the audiovisual tau effect is a kind of bypassed distortion of visual spatial judgments by auditory temporal information.

Interestingly, the effect is observed even in infancy. When infants are exposed to a novel stimulus, they continue to gaze upon it for a while. However, when they are exposed to a familiar stimulus, they get bored and exhibit less frequent gazes at it. By exploiting this characteristic of infants, a previous study [3] investigated whether the audiovisual tau effects occurs in infancy. Specifically, infants were exposed to an audiovisual stimulation that invokes the compelling audiovisual tau effect in adults. After a prolonged exposure to the stimulation, the infants were simultaneously presented with two spatial patterns: one involving the audiovisual tau effect and the other not involving it. The durations of gazes at each pattern were measured and it was found that the infants looked at the spatial pattern free of the audiovisual tau effect for less time than the spatial pattern with the

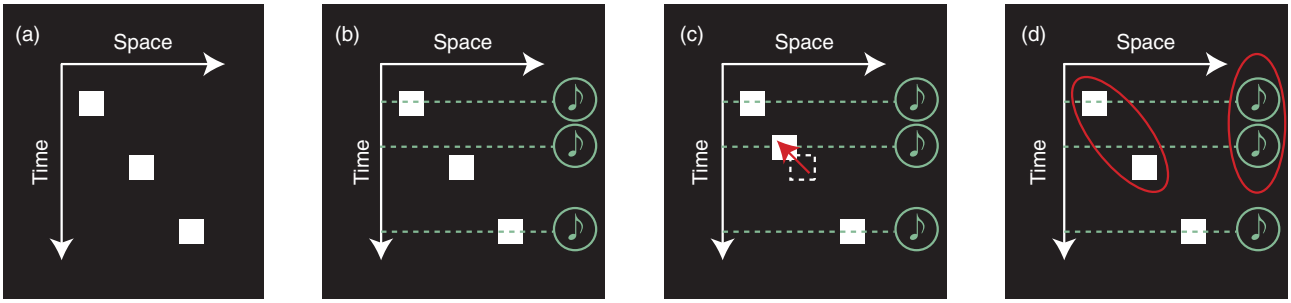


Fig. 1. Audiovisual tau effect.

effect, indicating that the audiovisual tau effect does occur even in infancy. These results indicate that audiovisual integration related to the audiovisual tau effect is located at an early level of processing in the brain, which is free from a knowledge-based interpretation of the relationship between audiovisual signals because it is implausible to assume that infants utilize some kind of knowledge.

3. Perceptual grouping across senses

The initial interpretation was that the audiovisual tau effect occurred because visual time representation in the brain was modulated by sounds. On the other hand, no study has ever demonstrated what kind of change in neural representation corresponds to the change in subjective experiences of time. Therefore, we should propose an explanation for the audiovisual tau effect without assuming modulation of the neural time presentation through audiovisual integration.

We think that perceptual grouping is a promising idea for sufficiently explaining the audiovisual tau effect. Perceptual grouping refers to one of the brain's strategies for treating elements with high similarity or high proximity as a group. As shown in Fig. 1(b), when two among three tones are temporally proximate, they are perceptually grouped together and the remaining one is perceptually assigned to another group.

We think that perceptual grouping in the temporal dimension affects perceptual grouping in the spatial dimension (**Fig. 1(d)**). Previous studies have demonstrated that elements within the same perceptual group were judged to be spatially proximate more than elements each belonging to different perceptual groups even though the physical spatial distance was identical [4]. Temporal grouping in audition might dictate spatiotemporal grouping in vision, and this might cause the misjudgment of spatial intervals between flashes, leading to the audiovisual tau effect.

4. Sound alters visual motion perception

When viewing the alternate presentation of two flashes at the right and left sides of a display, we perceive the flashes to alternately move leftward and rightward (see **Fig. 2(a)**). This is a visual illusion called apparent motion. We see moving objects in the display of television; this is due to apparent motion. A recent study [5] discovered an interesting phenomenon whereby the apparent motion direction was subjectively fixed by presenting sounds with a non-uniform temporal interval (**Figs. 2(b)** and **2(c)**). Specifically, a motion direction made of flashes that were temporally proximate to sounds tended to be dominant over the other. This phenomenon has been explained in terms of neural timing modulation. Specifically, sound timing might alter the timing of visual flashes, and the brain might determine the motion direction by relying on the altered visual timing.

On the other hand, we believe that this phenomenon can also be explained with perceptual grouping [6]. Our explanation is shown in **Fig. 2(d)**. When temporally proximate sounds are grouped together (as represented by dotted ellipses), the grouping of sounds dictates the temporal grouping between flashes, and this consequently alters the spatiotemporal grouping between flashes, leading to a unidirectional apparent motion. The explanation based on perceptual grouping does not require modulation of the visual timing by sounds. Rather, the judgment for spatiotemporal *nearness* between flashes is affected by the grouping of sounds.

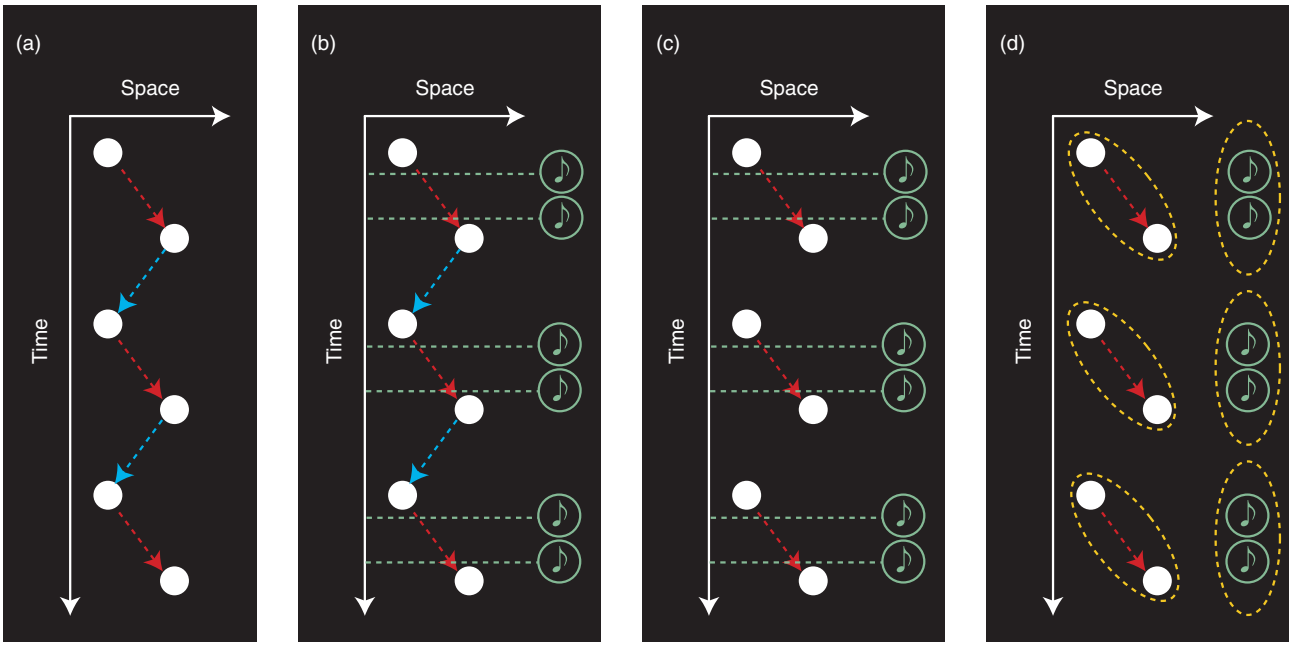


Fig. 2. Change in appearance of apparent motion direction caused by adding sounds.

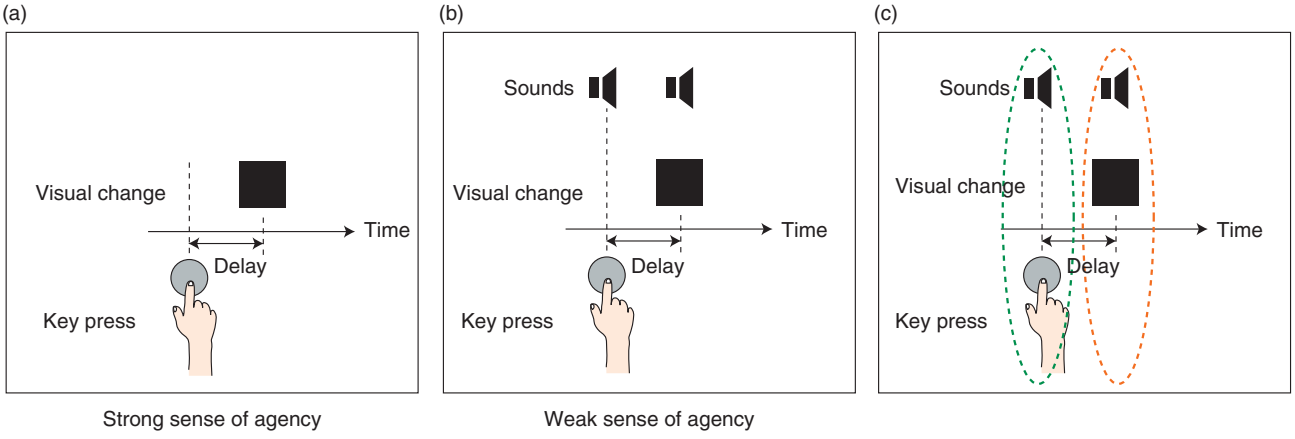


Fig. 3. Sound alters the sense of agency.

5. Sound alters the sense of agency

When you press a key on a computer keyboard, a visual change is often triggered on the display. In this situation, you feel as if there is a causal relationship between the key press and the visual change on the display. Counterintuitively, however, the sensation of causal relationship is also an illusion (or a product of mental processing). The sensation disappears when a large temporal delay is inserted between the key press and the visual change on the display even though there is still a mechanical causal relationship between them. The sensation that an agent's action triggers a change in the external world is called the sense of agency (**Fig. 3(a)**) [7].

We found that adding sounds reduced the sense of agency [8]. Specifically, when both the key press and the visual change on the display were accompanied

by pure tones, the sense of agency decreased (**Fig. 3(b)**). However, the sense of agency was not affected when only one action—either the key press or the visual change on the display—was accompanied by a pure tone.

We would like to interpret the modulation of the sense of agency by sounds in terms of perceptual grouping. When both the key press and visual change are accompanied by tones, the first tone is probably grouped together with the key press while the second tone is grouped together with the visual change (**Fig. 3(c)**). Thus, the key press belongs to a different perceptual group from the visual change, and this may result in the reduction in the sense of agency.

So far, one previous study has proposed a sensory-motor model for the sense of agency [9]. Specifically, the model assumes that the brain first predicts the outcome of an agent's action and then compares the prediction with actual sensory feedback. If the consistency between the prediction and actual feedback is low, the sense of agency is also decreased.

On the other hand, our results indicate that the sense of agency involves not only sensory-motor processing but also sensory processing. In particular, perceptual grouping is a key mechanism for the sense of agency. Investigating the interaction between sensory-motor and sensory processing will lead to a better understanding of the mechanism underlying the sense of agency.

6. Conclusion

My colleagues and I in our research group are trying to clarify psychophysical mechanisms for the integration of multimodal sensory signals. The clarified mechanism may be used to develop and/or improve various types of user interfaces. For example, the addition of sounds to silent video might enable us to control subtle visual representation. The temporal resolution of video in one-seg* or augmented reality is often coarse, and the resultant subjective smoothness of the video is not high enough. Adding appropriate sounds to these videos may improve the smoothness of object motion in them. Moreover, we may be able to control the sense of agency by adding sounds in the interface. By applying the phenomenon that we have found, users can separate the information relevant to their action from irrelevant information. Through an understanding of the principle of information processing in the brain, we would like to propose unique, interesting, and user-friendly information technology and devices in future.

* One-seg. A TV broadcasting service using one segment of the thirteen used for a high-definition television broadcast signal [10], [11].

References

[1] R. B. Welch and D. H. Warren, "Immediate perceptual response to intersensory discrepancy," Psychological Bulletin, Vol. 88, No. 3, pp. 638–667, 1980.

[2] T. Kawabe, K. Miura, and Y. Yamada, "Audiovisual tau effect," Acta Psychologica, Vol. 128, No. 2, pp. 249–254, 2008.

[3] T. Kawabe, N. Shirai, Y. Wada, K. Miura, S. Kanazawa, and M. K. Yamaguchi, "Audiovisual tau effect in infancy," PLoS ONE, Vol. 5, No. 3, e9503, 2010.

[4] S. Coren and J. S. Girgus, "Principles of perceptual organization and spatial distortion: The gestalt illusions," Journal of Experimental Psychology: Human Perception and Performance, Vol. 6, No. 3, pp. 404–412, 1980.

[5] E. Freeman and J. Driver, "Direction of visual apparent-motion driven solely by timing of a static sound," Current Biology, Vol. 18, No. 16, pp. 1262–1266, 2008.

[6] W. Roseboom, T. Kawabe, and S. Nishida, "Changes in visual apparent motion direction by cross-modal interaction are not dependent on temporal ventriloquism," Poster presented at the Vision Sciences Society Annual Meeting, Naples, FL, USA, May 2012.

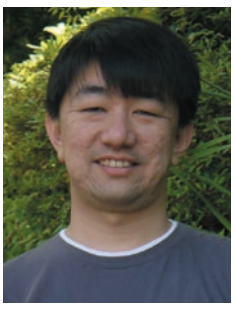
[7] P. Haggard and V. Chambon, "Sense of agency," Current Biology, Vol. 22, No. 10, pp. R390–R392, 2012.

[8] T. Kawabe, W. Roseboom, and S. Nishida, "Crossmodal perceptual grouping modulates subjective causality between action and outcome," Poster presented at Asia-Pacific Conference on Vision, Incheon, Korea, July 2012.

[9] C. D. Frith, S. -J. Blakemore, and D. M. Wolpert, "Abnormalities in the awareness and control of action," Philosophical trans. of the Royal Society of London, Series B, Vol. 355, No. 1404, pp. 1771–1788, 2000.

[10] One-seg. NTT DOCOMO. http://www.nttdocomo.co.jp/english/service/entertainment/1seg/index.html

[11] One-seg. Wikipedia. http://en.wikipedia.org/wiki/1seg



Takahiro Kawabe

Research Specialist, Sensory Representation Research Group, Human Information Science Laboratory, NTT Communication Science Laboratories.

He received the Ph.D. degree in psychology from Kyushu University, Fukuoka, in 2005. Since joining NTT Communication Science Laboratories in 2011, he has been studying cross-modal integration between vision and audition and multimodal integration to cause the sense of agency. He is also interested in material perception of fluids. He is a member of the Vision Science Society.

ICT Infrastructure Technology Underlying Business & Service Innovation

Yoichi Kihara

Abstract

This article explains how information and communications technology (ICT) infrastructure technology will help corporations overcome the new challenges facing them as the business management environment changes rapidly and how it will support business and service innovation. Examples of actual projects handled by NTT DATA are given and future technological prospects are discussed.

1. Introduction

The information and communications technology (ICT) strategies of companies are closely intertwined with their business strategies, and ICT is now indispensable to a wide range of business activities. The business management environment is currently undergoing various changes including shorter product life cycles in response to more diverse and complicated market demands, rapid globalization due to changes in the domestic and international economic situations, stricter legal compliance, business continuity plans for the aftermath of the Great East Japan Earthquake, and energy saving efforts. To keep up with such changes, it is essential to update the ICT systems that support business management. It is even said that the speed of ICT system updates determines whether a business is successful or not. However, the capabilities of current ICT systems are insufficient to meet these corporate demands.

Although business growth and customer expansion have been becoming a higher priority than cost reduction in both domestic and international business since last year, cost and business processes improvements still remain as important issues for many companies. While global corporations place higher priority on developing new products and services (innovation), Japanese corporations regard the cultivation of new markets and business expansion to wider areas as important. This trend in Japan suggests that a global rollout is urgently required to support the international business of Japanese corporations [1].

2. ICT infrastructure technology to overcome business challenges

Key concepts to help corporations overcome their challenges are streamlining, agility, globalization, and value creation [2]. To implement these concepts, NTT DATA is focusing on the following three ICT infrastructure technologies.

(1) Cloud computing technology
(2) Robotics integration technology
(3) Communication advancement technology

Cloud computing technology features resource sharing, on-demand self-service, and speedy scalability, all of which lead to streamlining and agility. It is also helpful for accelerating globalization because it is designed to be used via a wide range of network access methods. Communication advancement technology is expected to contribute to globalization by bridging communication gaps caused by remote locations and different languages. It will also create new value by enabling not only person-to-person communication but also person-to-machine communication. Robotics integration technology is related to the Internet of Things (IoT) and controls the actual machines by capturing changes in the business environment in real time through advanced sensing technology. This is called M2M2A (machine to machine

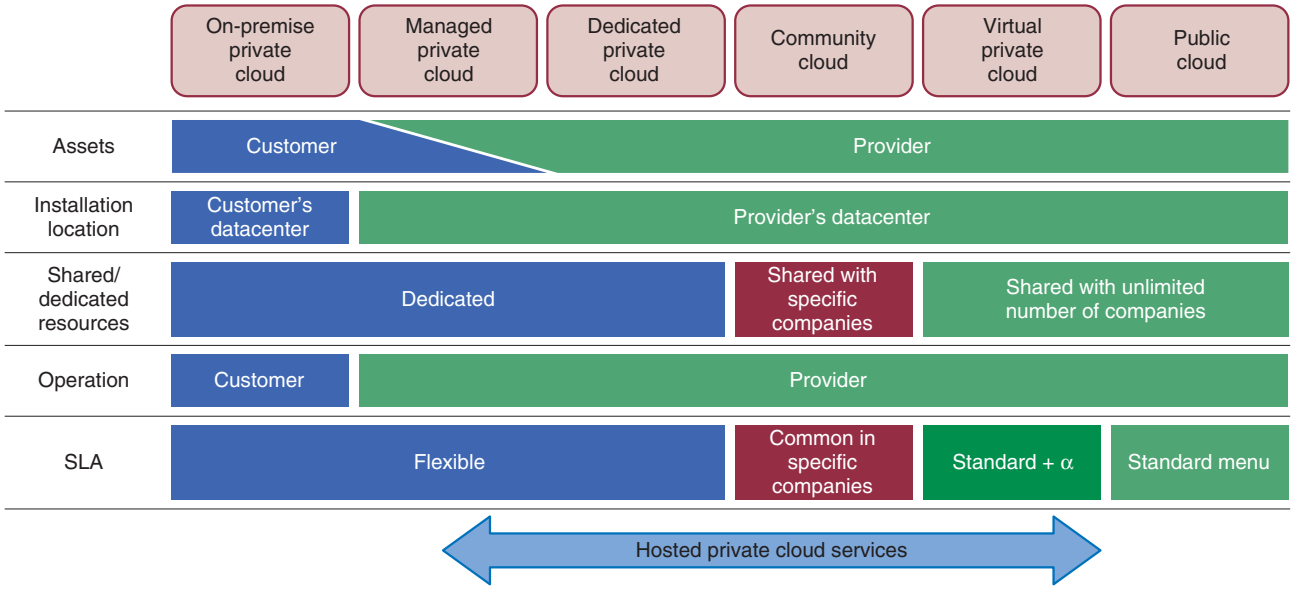


Fig. 1. Various types of private cloud are available.

to actuator), an outgrowth of M2M (machine to machine). Robotics integration technology not only gathers real-world information, but can also provide feedback to the real world based on the collected information. This should lead to the creation of new value.

2.1 Cloud computing technology

Prior to the Great East Japan Earthquake of 2011, it was generally believed that core business systems should remain within private clouds because of security and service level agreement (SLA) issues, while peripheral business could be migrated to public cloud services. However, this earthquake changed opinions: more and more corporations are now transferring their existing business applications to public clouds, and not only non-core applications but also business-to-consumer (B2C) applications and developmental test environments are more frequently being migrated to infrastructure as a service (IaaS) or platform as a service (PaaS) environments.

Many companies are performing server integration of their intra-company ICT systems through virtualization for efficient resource usage. However, such attempts tend to end up with mere hardware integration, so the load of operation management tasks is not reduced. They even complicate isolating the problem in the virtual environment. A solution would be a private cloud with enhanced flexibility, speed, and operability. However, such a cloud requires additional investment and new skills to handle new technologies. This is why migration to private clouds is limited to advanced corporations.

Private clouds have generally been on-premise[*1] deployments so far, but new types of private cloud services, in which users can choose the location of facilities and whether or not to share resources, are now available (**Fig. 1**). Such new private clouds overcome concerns related to the conventional type of private clouds. In the future, corporations will be able to use both private and public clouds either on their own premises or hosted elsewhere for different purposes in the most suitable manner. Furthermore, this mixed usage of private and public clouds will develop into an inter-related hybrid cloud.

Four of the five companies shown in the leaders quadrant by Gartner in its Dec. 2011 report "Magic Quadrant for Public Cloud Infrastructure as a Service" [3] use a cloud infrastructure built using proprietary software[*2]. On the other hand, IaaS, a new public cloud based on an open source software (OSS) cloud infrastructure such as NTT Communications' Cloud^n and HP's Cloud Services, came into full

*1 On-premise deployment: A company operates and manages software, servers, and network equipment prepared by itself on its own premises.

*2 Proprietary software: Software that is legally and technically limited in its usage, alteration, and reproduction.

service in the latter half of 2011. Functions offered by the OSS-based cloud infrastructure are becoming mature, and its governance model is also shifting towards a community-driven one from one led by a single company. The OSS-based cloud is getting ready to be used for commercial purposes. Anticipating that private cloud services will also be based on OSS, since the foundation of the public cloud IaaS infrastructure is evidently moving from proprietary software to OSS, NTT DATA is directing its business efforts towards establishing OpenStack, which is OSS-based cloud infrastructure software, and its support services. We are also working together with NTT's research and development laboratories to establish a proof of concept of a cloud built entirely with open source technology—including the hardware, storage, network, and monitoring functions (**Fig. 2**).

Requirements for datacenter networks have changed as follows as server virtualization technology has become common.

(1) When multiple tenants use a cloud, they share the physical networks. Therefore, the networks themselves need to be virtualized and the virtual local area network (VLAN), virtual routing and forwarding (VRF), virtual firewall, and virtual load balancer must be set up interdependently on network devices.

(2) Because virtual machines can be moved among physical servers, the machines' virtual network settings must also be dynamically changeable.

(3) To provide an on-demand cloud service, both virtual machines and their virtual networks must be centrally controlled and managed. Mapping between physical and logical network configurations is also necessary because they are different entities.

(4) Conventional datacenter networks do not allow a loop structure. This limitation means that the network structure must be a series of tree structures, resulting in an inflexible and inefficient network. New technology must be introduced to enable traffic to take different routes to create a highly efficient and flexible network.

Two possible candidates that might meet these requirements are an SDN/OpenFlow-based network (SDN: Software Defined Networking) and a VLAN-based network (**Table 1**). SDN is an approach that enables network configurations and behavior programmable, and OpenFlow is a candidate protocol

Fig. 2. Full Open Source Cloud.

Table 1. Responses to datacenter network requirements.

Network requirements	SDN/OpenFlow-based	VLAN-based
(1) Support for multiple tenants	SDN/OpenFlow + virtual appliances	VLAN/VXLAN/NVGRE + virtual appliances
(2) Support for virtual machine mobility to different physical servers		VEPA/VNTag + virtual chassis + automatic management of port profiles
(3) Automatic operation & centralized operation and server management		Virtual chassis + virtual appliances
(4) Streamlining of network bandwidth usage		TRILL/SPB/MLAG

(1) VXLAN, NVGRE: network virtualization technology or method for Layer 2 over Layer 3. Technologies for logically dividing the physical network among tenants.

(2) VEPA, VNTag: technology or methods to help reduce server loading by substituting external switches for virtual functions in the hypervisor.

(3) TRILL, SPB, MLAG: technology that achieves multipath routing. It enables pathway and device redundancy.

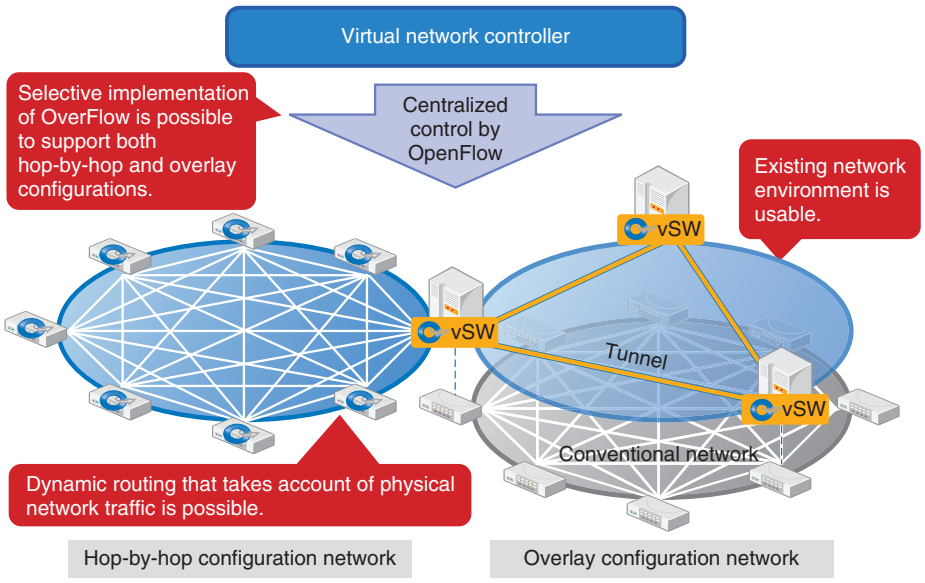


Fig. 3. Virtual network controller that supports hop-by-hop and overlay configurations.

that implements SDN. On the basis of these technologies, we have developed a Virtual network controller, which is an OpenFlow-based controller, that can be used within a datacenter and between datacenters by supporting both the hop-by-hop and overlay configurations [4] (**Fig. 3**). We are currently testing interconnectivity of this controller with switches from major manufacturers through standardization activities.

2.2 Robotics integration technology

As sensor technology and network technology develop, a computer system (or a machine) is becoming able to collect information about the real world by communicating with other machines without any help from human beings. This advance materializes concepts such as IoT and M2M, which improve efficiency, convenience, and sustainability by utilizing such real-world information. In conventional M2M systems, servers for collecting and analyzing information from machines were developed individually. Now, virtualization and cloud technologies have enabled the M2M cloud, in which all the functions such as communication with various devices and the collection, storage, and analysis of data are managed centrally. Examples of M2M clouds are NTT DATA's Xrosscloud and NEC's CONNEXIVE. These are expected to be used for various applications (**Table 2**). M2M clouds have the following features.

(1) Data collection and storage: An M2M cloud has functions for communicating with devices, collecting and storing data, and managing devices. A data collection mechanism is ready within a short time. Moreover, a combination of multiple data sources can be analyzed easily because the large amount of information can be stored across the cloud.

(2) Data analysis: The M2M cloud offers statistical calculations such as multivariate analysis as a service. Users benefit from distributed computing technologies such as the Hadoop Distributed File System (HDFS) and MapReduce[*3] for conducting data analysis without a large capital investment. Since this cloud also allows data trading, it will be possible in the future for a corporation to analyze its own data together with data from others.

The current M2M system is used to collect and analyze data from devices and to visualize the analysis results or utilize the results for data mining. In the future, this system will probably develop into an M2M2A system where the system acts on the real world through devices (actuators) on the basis of

*3 MapReduce: A software framework introduced by Google in 2004 to support distributed computing for massive datasets on clusters of computers.

Table 2. Examples of M2M cloud applications.

Areas	Data to be collected	Possible services
Medicine	Physical information (height, weight, body temperature, blood pressure, pulse), eating information (calories, nutrition), and activities (pedometer counts, walking distance, hours of sleep)	Health management, disease prevention, obesity prevention, diet, diagnosis support, stress measurement, remote monitoring of patients, and life insurance discounts
Agriculture	Weather information (weather, temperature, humidity, hours of sunshine, wind direction, wind speed), density of air (CO_2, O_2), water content in soil, soil fertility, images from cameras	Vegetable factories, automated agriculture (sowing seeds, watering, disinfecting, monitoring, harvesting), controlling shipping timing, automated gardening, and trading of vegetation information
Energy	Power consumption amount, water consumption amount, gas consumption amount, power generation amount, amount of stored energy, power outage information, water supply control information, and gas supply control information	Smart grid, BEMS, remote meter reading, remote control of energy consumption, control of energy storage, and support for creating power generation plan
Transportation	Traffic information (locations, speeds, distances between vehicles, breakdowns), people movement information, actual and scheduled public transportation movements, EV charging station information (locations & usage status)	Traffic congestion prediction, public transportation rush prediction, freight management, emergency warnings concerning failure or accident, EV charging station availability, and support for city planning
Construction	Positioning, torsion & rotation, vibration & shaking, tilting, erosion, freezing, and contamination.	Structural monitoring (buildings, pylons, bridges, highways), road monitoring, monitoring of underground pipes (for electricity, water, drains, gas), and monitoring of high-voltage electricity cables (current leakage, cable disconnection, snowfall)
Home appliances	Equipment information (model & production date), equipment status (on/off, in service or not in service, settings, failures), and operation environment (temperature, humidity, vibration)	Remote product support, remote control of home appliances (air conditioners, videos, feeding of pets), and automatic updating of firmware
Disaster prevention	Weather information (rainfall amount, wind speed, snowfall amount, lightning), water level in rivers, amount of soil carried (by rivers etc.), heat, flames, smoke, gas, seismic intensity, and radiation levels	Prediction of damage from natural disasters, alerts and warnings, creation of hazard maps, and support for creating evacuation plans

BEMS: building management system
EV: electric vehicle

information about the real world gathered from other devices. To make an M2M2A system, we need to add four functions to the current M2M cloud in order to upgrade it into an M2M2A cloud.

- A function for analyzing information from devices and determining how to act on the real world on the basis of the analysis
- A function for choosing which actuators and robots to activate
- A function for creating an optimum command set for delivery to the actuators and robots
- A function for controlling the actuators and robots

We are working to overcome various challenges to achieve a smart society supported by the M2M2A system (**Fig. 4**).

2.3 Communication advancement technology

Apps (applications) with a voice-based interface on smartphones have recently been appearing, including NTT DOCOMO's Shabette Concier (talking concierge) service and call interpretation service and Apple's Siri. Furthermore, communication between things and people is now possible. Such apps are made possible by media analysis technology—such as voice recognition and machine translation—and the functions that the cloud possesses, including processing capability, massive databases used for voice recognition and translation, and service coordination according to the analysis results.

As the corporate environment changes, there are demands to reduce linguistic barriers to help domestic corporations expand their business overseas or manage offshore outsourcing. To help such corporations, NTT DATA is currently developing a tool for creating design documents in Japanese for offshore outsourcers to reduce the burden of creating Japanese documents and improve document quality and a

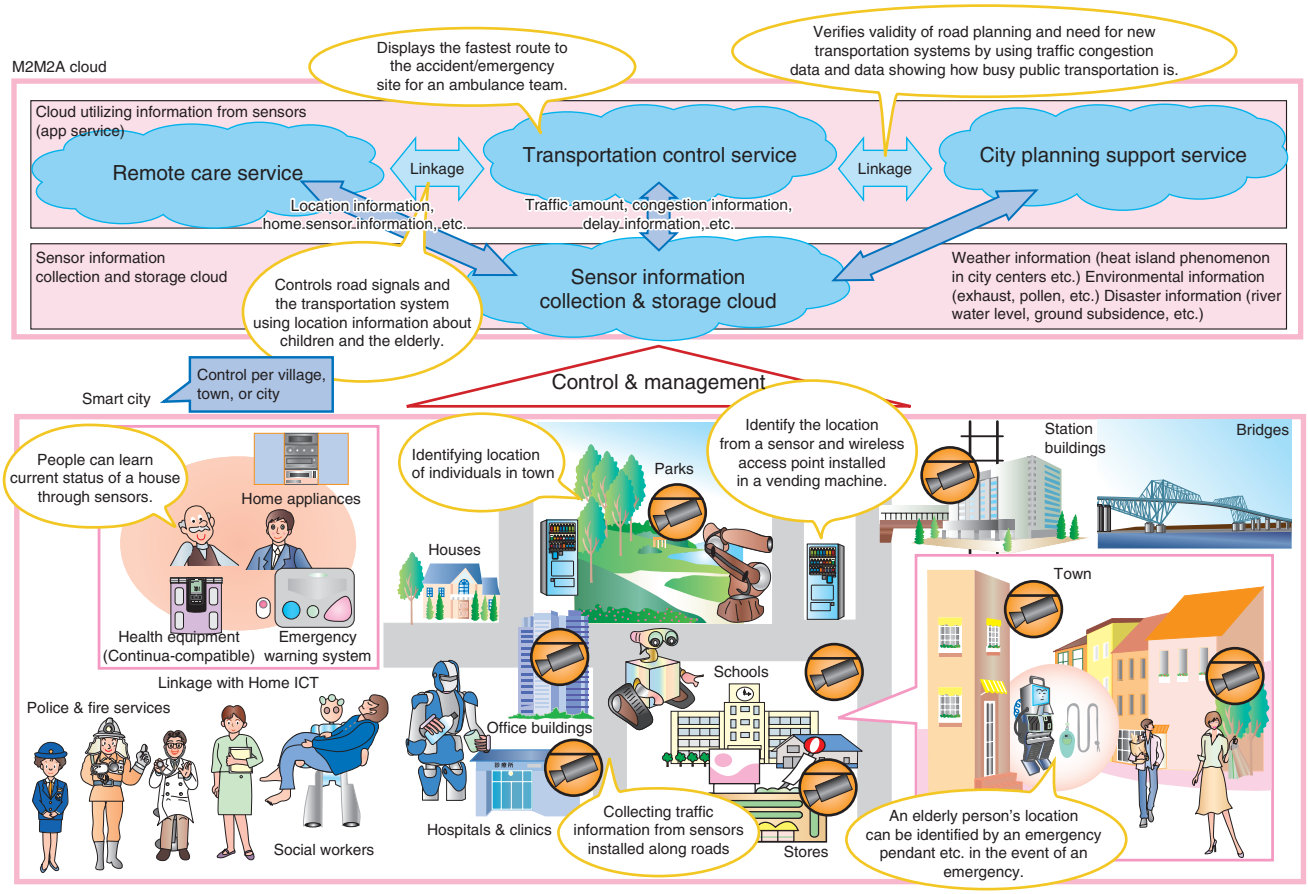


Fig. 4. Smart society achieved using M2M2A systems.

global meeting support system to provide smooth communication in multi-language meetings.

3. Future prospects

ICT systems are expected to contribute to value creation, providing efficiency and speed. To create services with new value, everyone involved in the business including developers and operators must cooperate. This idea has generated the word DevOps [5]. To repeat development and operation in a short cycle, it is necessary to establish a DevOps infrastructure that enables the distribution of applications. We will continue our research and development to create an environment where service providers can offer any services without being aware of the ICT infrastructure itself.

References

[1] Y. Kihara, "Current Movement and Future Perspective of Enterprise Cloud Computing—The pursuit of efficiency and agility, and the value creation by big data practical use—," Technical Report of the Proc. of the Institute of Electronics, Information and Communication Engineers, Vol. 111, No. 408, NS2011-162, pp. 99–102, Jan. 2012 (in Japanese).
[2] Press release (in Japanese). http://www.gartner.co.jp/press/html/pr20120309-01.html
[3] "Magic Quadrant for Public Cloud Infrastructure as a Service," Gartner, Dec. 2011. http://www.savvis.com/en-us/info_center/documents/magic_quadrant_for_public_cloud.pdf
[4] H. Kitazume, T. Koyama, T. Kishi, and T. Inoue, "Network Virtualization Technology for Cloud Services," NTT Technical Review, Vol. 9, No. 12, 2011. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201112fa4.html
[5] Wikipedia. http://en.wikipedia.org/wiki/DevOps

Yoichi Kihara
Executive Manager, Personnel Department, NTT Software Corporation.

He received the B.E. degree in electrical and communication engineering from Waseda University, Tokyo, in 1983. Since joining the Musashino Electrical Communication Laboratories of Nippon Telegraph and Telephone Public Corporation (now NTT), as a research engineer, he developed network and information technology service systems, such as network-based Karaoke, e-commerce, tele-education, 3D web browsing, web recommendation, authentication and accounting, and security system. From 2009 to 2012, he led technology development projects for cloud computing, Internet of things (IoT), and security system in NTT DATA R&D. He became Deputy Senior Executive Manager, Research and Development Headquarters, NTT DATA Corporation in 2010. In 2012, he moved to NTT Software Corporation, where he is currently engaged in human-resource management. He is a member of the Information Processing Society of Japan, and a visiting professor of Shizuoka University.

Latest Trend of OpenStack, Open Source Software for Infrastructure as a Service, and NTT DATA's Activities

Masayuki Hanadate

Abstract

NTT DATA is researching and developing cloud infrastructure technology utilizing OpenStack. This article explains how OpenStack works and introduces some of NTT DATA's recent projects including OpenStack component development, OpenFlow collaboration technology development, cloud security component development, and the use of Swift.

1. OpenStack

1.1 Overview

OpenStack [1] is open source software (OSS) for centrally and efficiently operating and managing physical servers and devices comprising the cloud. It is under collaborative development by more than 160 companies around the world (as of April 26, 2012). The latest version (version 6, Folsom) was released in September 2012. Moreover, OpenStack is included in Ubuntu 12.04LTS, a Linux distribution.

NTT DATA has been participating in the OpenStack project since 2010 as one of the project startup members. It is conducting research and development (R&D), as well as promoting cloud services using OpenStack together with NTT's R&D laboratories.

The service model offered by OpenStack is infrastructure as a service (IaaS), which is a cloud service model suggested by the National Institute of Standards and Technology (NIST) of the USA [2].

OpenStack provides users with virtual machines that run on a hypervisor such as kernel-based virtual machines (KVMs) and XenServers. Any OpenStack users can access their virtual machines through networks and operate their computing resources (e.g., central processing units (CPUs), memory, hard disk drives, and IP (Internet protocol) addresses) allocated to their virtual machines.

The features of OpenStack are as follows:

(1) Multiple tenants

A single physical machine can host multiple virtual machines belonging to different cloud computing users. This reduces redundant computing resources leading to lower physical machine costs.

(2) On-demand self-service

Cloud computing users can manage their virtual machine operations (e.g., starting and stopping virtual machines) via a web-based graphical user interface (WebGUI), Amazon EC2-compatible application programming interface (API), and OpenStack API. Using these interfaces, users can agilely start their services without having a cloud computing administrator (provider). Furthermore, because cloud computing users actually carry out some operation management tasks that used to be conducted by the cloud provider, the cloud provider's management costs can be lower.

(3) Live migration

Live migration is a function that transfers values stored in a physical memory to another physical memory without interrupting the physical machines containing these memories. This interruption-free replacement of physical machines improves maintainability.

(4) Security

OpenStack has the following basic security functions: authentication of cloud computing providers or users, the hash-based message authentication code (HMAC) specified in the Amazon EC2-compable

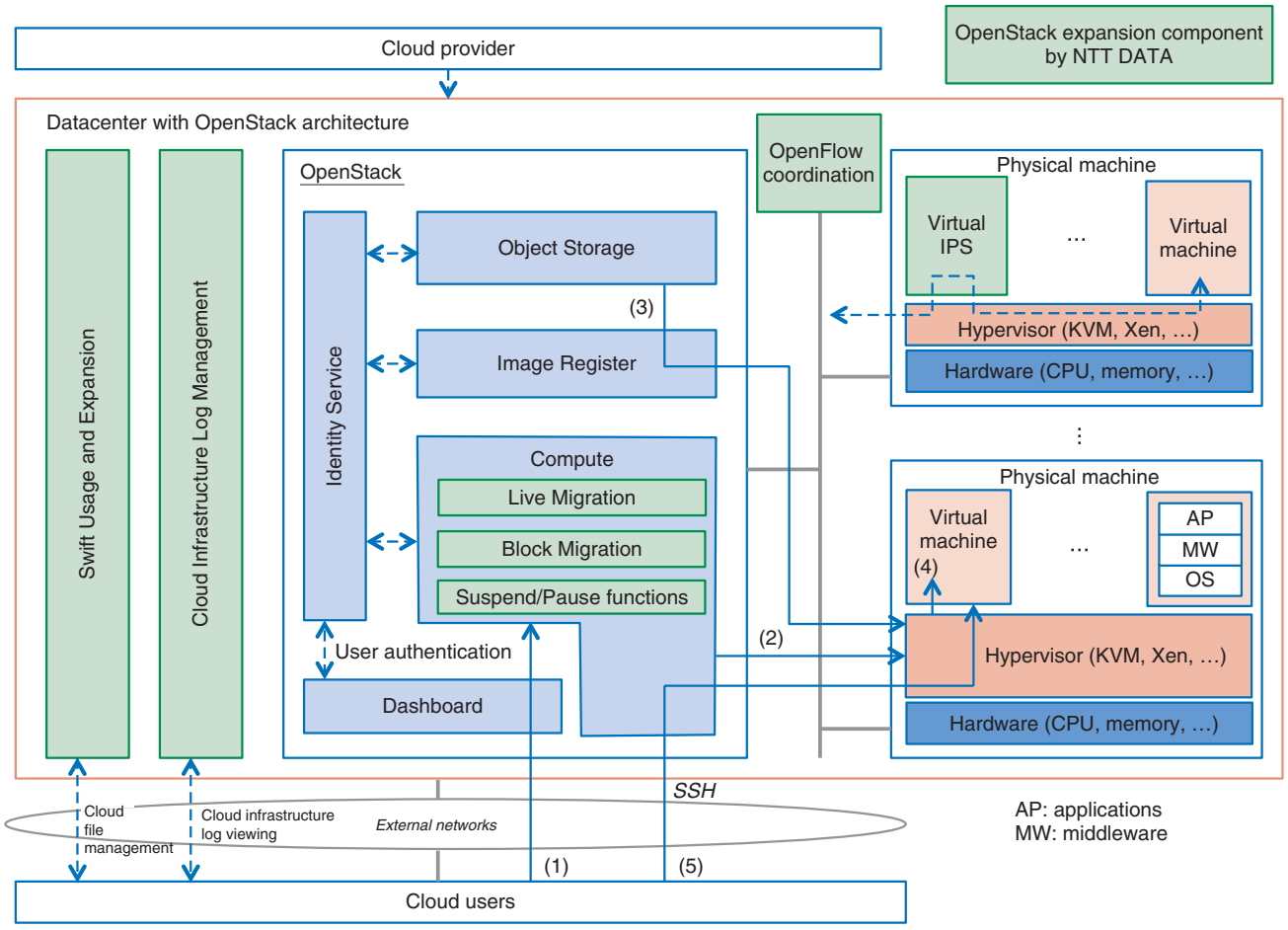


Fig. 1. Configuration of OpenStack.

API, an iptables*-based firewall to isolate virtual machines of different tenants, a virtual private network (VPN), and encryption of communications between a user terminal and a virtual machine by SSH (secure shell).

1.2 Operation

OpenStack comprises multiple components [3], [4] and provides IaaS functions by coordinating them (**Fig. 1**). These components (and their respective code names) are as follows:

The Compute (Nova) component provides computing resource management, computing resource allocation, and message transfer. In the computing resource management, the Compute component manages physical resources (e.g., CPUs and memory) in OpenStack. In the computing resource allocation, the Compute component determines which physical resources should be allocated to the cloud computing user; and in the message transfer, the cloud computing user sends and receives cloud control messages (e.g., to invoke, terminate, or pause the virtual machine).

The Object Storage (Swift) component stores the template information of some available virtual machines (as VM images).

The Image Registry (Glance) component reads the VM image selected by the Compute component from the Object Storage component and transmits it to the physical machine.

The Identity Service (Keystone) component centrally stores identities (IDs) and passwords of the cloud computing users and providers. It also provides user authentication and component authorization.

The Dashboard (Horizon) component provides the WebGUI to the cloud computing users.

* iptables is the name of an application program.

As a brief introduction to the OpenStack process, we describe an example of the procedure for invoking a virtual machine by using the abovementioned components.

(1) Receiving a message: A cloud computing user submits a request to the Compute component to start a virtual machine. At this point, the user selects the computing resources (CPUs, memory amount, disk capacity, etc.), virtual machine types (operating system (OS), etc.), and other options.
(2) Allocating computing resources: The Compute component decides the physical machine appropriate for the requested computing resources, as well as IP addresses to be used. It then notifies the hypervisor for the selected physical machine to start virtual machine operations.
(3) Loading the VM image: The hypervisor requests the Image Registry to transmit the VM image. The Image Registry identifies the VM image to invoke from among several VM images stored in the Object Storage and transmits this identified image to the hypervisor.
(4) Starting the virtual machine: The hypervisor invokes this transmitted VM image.
(5) Using the virtual machine: The cloud computing user accesses the started virtual machine through networks and uses the computing resources of the virtual machine.

2. NTT DATA's projects

NTT DATA's development efforts related to OpenStack involve the following functions (Fig. 1).

(1) We developed the Live Migration, Block Migration, KVM Pause/Suspend functions, as well as other functions, which we contributed to the OpenStack community.
(2) In response to security concerns about cloud computing, which is the top issue hindering cloud computing introduction into the Japanese market, we developed some security functions (the integrated log management system of the cloud computing infrastructure and the virtual intrusion prevention system (virtual IPS)), which complement the default security functions of OpenStack.
(3) We are working to coordinate NTT DATA's OpenFlow Controller with OpenStack.
(4) Using the Object Storage (Swift), we are working to establish technology (Swift Reference Architecture) for building a highly reliable petabyte-class distributed storage system at a low cost.

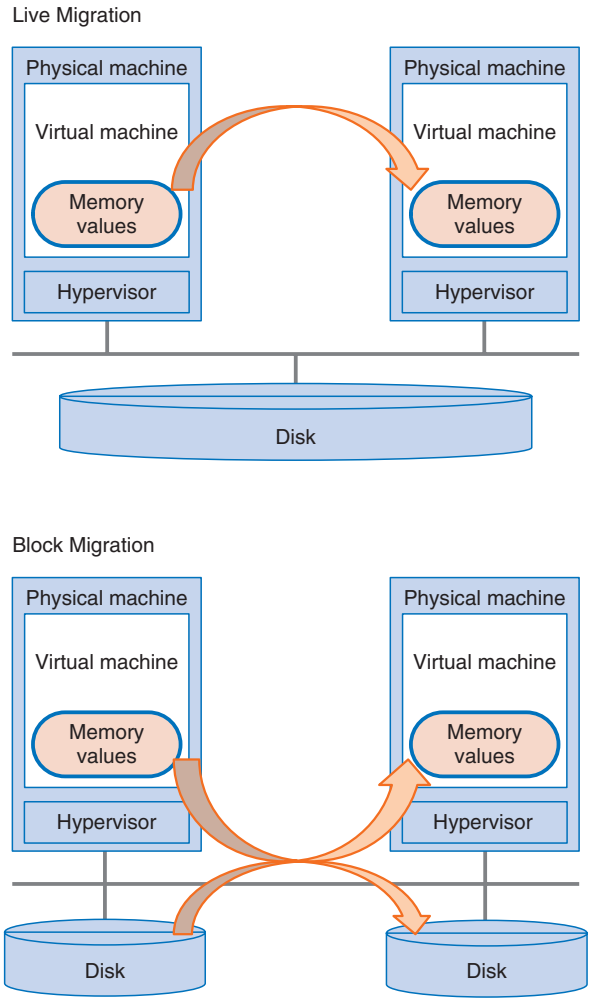


Fig. 2. Mechanisms of Live Migration and Block Migration.

2.1 Block Migration

Block Migration, released in OpenStack Version 4 (Diablo), is a function for moving the virtual machine including its VM image and memory contents, on one physical machine (the original physical machine) to another (the target physical machine) (**Fig. 2**).

Live Migration transfers memory contents but not disk contents, so both the original and target physical machines must access the same disk. Therefore, it is impossible to execute Live Migration if the original and target physical machines cannot access the same disk (e.g., if they are in different datacenters). By contrast, Block Migration transfers both memory and disk contents on one physical machine to another. This makes it possible to move the virtual machine

between different datacenters. By combining Live Migration with Block Migration, we expect to improve the maintainability of virtual machines.

2.2 KVM Pause/Suspend support

This function was released in OpenStack Version 4 (Diablo). The procedure for pausing or suspending a virtual machine is as follows: (1) save the virtual machine, (2) stop the virtual machine, and (3) restart the virtual machine stored in step (1). However, up until OpenStack version 3 (Cactus), it was impossible to pause or suspend virtual machines because the status of a virtual machine was not stored on a disk or in memory. Therefore, NTT DATA developed the Pause function to store the virtual machine status in memory and the Suspend function to store the virtual machine status to disk.

2.3 Infrastructure Integrated Log Management for cloud computing

Because cloud computing users can directly access their own virtual machine, they can also read the virtual machine's log file (e.g., the OS, middleware, and application logs) by themselves. However, viewing this cloud computing infrastructure log file (e.g., virtual environment log and OpenStack log (such as that of the Compute component and Image Registry component)) is not an easy task because a single log file contains all of the logs for the cloud users and each log record requires independent access control.

In response, NTT DATA has developed a cloud computing infrastructure log management system that can extract the logs that are relevant to a specific user in real time from the single log file. This system enables cloud users to understand the status of their cloud computing infrastructure operation through the logs in a safe and speedy manner. This reassures them about the safety of the cloud computing infrastructure, which is not readily visible to users.

2.4 Virtual Intrusion Prevention System

Virtual IPS is the system that monitors communication packets in Layer 3 or higher, detects unauthorized or malicious packets, and prevents their transfer to virtual machines. When an existing IPS product is being installed in a cloud computing system accessed and shared by multiple users, it is necessary to set a different signature file for each user in the existing IPS product, and we must consider user requirements for system performance and security level when making these signature files.

Moreover, in the case of monitoring the communication of virtual machines residing on one physical server, we must monitor the hypervisor on the physical server. However, the existing IPS product cannot monitor the communication of virtual machines inside the hypervisor. Therefore, when this communication is being monitored using the existing IPS product, the virtual machines must communicate through the network outside the physical machine monitored by the existing IPS product.

To solve these inefficiencies, NTT DATA provides a virtual machine equipped with Suricata, an OSS IPS product, as a virtual IPS. For communication via this virtual IPS, we are currently developing technology for monitoring a variety of communication packets, which are exchanged among virtual machines and other devices, for every virtual machine.

2.5 Swift Reference Architecture

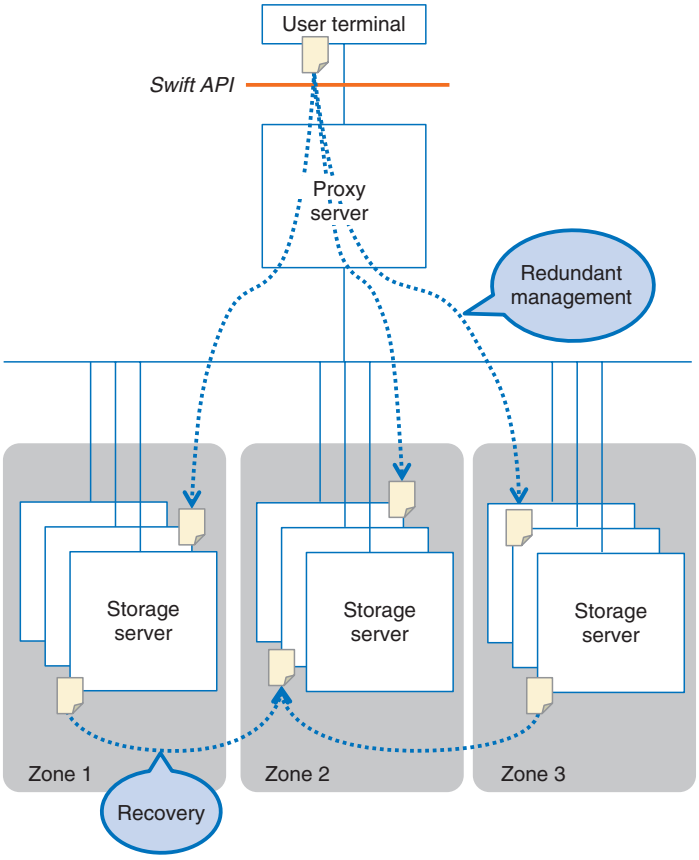
Swift has been developed by US Rackspace for its cloud file storage service. Its main components (**Fig. 3**) are a proxy server and a storage server. The storage server stores objects (which are text files), image files, movie files, and directory metadata files. The proxy server mediates the communication between a cloud computing user and the storage server using Swift API.

There is usually more than one storage server, and these servers are managed as a group called a zone. One object is stored on multiple storage servers, which are assigned to different zones for redundancy. All objects are always monitored by the monitoring processes. If an object is removed because of file or disk trouble, then this lost object is recovered automatically as soon as possible by copying other redundant objects.

To integrate a petabyte-class distributed storage system using Swift, we must consider the system architecture, network design, parameter setting, and system configurations, which must suit user requirements (e.g., the size of files to be stored, network bandwidth to be used, hardware processing capability, failure rates, permissible recovery time, and hardware costs). For this purpose, NTT DATA has standardized design and setting knowhow as the Swift Reference Architecture in order to offer fast and stable large-scale distributed object storage at a low cost.

3. NII projects

Through the dodai project conducted together with the National Institute of Informatics (NII), NTT



Each storage server comprises an account server that manages metadata, a container server, and an object server that manages files.

Fig. 3. Basic configuration of Swift.

DATA has built a mechanism for implementing a cloud API on a leased physical machine for clients who particularly value high machine performance for cloud computing or ones who are unable to access a virtual environment for licensing reasons. Building upon this dodai project, we created a prototype of the Academic Community Cloud System for research purposes for NII in FY2011. We are now working to put this physical machine cloud system into practical use after appropriate operational assessment.

References

[1] OpenStack. http://openstack.org
[2] T. Grance and P. Mell, "The NIST Definition of Cloud Computing," NIST Special Publication, No. 800-145, Sept. 2011.
[3] M. Noguchi, "Basics of Cloud Building, Learning with OpenStack, Currently Popular," Nikkei Linux, No. 151, pp. 120–124, 2012 (in Japanese).
[4] News report (in Japanese). http://techtarget.itmedia.co.jp/tt/news/1101/13/news06.html

Masayuki Hanadate
Manager, NTT DATA Corporation.
He received the B.E. degree in electrical and communication engineering from Tohoku University, Miyagi, in 1997. Since joining NTT Information and Communication Systems Laboratories in 1997, he has been engaged in R&D of information security systems, such as ones for NTT's e-ticket/e-money, smartcard applications, security protocols, and so on. Since moving to NTT DATA in 2010, he has developed some cloud security solutions, such as the virtual IPS/IDS and the distributed object storage and its security option products. He is a member of the Information Processing Society of Japan.

NTT DATA's Efforts for OpenFlow/SDN

Hiroshi Nagasono

Abstract

This article discusses the recent technological trend of OpenFlow/SDN and NTT DATA's efforts related to it for future business. Software Defined Networking (SDN) is currently in the spotlight for its capability to control networks flexibly without being bound by conventional network technologies. NTT DATA has been researching and developing OpenFlow, which is technology for achieving SDN, and has developed its own OpenFlow controller.

1. Introduction

TCP/IP (transmission control protocol, Internet protocol) and Ethernet are currently the current main computer network technologies. To exchange signals, these technologies use various transmission protocols standardized by international organizations such as the Internet Engineering Task Force (IETF) and the Institute of Electrical and Electronic Engineers (IEEE), and such protocols are based on the OSI (Open Systems Interconnection) Reference Model*1 that describes the communication layers and their functions. Network devices produced by different manufacturers can communicate with each other by complying with these standards.

Most device manufacturers develop network device software along with the hardware to enable the hardware to perform operations appropriate for each communication layer. The device manufacturers also implement their own networking functions in addition to the standard protocols to add value to their products and thus make them more competitive. However, this makes it difficult to replace devices that have exclusive functions with ones from other manufacturers, so network owners may need to keep on using the same manufacturer's devices. This situation is called *vendor lock-in*.

Another issue with the current network is the difficulty of functional enhancement. The requirements for information systems are becoming more complicated and advanced, and you may want to upgrade your network to respond to such requirements. However, the functional limitation of network devices sometimes inhibits network upgrading because the devices do not allow network configurations unsupported by the device manufacturers. For network device users who wish to configure the network as they wish, the only option is to wait for the device manufacturers to support their desired functions.

One solution that can address these problems with conventional networks is Software Defined Networking (SDN) based on OpenFlow [1].

2. OpenFlow/SDN

2.1 Definition of SDN

Let us clearly define SDN before proceeding to discuss OpenFlow. As the name implies, SDN is a system architecture in which networking is defined by software. Although the name SDN has become widely known only recently, the software defining architecture itself has been in use for some time. A conventional network device comprises hardware and software. However, users could not previously define a network by themselves because the device manufacturers did not publish the application programming interface (API) for controlling the hardware.

OpenFlow was introduced in response to such situations. It is one of the technologies capable of

*1 OSI Reference Model: A model with a set of layers that define the different communication functions to be used on the networks and devices connected to the networks. Suggested by the International Organization for Standardization.

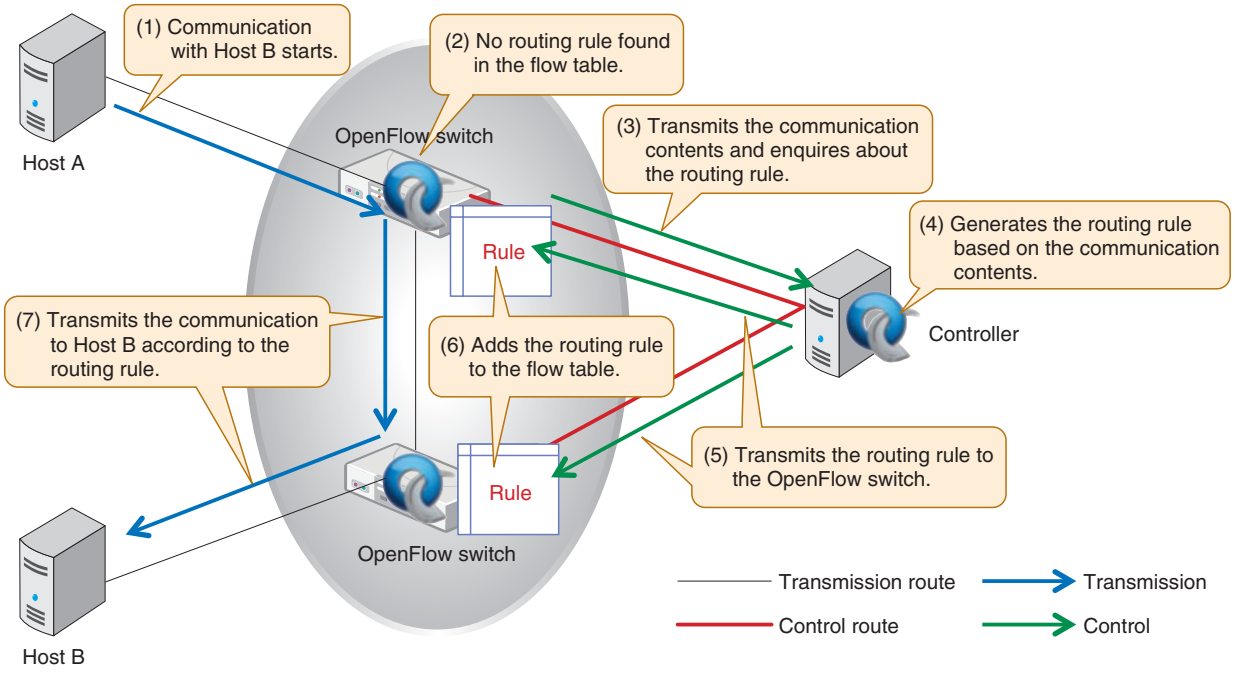


Fig. 1. Basic behavior of OpenFlow.

achieving SDN. It is not a method of networking that provides specific functions, such as L2 (Layer 2) switching or IP routing. In fact, OpenFlow is simply an interface to be installed on a network device. OpenFlow specifications are defined by the Open Networking Foundation (ONF) [2], which promotes the spread and standardization of SDN. The specifications cover the components and the basic functions of a switch as well as the OpenFlow protocol for managing an OpenFlow switch from a remote controller. OpenFlow opens up access to the API controlling the hardware—even though the API is not disclosed by the device manufacturers—and enables users to define networking by themselves.

2.2 Control of OpenFlow switches

Although the OpenFlow switch specifications are standardized, the controller's behavior is not. The communication unit handled by OpenFlow is called a *flow*. The behavior of a conventional network device is as follows (here an L2 switch is used as an example): the L2 switch identifies the transmitting destination's physical port using a destination media access control (MAC) address[*2] as an identifier and submits frames from the physical port (In the case of an L3 switch, the destination IP address is used as an identifier.). Like conventional network devices, OpenFlow uses the same header information including a destination IP address and an originating MAC address as components; however, the difference is that OpenFlow can handle the combination of such components as an identifier. For example, in OpenFlow it is possible to configure an identifier as a flow submitted from a specific originating MAC address to a specific destination IP address, or as TCP port 443 communication received from a specific physical port. When a flow is identified, the OpenFlow switch performs control operations such as packet transmission and packet correction on the flow. The information for controlling this flow is called the *flow entry*, and the information is stored in a *flow table* within the OpenFlow switch. Controllers usually create a flow entry in response to a request from an OpenFlow switch and update the flow table in the switch (**Fig. 1**).

A series of such control operations enables an OpenFlow switch to emulate the functions of a conventional network device such as an L2 switch and router. Furthermore, it also makes possible network configurations that used to be unachievable by conventional network devices such as L2 segments

*2 MAC address: A physical address assigned to network device hardware such as a local area network (LAN) card to identify each node on the networks.

configured using only MAC addresses without using a virtual local area network (VLAN).

2.3 Implementation of OpenFlow/SDN

Practical usage of OpenFlow is now available to provide flexible network control. There are two configurations for implementing OpenFlow: hop-by-hop and overlay.

2.3.1 Hop-by-hop configuration

In the hop-by-hop configuration, the controller centrally manages the end-to-end communication by setting all the network devices via the OpenFlow switch. This configuration makes best use of the following OpenFlow features.

(1) Flexible routing

Flexible routing appropriate for the communication characteristics is possible such as routing a high-priority communication to a physical path that has sufficient bandwidth while routing lower priority ones to other paths.

(2) Traffic monitoring

Since an OpenFlow switch is capable of managing the flow status by using a flow table, it is possible to recalculate routes when traffic congestion is anticipated.

2.3.2 Overlay configuration

In the overlay configuration, servers and network devices at the edge of datacenters are managed by OpenFlow, and communication to the edge is managed by the conventional networks. The overlay configuration also uses tunneling for communication between the originating and destination edges.

(1) Existing facilities can be used.

Not every network device on the communication path needs to be configured by the OpenFlow switch. OpenFlow can be used partially within the conventional network. If Open vSwitch, which is an OpenFlow-compliant software switch, is embedded in a virtual server, such servers can function as edge devices; this allows network management by OpenFlow.

(2) The amount of routing information can be limited.

Network controllers store edge information and logical network routing information, but not physical routing information between edges. This reduces the number of items of routing information managed by the controller.

3. Virtual network controller

NTT DATA is in a position to provide information system solutions as a global information technology (IT) innovator. We have recently received requests from our corporate customers for network configurations that cannot be achieved with conventional network devices. We intend to utilize OpenFlow/SDN to achieve the optimum network solution that can accommodate such requests. To achieve this, we have developed our own virtual network controller and started providing it. It has the following features:

(1) Easy to customize

The controller is composed of two types of software: NetworkOS (NOS) and NOS-Application (NOS-AP) (**Fig. 2**).

NOS-AP is software for defining networking; it can perform routing based on its own algorithm and provide logical configuration management. NOS-AP can also be coordinated with higher systems that handle other network management tasks such as OpenStack. On the other hand, NOS provides functions commonly required in controllers based on the OpenFlow specifications, such as functions for generating OpenFlow protocols and handling events. Along with the software, we provide NOS-API for efficiently controlling NOS from NOS-AP. NOS-API helps reduce the controller development costs and shorten the development period.

(2) Enables a small start

As mentioned earlier, we offer the hop-by-hop and overlay configurations for implementing OpenFlow. One barrier to the upgrading of network technologies is that existing network devices with limited capabilities are still usable, so delaying upgrading is seen as a viable option. The overlay configuration of OpenFlow/SDN can solve such problems by enabling network technology to be upgraded while leaving existing network devices untouched. These network devices can be changed into OpenFlow-compliant devices one by one as required, so it is possible to configure the network with advanced technologies in stages (**Fig. 3**).

(3) High reliability

Because the OpenFlow/SDN architecture allows central control of networks, the controller can become a single point of failure. Large, complicated networks such as the public cloud or the networks of service providers require a mechanism for flexibly upgrading performance and functions. A virtual network controller secures network availability through the high-availability cluster configuration. Furthermore, if controller functions are deployed in a distributed manner, the controllers can also offer scalability to the network.

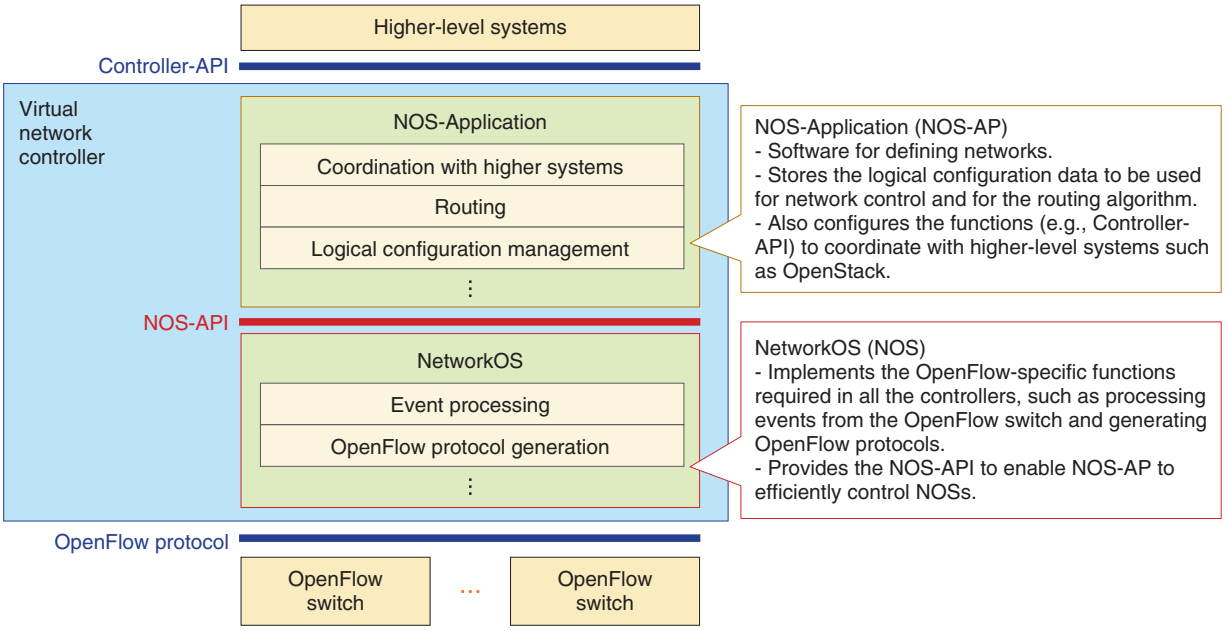


Fig. 2. Software configuration of the virtual network controller.

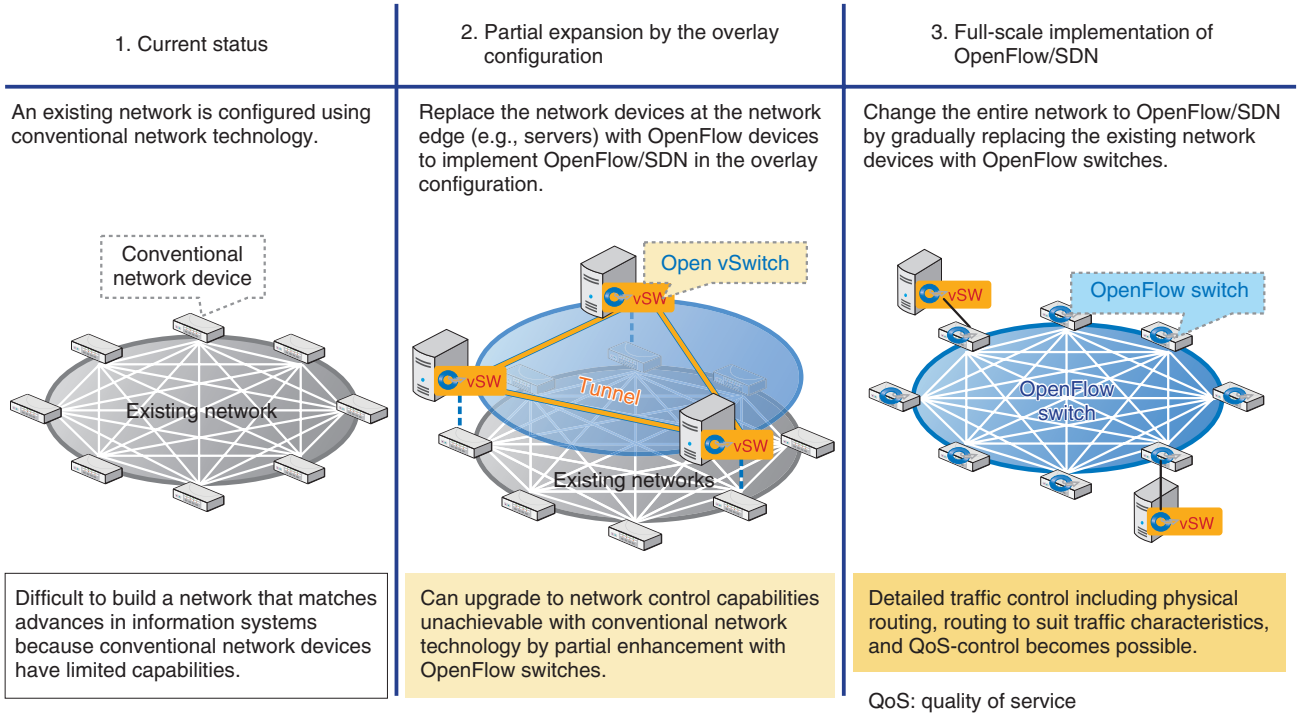


Fig. 3. Stepwise implementation of OpenFlow/SDN.

4. Concluding remarks

OpenFlow/SDN used to be applied mainly to data-center networks as technology for virtualizing networks in order to catch up with more advanced server technology. However, there are a number of other areas where OpenFlow/SDN technology can be applied to solve problems that cannot be solved using conventional network technology. NTT DATA will enhance the functions and capabilities of its virtual network controller to solve such problems.

References

[1] OpenFlow: Realization of the New Network which Reverses an Old Concept (in Japanese). http://thinkit.co.jp/book/2012/02/01/3150
[2] Open Networking Foundation. https://www.opennetworking.org/

Hiroshi Nagasono
Senior Expert, Systems Engineering Business Unit, System Platform Sector, NTT DATA Corporation.

He received the M.E. degree in mathematics from Nagoya University, Aichi, in 1998. He joined NTT in 1998. From 2002 to 2007, he was in the R&D Center of NTT EAST. He developed network service systems, such as IP-telephony. Since moving to NTT DATA in 2007, he has developed some network solutions, such as the CTI system, IP-FAX, and OpenFlow/SDN controllers.

Strategy and Efforts for Robotics Integration Aiming at Combining Information and Communications Technology with Robots

Toyoaki Kagaya

Abstract

This article introduces the strategy and research & development for fusing information and communications technology (ICT) with robots in relation to the open source trend. As machine-to-machine (M2M) communication and sensor networks become widespread, we are entering a time when the robots coordinated with clouds can offer new added value.

1. Introduction

When people hear the word *robot*, they tend to think of industrial robots and humanoid robots. The Committee on Robot Policy Study under the Ministry of Economy, Trade and Industry defines a robot as a mechanism composed of three technical components: sensors, an intelligent control system, and a power train. The committee regards a robot as a mechanical system with intelligence[*1]. However, because the areas where robotic technology is applied have been expanding, the above definition is also being extended.

As a systems integrator, not a manufacturer, NTT DATA also takes a broad view of robots, treating them as systems that utilize robotic mechanisms or robotic technology. We aim for robots that work safely and efficiently within a smart space by coordinating resources and services in the cloud or robots that deliver services or perform actions within such an intelligent space, rather than robots that work in a standalone manner.

*1 The definition of a robot by the Patent Agency is (1) a machine with manipulating functions, (2) a machine with mobility, ability to acquire external information, and functions for determining its own behavior, or (3) a machine with communication functions, ability to acquire external information, and functions for determining its own behavior and behaving accordingly.

2. Motivation for robot technology research

Faced with the persistent recession in the domestic market, the information and communications technology (ICT) industry expects smart grids and smart cities to develop as a new growth business area (**Fig. 1**). Because new information infrastructures and solutions, such as machine-to-machine (M2M) communication and sensor networks, are expanding, we can make robots into contributors to society and business by fusing robotics and ICT by organizing an environment in which robots and the new information infrastructures and solutions can work together.

For example, it will soon become possible to have an intelligent mobile robot that understands its surroundings from sensor networks and moves safely and efficiently or a robot that can feed the highly valuable information acquired from big data to the real world. We expect that robot technology that senses the world more efficiently and automatically will become more important for efforts to establish big data in the cloud.

Robots for manufacturing made in Japan account for nearly 70% of the world's market; however, competition with Europe and the USA is already harsh,

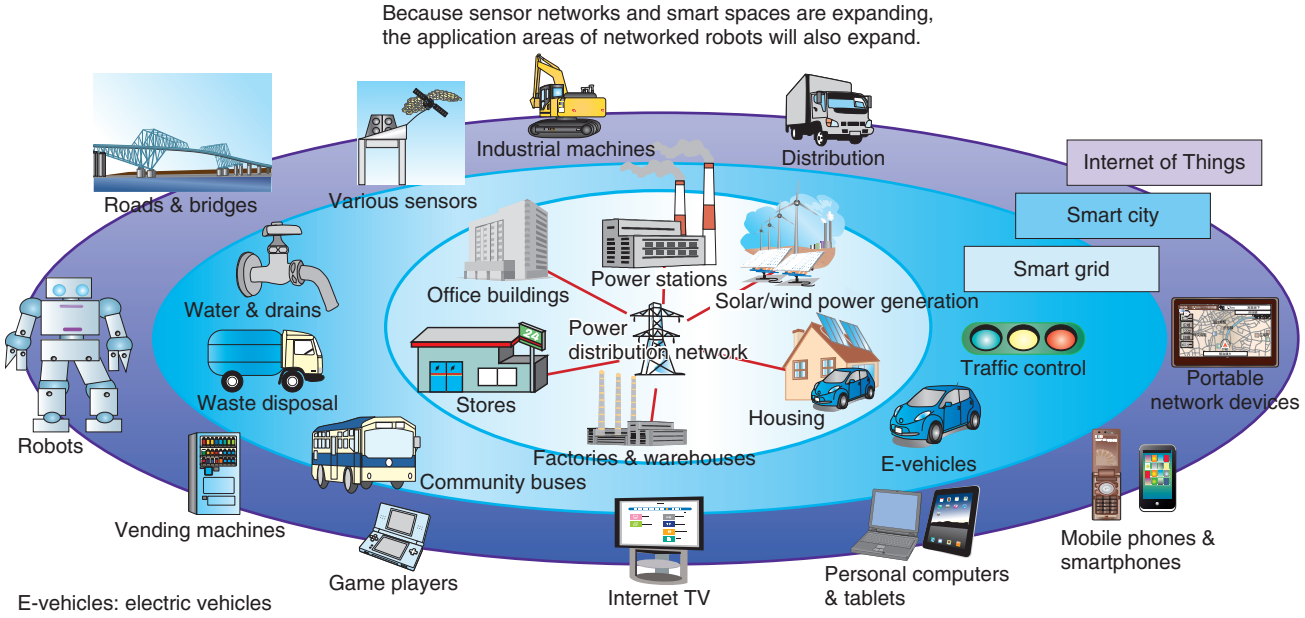


Fig. 1. The expanse of smart space and the application areas of robots.

and China and Korea are catching up with the assistance of strong support from their governments. Japan is now in the stage of growing beyond robots for production and is developing the robot application business through the creation and spread of robots for the service industry.

NTT DATA will expand the ICT business with wider application by working with a number of robot manufacturers to open up the M2M market utilizing Japan's world-class robot technology.

3. From M2M to M2M2A

M2M is a system in which multiple machines connected to a network exchange data without human mediation to automatically provide optimum control. NTT DATA regards a robot as a critical machine that contributes in this system. We predict that the machine-to-machine-to-actuation (M2M2A) solutions, where sensors and robotic technologies are combined to fuse the real and cyber (or virtual) worlds, will become common (**Fig. 2**). In such solutions, services based on visualized information generated from information gathered by sensors are linked and the formulated actions are performed by machines.

4. Likely areas where ICT and robotics can fuse

NTT DATA set up the M2M Cloud Promotion Office in October 2011 in order to accelerate M2M-related business across the company. The office aims to advance ICT services by building an M2M cloud and offering new cloud-based services. Our target areas for applying robotics are life support and social infrastructure maintenance, taking account of the direction of our M2M business expansion. These areas are also our existing business forte.

4.1 Robotics applications in the life support area

Nursing care is one area with high potential for robotics application. There is currently strong demand for robots able to provide support for the independence of care receivers or able to assist nursing staff. In fact, the government announced in 2012 that care robots will be covered by long-term care insurance from 2015. This will further accelerate the research and development (R&D) and field trials of such robots.

Through the combination of M2M technology with devices used by the elderly or patients at care facilities, data concerning the physical conditions and daily routines of people under nursing care can be collected. This data can then be used to reduce the

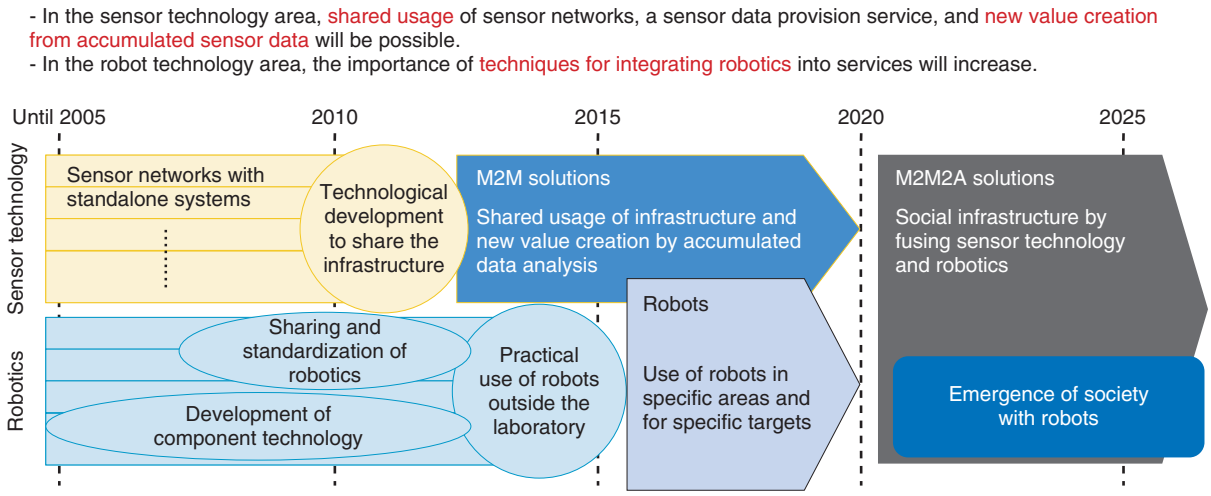


Fig. 2. Forecast of trend for sensor and robot technologies.

workloads of caregivers and to provide comprehensive health management in combination with personal health records and lifelogs.

Utilizing robotics will make it easier to sense their physical conditions. With such data, it will be possible to create practical appliances such as rehabilitation equipment and programs that are tailored to individuals' conditions and drug management machines that prevent overdoses or failures to take medicine.

There is already some R&D of robots for supporting communication between hospital staff members and patients, including their family members, for patients living in remote areas or for hospitals with staff shortages. These robots include an in-hospital guide robot that takes a patient to the appropriate place within a hospital or care home safely and efficiently and an autonomously mobile robot that provides videoconferencing functions. These robots can be improved to offer further convenience to service users and providers by enabling them to provide medical consultation and diagnosis and to recommend healthy habits by utilizing personal health records in the cloud.

4.2 Robotics applications in the social infrastructure area

The maintenance of social infrastructure, including large buildings and lifeline services, has recently been becoming more important. Infrastructure breakdowns cause massive social and economical loss, so breakdown prevention is vital. A large part of Japan's current infrastructure was built during its high growth period and many structures are approaching renewal time. This renewal must be executed to provide optimum results at a low cost because the government's tax income is declining owing to the aging of society. Construction and maintenance work is ideal for robots because it tends to be tough, dirty, and dangerous and hence unpopular with human workers. Indeed, robots are gradually starting to be used in such workplaces.

As one of the M2M cloud solutions for social infrastructure maintenance, NTT DATA offers a bridge monitoring solution called BRIMOS. It monitors a bridge constantly in real time using sensors installed on the bridge. Although it does not use robots, it would be possible to create a solution with higher added value by combining BRIMOS, which already offers sophisticated service, with robotics. Some structures have complicated sensor installation designs or require monitoring over a long distance or wide area. Robotics can help in such cases by offering advanced and autonomous mobility and location identification functions by using a geographical map service in the cloud. Robots can perform structural checks efficiently. As for the visual inspection tasks carried out by human beings, robots can collect images of a structure while autonomously moving in and around it. They can then amalgamate the mass of collected images and thus improve the quality of images in the cloud to help identify any faults.

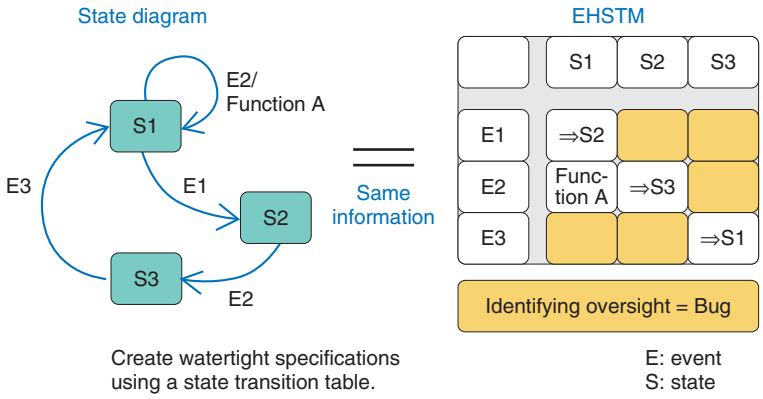


Fig. 3. Extended hierarchy state transition matrix (EHSTM).

5. Organizing the development environment

5.1 Robot technology middleware

Making architectures open is rapidly becoming common in the robot industry, and establishing a development environment that utilizes such open architectures and resources is important for creating the contact point between robotics and ICT.

As part of the open architecture strategy for promoting the common use of service robots, the National Institute of Advanced Industrial Science and Technology (AIST), which leads R&D in Japan's service robot industry, suggested the standard interface specifications of robot technology components (RTCs) to the Object Management Group (OMG)[*2] and they were issued as the OMG's standard in 2008. AIST also developed and published RT middleware[*3] to control RTCs compliant with the standard specifications.

In collaboration with one of our group companies, CATS Co., Ltd., NTT DATA altered the CATS's computer-aided software engineering (CASE) tool, called ZIPC, for embedded software development to comply with RTCs and RT middleware, which contributed to improvements in the design quality (**Fig. 3**).

5.2 Safety functions

Because robots are automatically driven, they must be equipped with advanced functional safety[*4] measures to avoid any harm to human beings and the surrounding environment.

Robot manufacturers are often the manufacturers of existing home appliances and other equipment. For this reason, they fear damage to their brand image by accidents caused by their robots, so they do not launch their robot products on the market unless they are absolutely confident in the safety of the robots' behavior. As a result, robots remain in the field-trial stage and there are few cases of practical use. This is hampering the business from budding into commercialization, causing a vicious circle in the industry. At the same time, securing high reliability in robots is a strategically important long-term competence as a part of corporate branding.

The robot industry is also waiting for the release of ISO13482 in 2013. ISO13482 is the international safety standard for personal care robots (robots designed to communicate with human beings) created on the basis of the functional safety standard for safety-related systems IEC61508[*5].

If we are to use robots as part of our business, then close collaboration with manufacturers is important. We must also keep an eye on the industry's approach to safety issues that may affect procurement requirements or lawsuits. This is why NTT DATA is discussing with AIST about a comprehensive development

*2 OMG: An industry organization that sets standards for distributed object-oriented systems. It promoted the standardization of UML and CORBA. Since 2005, OMG's Robotics Domain Task Force has been working on robot-related standards.

*3 RT middleware: Middleware for making sensors, actuators, and software control robots into components.

*4 Functional safety: An acceptable level of safety secured by implementing some functional contrivance. The antonym of intrinsic safety.

*5 IEC61508: An international standard for the Functional Safety of Electrical/Electronic/Programmable Electronic Safety-related Systems specified by the International Electrotechnical Commission (IEC). This is used as the basis for functional safety standards for many industries.

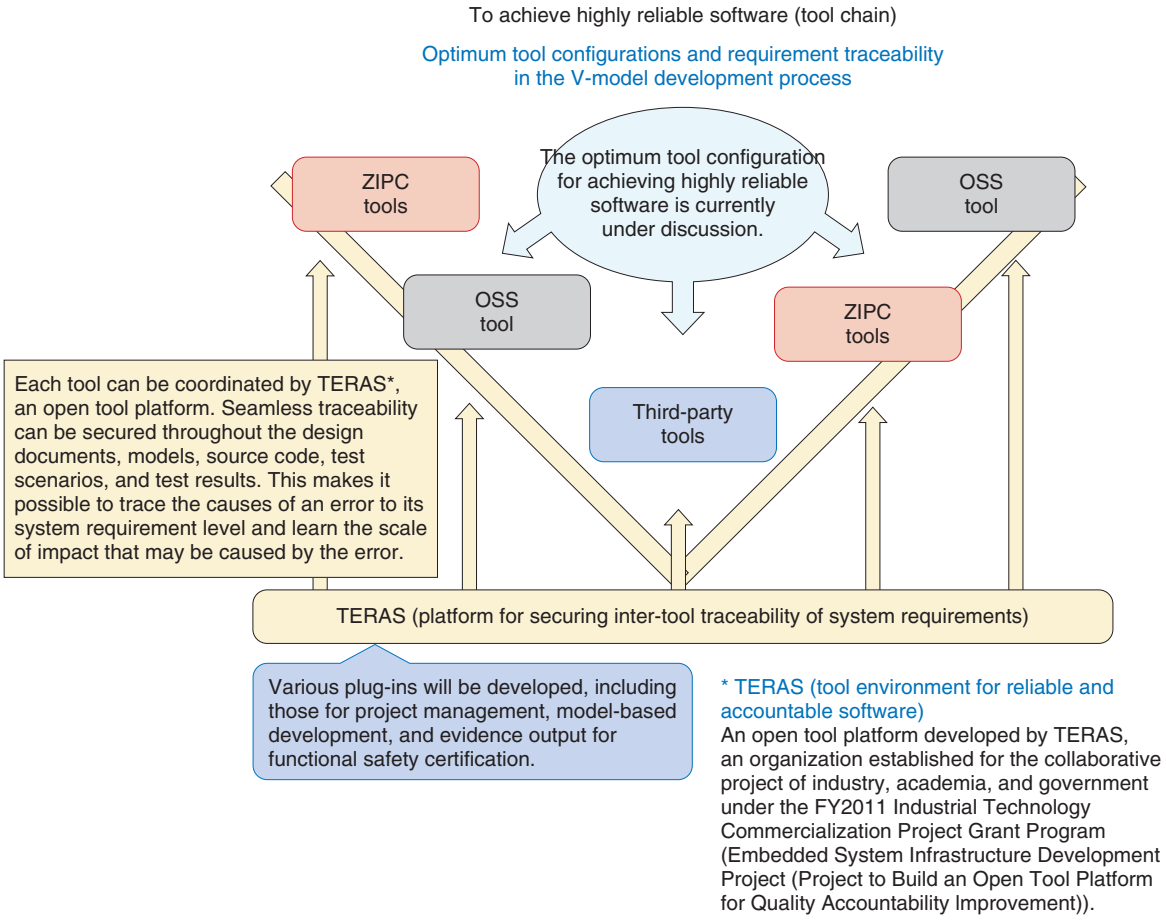


Fig. 4. Promotion of creation of reliable software (tool chain).

environment for delivering highly reliable software for robots to accelerate IEC61508 certification.

We are researching configurations and methods that enable seamless tracing of requirements across the series of ZIPC products and other major open source software (OSS) tools within the V-model development process of embedded software. This lets developers provide accountability for risk assessment across their development process from requirement decisions, through design and implementation, to verification and lets them produce highly reliable software more easily (**Fig. 4**)

5.3 Coordinating with the cloud in the development environment

RT middleware offers the benefit of cloud coordination through its connectivity with RTCs distributed across networks. If a cloud service application programming interface (API) with the RTC-standard interface can be installed in the cloud, robots equipped with RT middleware can easily use cloud services through this API.

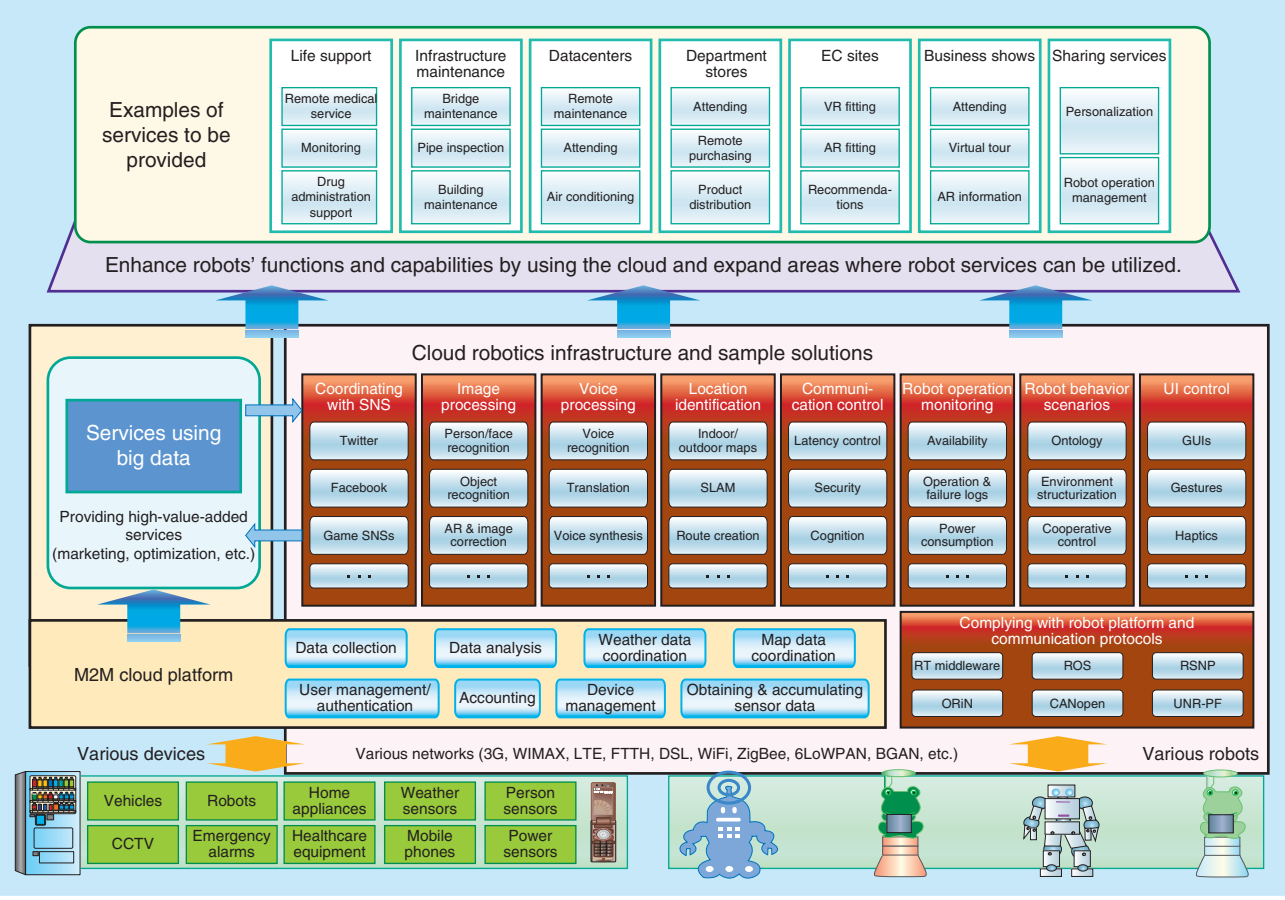
In the future, we aim to make an environment where resources in the cloud, application software, and robot software can be easily coordinated by making RTCs in the cloud easily accessible from the development environment or by building a development environment in which the RTC interface can be easily implemented.

6. Future prospects

6.1 Remote management of robots

Robot servicing technology and operating technology will be essential to encourage the wide implementation of robots and respond to situations in which robots are commonly used.

For example, in the nursing care industry, for which



AR: augmented reality
CCTV: closed-circuit television
EC: electronic commerce
GUI: graphical user interface
SLAM: simultaneous localization and mapping
SNS: social networking services
UI: user interface
VR: virtual reality

Fig. 5. Conceptual diagram of cloud robotics.

robotics application is promising, care support equipment and tools are often rented and their users are mainly elderly people. Therefore, close monitoring and maintenance are necessary. As the revised Long-Term Care Insurance Act 2012 obliges servicing planning of welfare tools, a maintenance service during a rental period and after sales for nursing care equipment are even more important.

NTT DATA aims to create high-value maintenance and operations services that can be included in nursing care equipment service plans. Such services include the provision of a development environment that makes it easy to implement the modules and interfaces to obtain a range of data such as operation logs, error and failure information, and availability or make it easy to obtain such information in the M2M cloud.

6.2 Cloud robotics

The idea that the cloud could make robots lighter, cheaper, and smarter was articulated by James Kuffner, one of the developers of Google's self-driving car, at Google I/O 2011. NTT DATA is continuing its R&D of the cloud-robotics infrastructure and solutions to provide a service robot system that is highly beneficial to users, as well as M2M cloud coordination. We intend to achieve such infrastructures and solutions by equipping robots with high-speed computation abilities, large storage, advanced functions, and knowledge, which cannot be achieved with conventional standalone robots (**Fig. 5**).

Toyoaki Kagaya
Senior Expert, Promotion Office for Robotics Integration, Service Innovation Center of Research and Development Headquarters, NTT DATA Corporation.

He received the bachelor's degree in economics from Chuo University, Tokyo, in 1996. He worked for embeded system software vendors and engaged in R&D and the software development process of embeded systems such as consumer electronics, medical electronics devices, and network module drivers. He moved to NTT DATA in 2009. He is currently engaged in R&D of systems integration of robots and ICT systems.

Technology Development for Communication Advancement

Toru Takaki

Abstract

This article introduces NTT DATA's research and development activities for advancing communications among people and between people and machines and mentions future prospects. High-quality communication is now receiving attention as a key element for improving corporate productivity and creating a better society.

1. Introduction

Corporations and public bodies hope to reinforce their competence in the domestic and international markets by improving productivity. One way to achieve this is to strengthen organizational bonds and improve productivity through advances in communication.

Efficient realtime communication among people is important. Furthermore, smoother person-to-machine communication and person-to-person communication achieved through machines is essential in the current communication environment, where information and communications technology (ICT) systems are becoming a prerequisite. At the same time, the utilization of unstructured data, such as documents, emails, and other textual information managed within systems, is increasingly expected to contribute to improvements in productivity. The means to bridge the communication gap in a diverse world is also awaited. This includes a way to overcome language barriers and support handicapped people to enable them to become more independent.

Human-machine interfaces keep changing to suit the devices that are common at a particular time, and the interfaces have a great impact on the usability of communications. Communications technology is now a component of the ICT infrastructure.

2. History of communications technologies

To enable a clearer understanding of the technological trend of communications technologies, a brief history of them is shown in **Fig. 1**.

In the 1990s, the dominant means of communication shifted from the telephone to email as the Internet became commonplace. It was the decade when person-to-person communication was linked via ICT systems. During the first decade of this century, the form of communication developed into communication available at anytime and anywhere as a result of the explosive spread of mobile communication devices such as cell phones and wireless broadband. Furthermore, as smartphones became popular, people became accustomed to touch-based user interfaces. In the 2010s, we assume that optimum communication, in which the means of communication is optimized according to the users' situations, will become the norm. The critical technologies for optimum communication are media analysis technology that utilizes linguistic and non-linguistic data and device technology that improves the human-machine interface. These technologies are now being vigorously studied from the research level to the practical level (**Fig. 2**).

Recent examples of advanced communication applications that utilize the rapidly spreading smartphones are NTT DOCOMO's Shabette Concier (talking concierge) and Apple's Siri. More new services that use advanced media processing and devices are expected to follow.

3. Aims of communication advancement

We believe that communication advancement can resolve the issues in the following four categories:

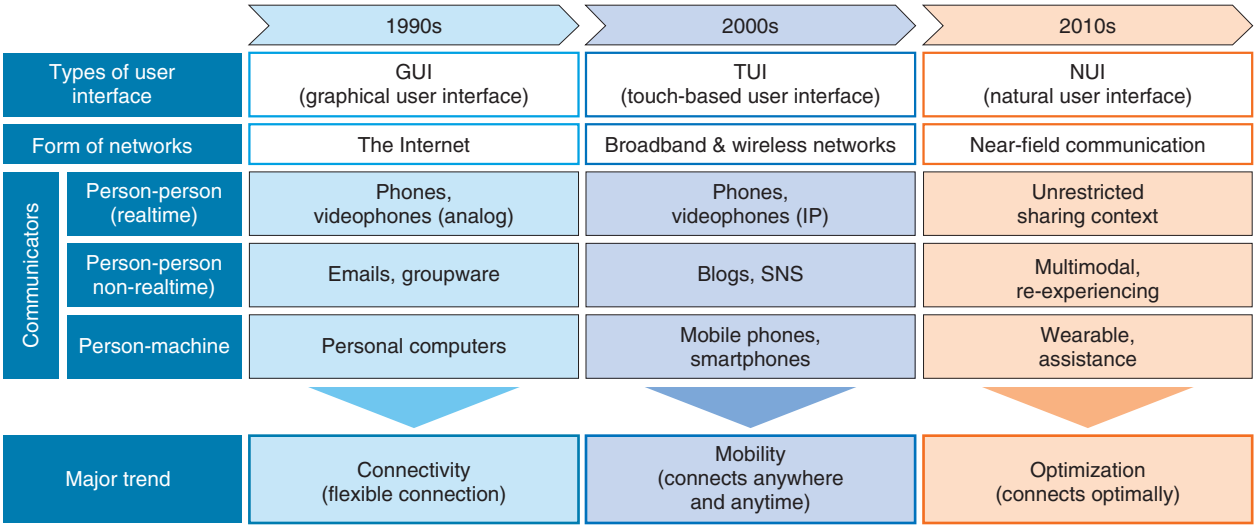


Fig. 1. History communications technologies.

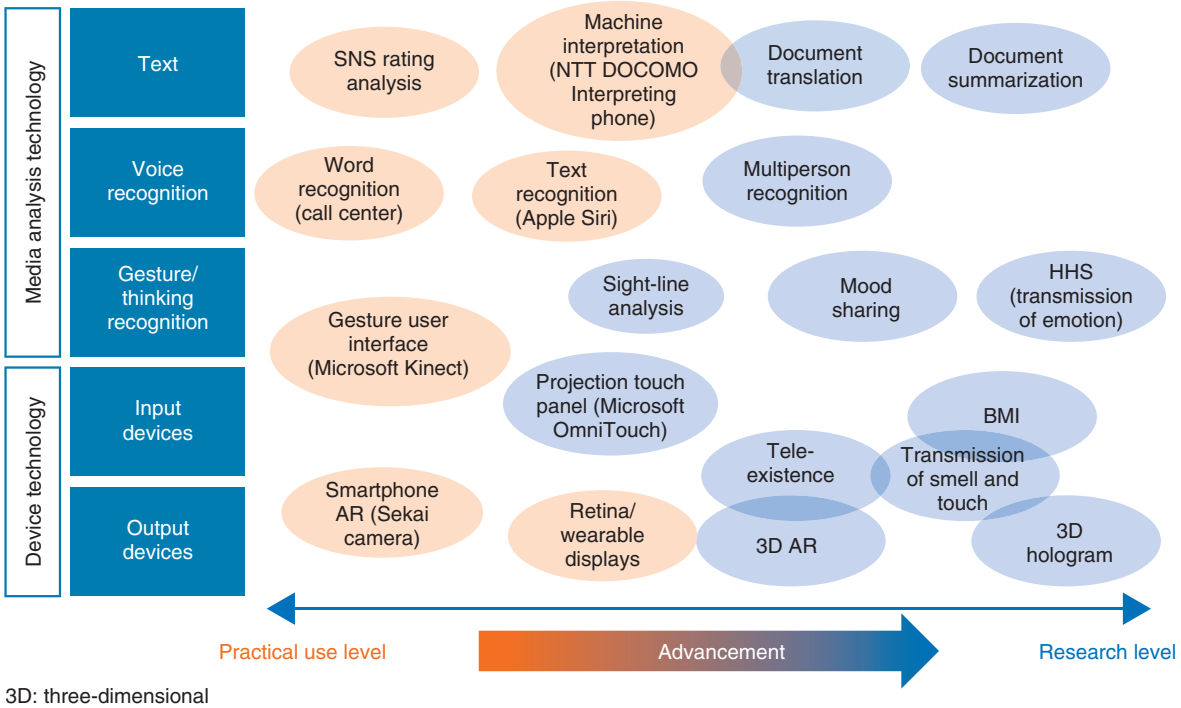


Fig. 2. Assessment stage of technological components.

Table 1. Issues that can be resolved by communication advancement.

	Issues	Examples of solutions
(1) Communication gap elimination	Offering communication without any feeling of stress	- Interpreting phone - Telepresence - AR and wearable devices
(2) Quality improvement	Adding value to communication	- Automatic information suggestion - Automatic minutes creation - Auto facilitation
(3) Organizing ways of communication	Supporting the provision of sufficient communication	- Electronic secretary - SNS utilization
(4) Improving experience value	Offering entertainment, security, and satisfaction	- Virtual world - Entertainment robot

eliminating the communication gap, improving communication quality, optimizing multiplexed communication, and improving the experience value (**Table 1**).

As the first step to tackle these issues, NTT DATA is currently working to develop application software to eliminate the communication gap. We aim to raise the current, insufficient, level of ICT-based communication to the level at which people can communicate without stress. Expectations for linguistic support services to overcome the language barrier are rising to cope with the rapid global expansion of businesses and social activities. We aim to develop application software that will help users understand foreign languages and create documents in multiple languages. Our ongoing projects include developing machine translation software optimized for different businesses and users and document creation software.

4. Examples of NTT DATA projects

4.1 Japanese document creation support tool for software development

With computer systems being increasingly developed offshore, such as in China, the proportion of development-related documents in Japanese created by non-Japanese speakers is also rising. However, this causes various problems such as a longer development period and larger number of grammatical and expressional errors in documents owing to the inadequate Japanese language skills of non-Japanese document writers. Countermeasures to address this linguistic gap are required.

NTT DATA is currently developing a foreign language document creation tool that can take account of error tendencies caused by differences in the linguistic characteristics of the language used in the document and the mother language of the document writers. This tool is designed particularly for Chinese people who create documents in Japanese. It has an example text search function and a Japanese checker function.

(1) Example text search function

The example text search function identifies grammatically correct text in a collection of previously created documents as a reference for the document writers to help them create correct sentences. One of the features of this function is that it displays examples of text containing the most frequent items first, such as Japanese particles and verb conjugation. It also displays simple sentences that are preferred in design documents and business letters near the top of the search results. This ensures that the most suitable examples are shown to the users. An output example is shown in **Fig. 3**. This example shows the search results for particles that can connect the words “message” and “display”. Underlining the particles makes it easier for the writers to compare candidate examples.

(2) Japanese checker function

The Japanese checker function checks whether the particles used in a specific design document match normal usage in design documents. Foreign words written in katakana, one of the Japanese phonograms, often confuse Chinese writers, leading to errors in the Japanese documents. This is because native Chinese speakers and Japanese speakers analyze the sounds of the original foreign words differently before converting the sound into katakana. For example, the English word “message” is commonly written as “メッセージ” (messeji) in Japanese, but sometimes written wrongly as “メセッジ” (mesejji) in documents created offshore. The Japanese checker function can identify phonemic errors and display the correct expression by taking account of the phonemic

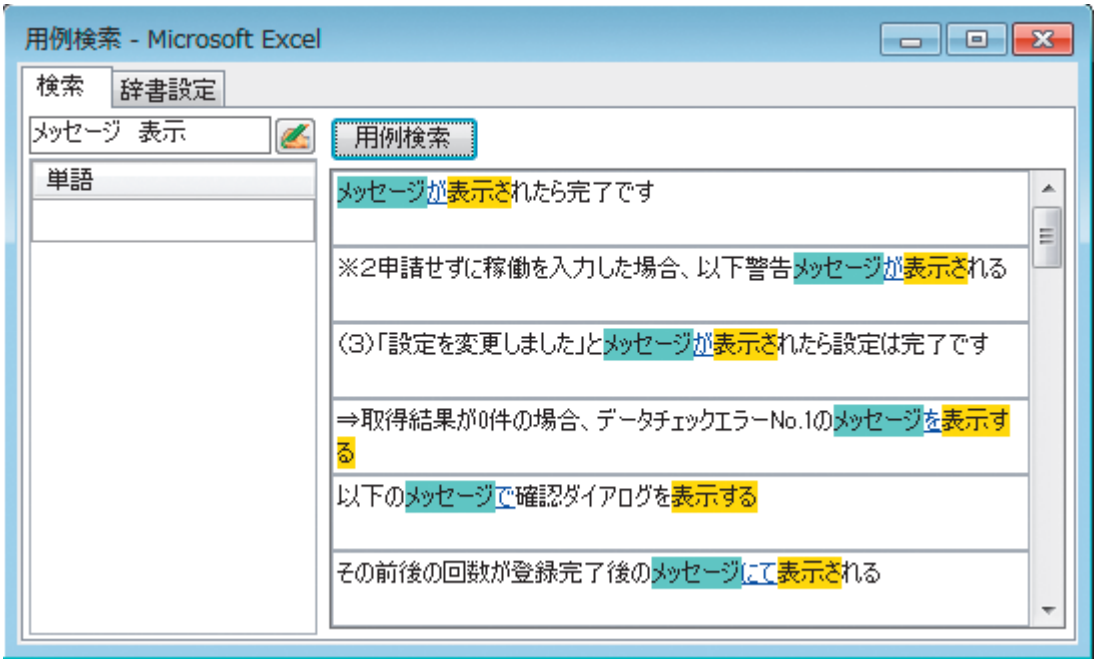


Fig. 3.　Output of the example text search function.

	Source content	Modified example	Indicated content
1	シットを参照する.	シートを参照する.	Error in katakana
2	「A」と「B」の２つフォルダの配下に…	「A」と「B」の２つのフォルダの配下に…	Particle omitted
3	ログ出力の処理開始する.	ログ出力の処理を開始する.	Particle omitted
4	同時に同じテーブルをアクセスすれば，データベース競合の…	同時に同じテーブルにアクセスすれば，データベース競合の…	Inappropriate particle

Fig. 4.　Output of the Japanese checker function.

similarity between the erroneous word and the candidate correct word. An example of the function's output is shown in **Fig. 4**. The checker displays sections with erroneous words, a candidate for the correct word, and the type of error in a table format. This example shows one katakana usage error and three particle errors. This output is achieved by applying research results from the NTT Media Intelligence Laboratories.

These functions are at the stage of field trial assessment in a document development office in China. The Chinese writers use them to create design documents or make enquiries in Japanese. The functions are also used to assess the Japanese quality for self-review or group-review, and they offer communication advancement in the global development environment.

5. Further applications

Communication advancement technology is regarded as the key component for corporations to adapt to the new business environment where the pace of change has accelerated in recent years (**Fig. 5**). This is indicated by the rising demands for support to achieve successful global communication in the multinational labor environment and to expand

Changes in business environment	Aging of society
Overseas business expansion by Japanese companies	Approx. 24% of the elderly live alone.
Multinational working environment	The number of care provider applicants is about half the number in demand.
Accelerating business speed	Society is also aging rapidly in China, India, and other countries.
Advancement of communication tools	The working-age population is decreasing.
(1) Improvements in business productivity - Improvements in global communication - Speeding up of business and improvements in quality	(2) Life support functions - Improvements in communication among medical personnel - Daily life support for the elderly and people with disabilities

Fig. 5. Challenges and application areas.

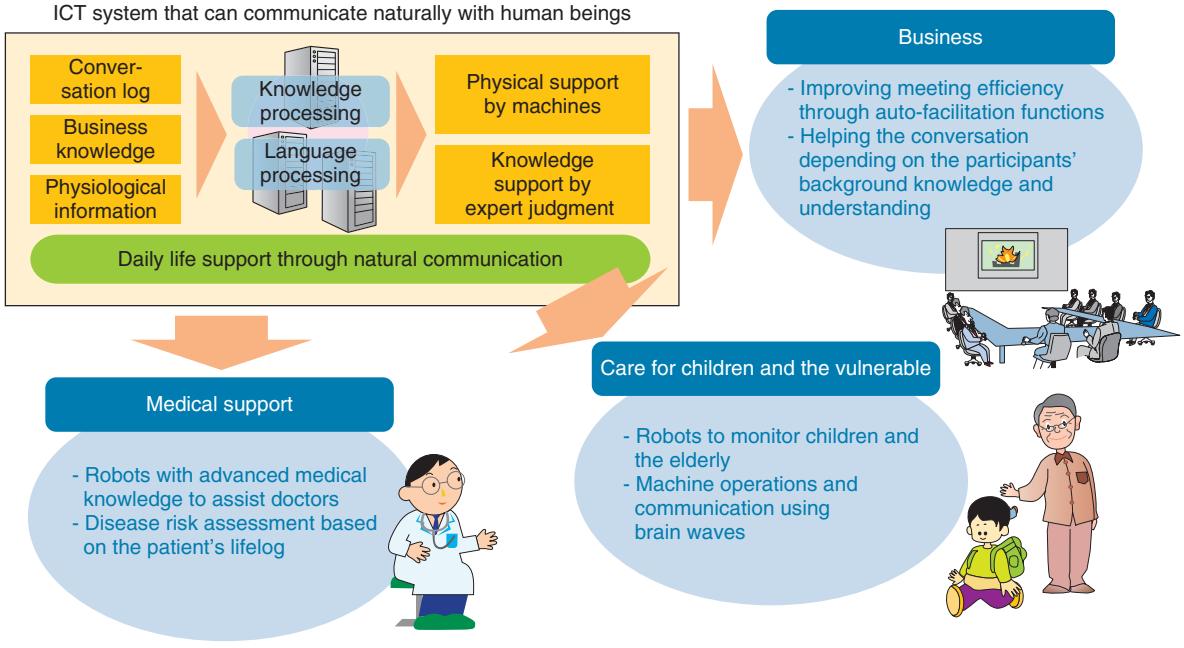


Fig. 6. Future vision achievable through communication advancement.

domestic business overseas.

Moreover, we are facing an unavoidable social challenge—a super-aging society. The number of elderly living alone is increasing while the working-age population is decreasing. The number of care provider applicants is about half the demand for them, and this will inevitably lead to the situation in which people in need of care cannot receive sufficient services. It is important for the coming Japanese society to make advances in communication over the ICT infrastructure. This will enable better communication between medical personnel and care providers to

increase their productivity and build an environment where the elderly and people with disabilities can live independently as much as possible.

6. Vision of the future

An ICT system that can communicate naturally with human beings is expected to become commonplace in all aspects of business and social activities (**Fig. 6**). It will be partially composed of robots and will provide a range of intelligent support based on special judgments obtained from the results of knowledge processing or language processing of various data, such as physiological information collected through sensors, business information, and voice information. For example, additional information helpful for a business meeting could be automatically added by the system according to the individual degrees of knowledge, opinions, behaviors, and levels of understanding of the meeting's participants. The system could even facilitate the smooth progress of meetings in the near future.

We also assume that new medical support services will gradually come into use, such as assistant robots with advanced medical knowledge and services to suggest a healthier life style on the basis of disease assessment results generated from the daily activities and lifelogs of individuals.

7. Future prospects

In a society where globalization and aging are progressing, the communication gap will become even bigger. To eliminate such a gap, NTT DATA is continuing to develop an ICT infrastructure and application software that utilize the optimum media analysis technology and device technology solutions in order to advance communications. Through these efforts, we aim to help improve productivity in business and society.

Reference

[1] T. Obi and N. Iwasaki, "ICT Innovation Targeted for the Elderly Will Save Super-aging Society," The Mainichi Newspapers, 2011.

Toru Takaki
Manager, Service Innovation Center, Research and Development Headquarters, NTT DATA Corporation.
He received the B.E., M.E., and Ph.D. degrees in information science from the University of Tsukuba, Ibaraki, in 1990, 1992, and 2005, respectively. He joined NTT DATA Communications Systems Corporation (since 1996, NTT DATA Corporation) in 1992. He has been working on R&D of information retrieval, text processing, question answering, and machine translation. He is a member of the Information Processing Society of Japan and the Association of Computing Machinery.

512 × 512 Port 3D MEMS Optical Switch Module with Toroidal Concave Mirror

Yuko Kawajiri, Naru Nemoto, Koichi Hadama, Yuzo Ishii, Mitsuhiro Makihara, Joji Yamaguchi, and Tsuyoshi Yamamoto

Abstract

We present a 512 × 512 MEMS (microelectromechanical system) optical switch module in a W-shaped layout with a toroidal concave mirror. The 512-array optical components are made by assembling four 128-array components. The concave mirror minimizes the tilt angle of the MEMS mirrors. All of the optical path connections were demonstrated in a prototype module.

1. Introduction

An all-optical cross connect will be a key component of large-capacity photonic networks and data-center networks with low power consumption because it provides large-scale switching without optical-to-electrical-to-optical (O-E-O) conversion. A promising way to make large-scale optical cross connects (OXCs) is to use three-dimensional (3D) microelectromechanical system (MEMS) optical switches because of the potentially large port count and the compact configuration that can be achieved using free-space optics [1]. A 3D MEMS optical switch basically consists of a pair of optical collimator arrays as the input and output (I/O) ports and a pair of two-axis MEMS tilt mirror arrays to steer the optical beams so that any input port can be connected to an arbitrary output port.

We previously reported a 100 × 100 port 3D MEMS optical switch module that provides good switching characteristics [2]–[4]. In this article, we describe a large-scale optical switch with a port count of over 500 and confirm the scalability of 3D MEMS optical switches.

2. Principle and design

When the port count exceeds a few hundred, the requirement for the maximum tilt angle of MEMS mirrors becomes exacting because of the expanded connection area. A cross-sectional view of the optics in a conventional 3D MEMS optical switch is schematically shown in **Fig. 1**. The I/O ports and a pair of MEMS mirror arrays are arranged in a Z-shaped layout. In this configuration, the maximum tilt angle for switching depends on a mirror's position in the array.

For example, let us denote by θ the angle that the central mirror needs to tilt by (dashed and dotted green lines). In that case, the mirror at the corner of the input MEMS mirror array connects to the opposite corner of the output MEMS mirror array when its tilt angle is 0° (solid blue line); however, it must tilt to one side on each axis with the maximum angle 2θ to connect to the diagonally opposite corner of the output MEMS mirror array (dot-dashed blue line). This position dependency makes the required maximum tilt angle excessively large.

One way to keep the tilt angle from increasing is to apply an optical Fourier transform between the two mirror arrays to make the required tilt angle the same

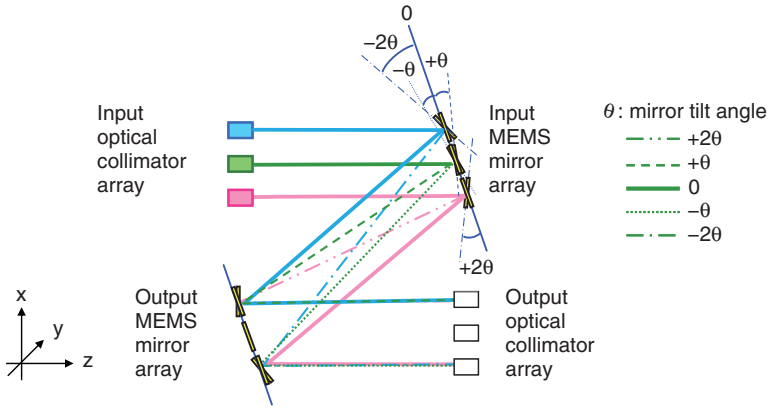


Fig. 1. Cross-sectional view of a conventional 3D MEMS optical switch.

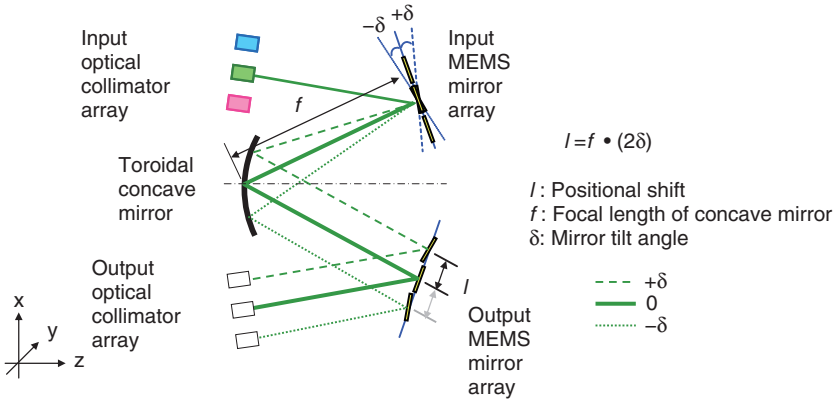


Fig. 2. Cross-sectional view of the 3D MEMS optical switch with the toroidal concave mirror.

for all the mirrors [5]. We use a toroidal concave mirror as the optical Fourier transform element. A concave mirror also enables a compact W-shaped layout with a folded optical path.

Another problem is how to produce large-scale optical components. As the array scale is increased, the process yield decreases significantly owing to the difficulty of fabrication. To solve this problem, we chose to use a 2 × 2 array of 128-port optical units to obtain a 512-port optical component. This achieves high fabrication yields with a small accumulated pitch error.

2.1 Basic switch structure

A cross-sectional view of the configuration of our 512 × 512 port 3D MEMS optical switch module is shown in **Fig. 2**. The input and output MEMS mirror arrays are placed on the concave mirror's two focal planes.

An optical beam deflected by an input mirror strikes the concave mirror at an incident angle that is related to the mirror's tilt angle δ. The concave mirror provides a Fourier transform, causing the optical beam to converge onto the output mirror array with a positional shift l. The angle of each output mirror is adjusted so that the beam is reflected into the proper output collimator.

The shift l is expressed by using the concave mirror's focal length f as

$$l = f \cdot 2\delta.$$

This expression means that the connecting output mirror is determined by the input mirror's tilt angle, without positional dependence in the mirror array. Thus, the increase in the maximum tilt angle of the

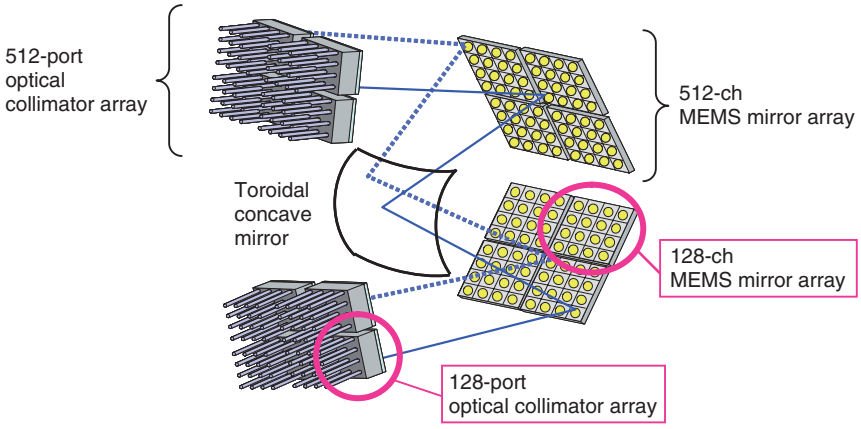
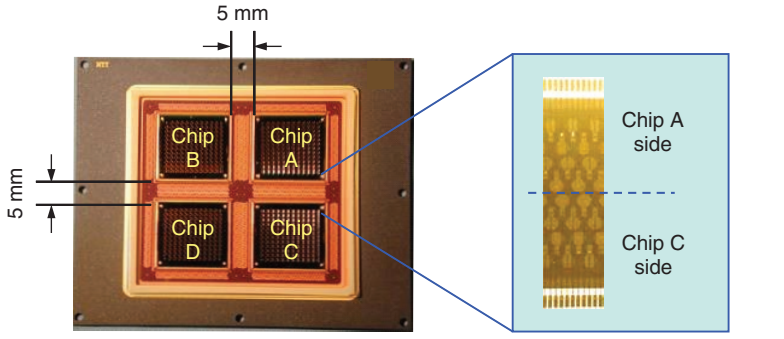


Fig. 3. Schematic of the 3D MEMS optical switch with the toroidal concave mirror.



(a) 512-ch MEMS mirror array (b) Zigzag alignment bonding pads

Fig. 4. 512-channel MEMS mirror array.

MEMS mirrors is minimized throughout the mirror array.

Folding the optical path into a W-shaped configuration enables a compact layout, but it also introduces the off-axis aberrations of the concave mirror in the x-z plane. We chose a toroidal surface for the concave mirror's geometry to reduce the aberrations. For the optimized mirror geometry, the calculated loss due to aberrations is below 0.5 dB.

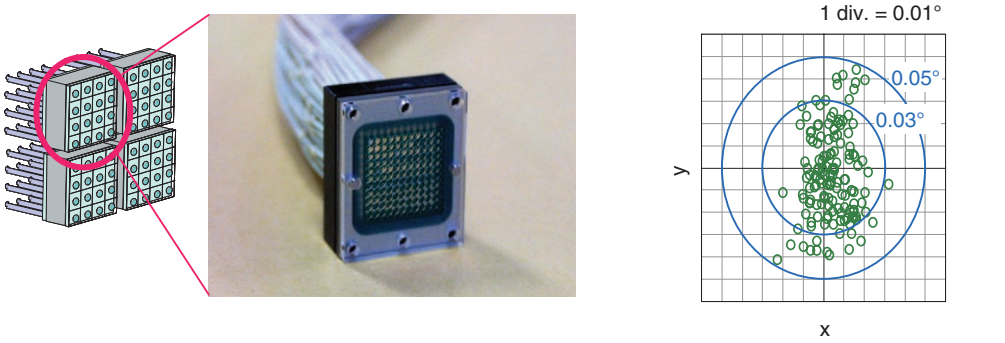
A schematic of our 512 × 512 port 3D MEMS optical switch module is shown in **Fig. 3**. The mirror and collimator arrays consist of 2 × 2 arrays, each handling 128 ports. The 128-port optical components provide high fabrication yields with a small accumulated pitch error.

2.2 512-channel MEMS mirror array

The 512-channel MEMS mirror array is shown in **Fig. 4**. It contains four 128-channel MEMS mirror array chips, which are precisely mounted in a ceramic package by multichip-module technology. The positioning accuracy of mounting is less than 50 μm.

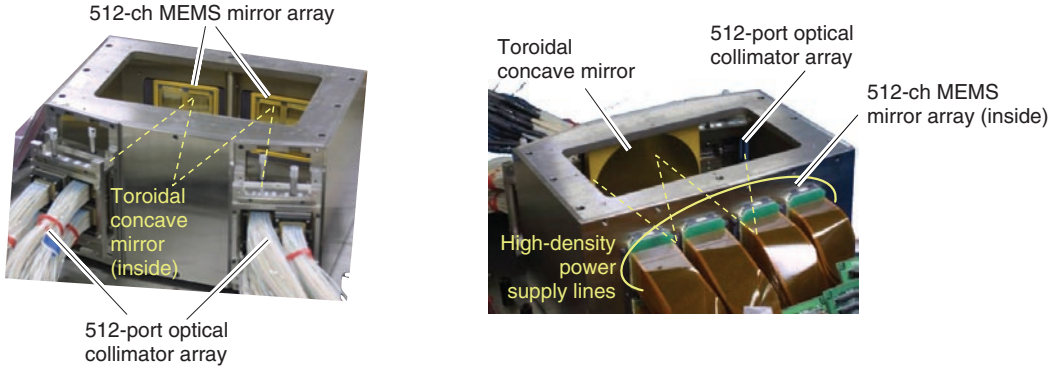
While the huge number of bonding pad electrodes should be assembled in a package, a compact ceramic package is needed to expand the possibility of the optical layout design. Our high-density bonding pads in a zigzag arrangement minimize the package size to 76 mm × 90 mm and the spacing between adjacent chips to 5 mm. The increase in mirror tilt angle needed to accommodate this spacing is only 0.6°.

The 128-channel mirror array chip has a two-dimensional (2D) arrangement of two-axis MEMS tilt mirrors. Each mirror has a gimbal structure that allows the mirror to rotate around two orthogonal axes [4]. The mirror is actuated electrostatically by four electrodes underneath it, resulting in a tilting



(a) 128-port collimator unit (b) Pointing accuracy of collimator unit

Fig. 5. 512-port optical collimator array.



(a) Front view (b) Rear view

Fig. 6. Prototype switch core module of the 512 × 512 port 3D MEMS optical switch.

range of 4.5° in any arbitrary direction.

2.3 Optical collimator array

The 512-port optical collimator array is also composed of four 128-port units. They are mounted on a frame with a positioning mechanism that helps align them with the mirror array. Each collimator array unit consists of an array of 128 fibers and an array of 128 microlenses, as shown in **Fig. 5(a)**.

The 128-port fiber array is made by inserting optical fibers with microferrules through a polymer substrate having precisely aligned holes. Each optical fiber is attached to a microferrule that had its end facet polished and antireflection-coated before assembly. This fabrication method is advantageous to improve the yield of a large-scale, highly accurate fiber array inexpensively [6], [7]. The 128-port microlens array is fabricated by a precise molding method using transparent optical polymer material. Both surfaces of the microlens array are also antireflection coated to reduce multiple reflections. The fiber array and microlens array are passively aligned using dowel pins with an accuracy of ±1 μm.

The mean pointing accuracy, which is the mean angular deviation caused by a lateral offset of the axis of a microlens from that of the corresponding fiber, is 0.03°. The uniformity of our optical collimator's pointing accuracy is shown in **Fig. 5(b)**. This is adequate for good optical coupling.

2.4 Switch module assembly

For switch module assembly, the critical issues are precise alignment and the assembly procedure itself: both are time consuming and costly. To reduce the cost, we use passive assembly to construct the optical system. A prototype 512 × 512 port switch module is

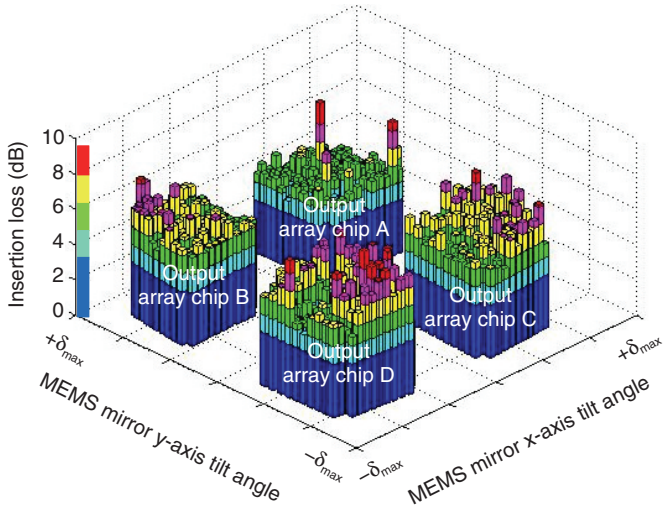


Fig. 7. Characteristics of optical path connections to all output ports from an input mirror.

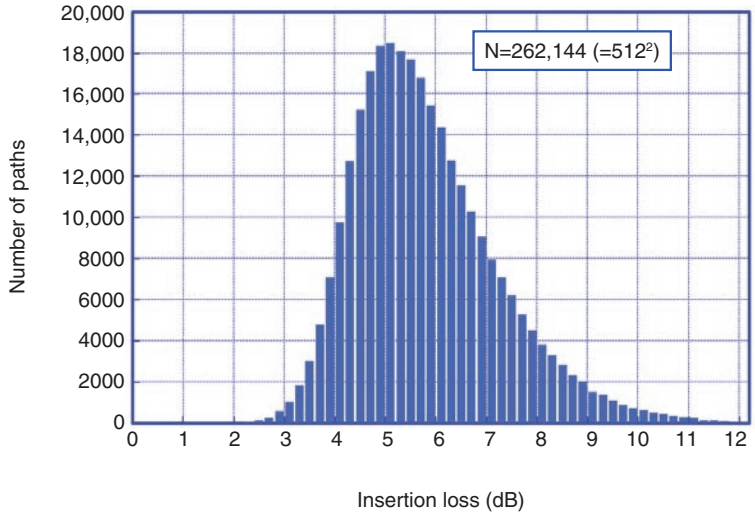


Fig. 8. Insertion losses for all the paths of the prototype module.

shown in **Fig. 6**. The module size is 197 mm × 243 mm × 140 mm. All components are passively mounted on the housing by using dowel pins and holes. Before the concave mirror is set in the housing, the collimator arrays are visually aligned to the corresponding MEMS mirror arrays by observation using an infrared camera.

3. Optical performance of the switch module

Regarding the optical performance, we measured the insertion loss of all the paths. A fast peak search algorithm enables accurate optical path connections at short measuring times [8]. The characteristics of the optical connections from an input mirror to all the output mirrors are shown in **Fig. 7**. This input mirror was located in the top-left corner of the array, which is at the maximum off-axis position. The vertical axis is insertion loss, and the horizontal axes are the tilt angles of the input mirror about the x- and y-axes. The maximum tilt angle is the same for both axes. The variation in mean insertion loss for the four

Table. 1. Breakdown of insertion loss sources.

Insertion loss sources	Min. (dB)	Max. (dB)
Beam clipping per mirror	0.5	3.8
Fiber connectors	0.0	0.8 for 2 points
Aberration of concave mirror (calculated)	0.1	0.5
Antireflection coat (calculated)	0.12 for 14 points	-

output MEMS chips was less than 1 dB, which means that there were no significant differences among them. We also found that the characteristics of all the optical connections were roughly uniform.

The distribution of the insertion loss for all the paths of the prototype module is shown in **Fig. 8**. The number of paths is 262,144 (i.e., 512^2). The mean fiber-to-fiber insertion loss is 5.3 dB. A breakdown of the insertion loss sources is given in **Table 1**. The main one is clipping by the MEMS mirrors. The misalignment of a collimator caused a maximum beam position error of 250 μm, which is equivalent to a clipping loss of 3.8 dB per mirror. It should be possible to reduce the misalignment to less than 50 μm by improving the assembly accuracy. This should reduce the maximum clipping loss to less than 1.5 dB.

4. Conclusions

We have presented a free-space 512 × 512 port 3D MEMS optical switch module featuring a W-shaped optical configuration with a toroidal concave mirror. This configuration keeps the increase in the maximum tilt angle of the MEMS mirrors small. We also devised 512-port optical components, each consisting of four 128-port units. This design provides both a large port count and a low cost. The results of optical path tests on a prototype switch module show its feasibility for constructing a large-scale optical switch with a port count of over 500.

References

[1] V. A. Aksyuk, F. Pardo, D. Carr, D. Greywall, H. B. Chan, M. E. Simon, A. Gasparyan, H. Shea, V. Lifton, C. Bolle, S. Arney, R. Frahm, M. Paczkowski, M. Haueis, R. Ryf, D. T. Neilson, J. Kim, C. R. Giles, and D. Bishop, "Beam-steering micromirrors for large optical cross-connects," IEEE Journal of Lightwave Technology, Vol. 21, No. 3, pp. 634–642, 2003.

[2] T. Yamamoto, J. Yamaguchi, N Takeuchi, A. Shimizu, E. Higurashi, R. Sawada, and Y. Uenishi, "A Three-Dimensional MEMS Optical Switching Module Having 100 Input and 100 Output Ports," IEEE Photonics Technology Letters, Vol. 15, No. 10, pp. 1360–1362, 2003.

[3] T. Yamamoto, J. Yamaguchi, R. Sawada, and Y. Uenishi, "Development of a Large-scale 3D MEMS Optical Switch Module," NTT Technical Review, Vol. 1, No. 7, October, pp. 37–42, 2003. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200310037.pdf

[4] J. Yamaguchi, T. Sakata, N. Shimoyama, H. Ishii, F. Shimokawa, and T. Yamamoto, "High-yield Fabrication Methods for MEMS Tilt Mirror Array for Optical Switches," NTT Technical Review, Vol. 5, No. 10, 2007. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200710sp5.html

[5] V. A. Aksyuk, S. Arney, N. R. Basavanhally, D. J. Bishop, C. A. Bolle, C. C. Chang, R. Frahm, A. Gasparyan, J. V. Gates, R. George, C. R. Giles, J. Kim, P. R. Kolodner, T. M. Lee, D. T. Neilson, C. Nijander, C. J. Nuzman, M. Paczkowski, A. R. Papazian, F. Pardo, D. A. Ramsey, R. Ryf, R. E. Scotti, H. Shea, and M. E. Simon, "238 x 238 Micromechanical Optical Cross Connect," IEEE Photonics Technology Letters, Vol. 15, No. 4, pp. 587–589, 2003.

[6] T. Yamamoto, J. Yamaguchi, N Takeuchi, A. Shimizu, R. Sawada, E. Higurashi, and Y. Uenishi, "A Three-dimensional Micro-electro-mechanical System (MEMS) Optical Switch Module Using Low-cost Highly Accurate Polymer Components," Japanese Journal of Applied Physics, Vol. 43, pp. 5824–5827, 2004.

[7] T. Yamamoto, J. Yamaguchi, T. Takeuchi, A. Shimizu, R. Sawada, E. Higurashi, and Y. Uenishi, "A three-dimensional MEMS optical switch module using low-cost highly accurate polymer components," Proc. of the 9th Microoptics Conference (MOC'03), paper-F2, pp. 92–95, Tokyo, Japan, 2003.

[8] N. Nemoto, M. Mizukami, Y. Kawajiri, and J. Yamaguchi, "Switching property of 3D-MEMS optical switch," Proc. of IEICE Tech. Rep., Vol. 110, No. 395, OPE2010-163, pp. 133–136, Osaka, Japan, 2011.

Yuko Kawajiri
Senior Research Engineer, Optical Network Device Research Group, Network Hardware Integration Laboratory, NTT Microsystem Integration Laboratories.

She received the B.S. and M.S. degrees in chemistry from Gakushuin University, Tokyo, in 1993 and 1995, respectively. She joined NTT Interdisciplinary Research Laboratories in 1995. She has been engaged in research on optical interconnection devices and passive wavelength-division-multiplexing filter modules. She is currently working on the development of free-space optical switches for optical networks. She received the Microoptics Conference Best Paper Award in 2009. She is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

Naru Nemoto
Researcher, Optical Network Device Research Group, Network Hardware Integration Laboratory, NTT Microsystem Integration Laboratories.

He received the B.E. and M.E. degrees in inorganic materials from Tokyo Institute of Technology in 2003 and 2005, respectively. He joined NTT Telecommunication Energy Laboratories in 2005. He has been engaged in research on MEMS motion control with free-space optical switches. He received the JSPE Best Presentation Award in 2009 and JSPE Young Engineer Award in 2010. He is a member of the Japan Society of Mechanical Engineers and the Japan Society for Precision Engineering (JSPE).

Koichi Hadama
Senior Research Engineer, Optical Network Device Research Group, Network Hardware Integration Laboratory, NTT Microsystem Integration Laboratories.

He received the B.E. and M.S. degrees in applied physics from the University of Tokyo in 1999 and 2001, respectively. He joined NTT Microsystem Integration Laboratories in 2001. He is currently engaged in the development of free-space optical modules for fiber communication networks. He is a member of IEICE and the Optical Society of Japan.

Yuzo Ishii
Senior Research Engineer, Optical Network Device Research Group, Network Hardware Integration Laboratory, NTT Microsystem Integration Laboratories.

He received the B.S., M.S., and Ph.D. degrees in precision machinery engineering from the University of Tokyo in 1995, 1997, and 2005, respectively. In 1997, he joined NTT Optoelectronics Laboratories, Tokyo, where he engaged in research on micro-optics for chip-to-chip optical interconnection. During 2005–2006, he was a visiting researcher at Vrije University Brussels, Belgium. Since 2006, he has been engaged in the development of a wavelength selective switch using MEMS technology. He received the Microoptics Conference Best Paper Award in 1999. He is a member of the Japan Society of Applied Physics (JSAP).

Mitsuhiro Makihara
Senior Research Engineer, Optical Network Device Research Group, Network Hardware Integration Laboratory, NTT Microsystem Integration Laboratories.

He received the B.E. and M.E. degrees in mechanical engineering from Kyushu Institute of Technology, Fukuoka, in 1987 and the University of Tokyo in 1989, respectively. He joined NTT Applied Electronics Research Laboratories in 1989. He has been engaged in research on the optical switching equipment for the optical communications. He received the JSPE Technology Development Award 1999. He is a member of IEICE and the Japan Society of Mechanical Engineers (JSME).

Joji Yamaguchi
Senior Research Engineer, Supervisor, Optical Network Device Research Group, Network Hardware Integration Laboratory, NTT Microsystem Integration Laboratories.

He received the B.E., M.E., and Ph.D. degrees in mechanical engineering, all from Tokyo Institute of Technology in 1988, 1990, and 1993, respectively. In 1993, he joined NTT Interdisciplinary Research Laboratories, where he engaged in research on OXC systems. During 2000–2001, he studied MEMS control technology as a visiting researcher at the University of California, Berkeley, CA, USA. Recently, he has been involved in research on 3D MEMS optical switches for large-scale OXCs. He is a member of the Japan Society of Mechanical Engineers and JSPE.

Tsuyoshi Yamamoto
Senior Research Engineer, Supervisor, Group Leader of Optical Network Device Research Group, Network Hardware Integration Laboratory, NTT Microsystem Integration Laboratories.

He received the B.E. degree in electrical engineering from Kansai University, Osaka, in 1991. In 1991, he joined NTT Communication Switching Laboratories, where he engaged in R&D of several optical interconnection systems using free-space optics. Recently, he has been involved in R&D of large-scale optical switches for next-generation ROADM systems. During 1998–1999, he was a visiting research engineer at the Department of Electrical and Computer Engineering, McGill University, Quebec, Canada. During 2007–2011, he was the director of the Optical Node System Department, Optical Communication Systems Group at NTT Electronics Corp. He received the Microoptics Conference Best Paper Awards in 2003 and 2009, respectively. He is a member of IEEE.

On the Security of the Cryptographic Mask Generation Functions Standardized by ANSI, IEEE, ISO/IEC, and NIST

Koutarou Suzuki and Kan Yasuda

Abstract

We revisit the security of mask generation functions (MGFs) in light of the indifferentiability framework. MGFs are a kind of hash function having variably long outputs and they are frequently utilized for designing public-key cryptographic schemes such as digital signatures. First, we clarify that there are weak and strong versions of indifferentiability, depending on the order of quantifies in the definition. Next, we prove that the classical, counter-based MGF standardized by ANSI, IEEE, and ISO/IEC satisfies only the weak version of indifferentiability, whereas the Double-Pipeline Iteration Mode specified in SP800-108 by the National Institute of Standards and Technology (NIST) satisfies the strong version. While our analysis does not necessarily imply that counter-based MGFs are insecure, our results show that MGF constructions have different levels of security (i.e., indifferentiability).

1. Introduction

1.1 Background

Ever since its establishment by Bellare and Rogaway [1], the notion of *random oracles* has played an essential role in the design of asymmetric cryptographic schemes [2], [3]. Informally, random oracles are objects that should behave like public random functions, accepting variable input length (VIL) data and returning variable output length (VOL) random strings. Random oracles are ideal objects: they cannot be implemented without additional assumptions. In practice, random oracles are replaced with concrete functions.

It is not an easy task to construct a random-looking VIL-VOL concrete function from scratch. So we usually start with a small concrete function that is restricted to a fixed input length (FIL) and fixed output length (FOL). Such functions are often called compression functions. We then iterate the compression functions in some way to obtain VIL and/or VOL functions.

Concrete functions that accept VIL strings but return only FOL strings are commonly called hash functions. The construction of *secure* hash functions has been theoretically investigated in various ways. In particular, the security of hash functions as VIL (but FOL) random oracles was studied by Coron et al. [4], where the underlying compression functions were modeled as FIL-FOL (restricted) random oracles in light of the indifferentiability framework [5]. Subsequent to the work reported in [4], the domain extension of random oracles has been analyzed in depth [6]–[11].

On the other hand, the range extension of random oracles has attracted less attention from the theoretical aspect. Despite the lack of formal treatment, VOL random oracles are regularly used in designing public-key cryptographic schemes, in particular digital signatures [3]. For the random oracles utilized in those signature schemes, which achieve full message recovery [12]–[16], the *variability* in output length becomes absolutely crucial.

There already exist several constructions for

Table 1. Summary of our results.

Definition of indifferentiability	Construction	
	Counter-based MGF	Chained MGF
Local ($\forall A \exists S$, Maurer et al. [5])	Secure [Theorem 1]	Secure [Theorem 3]
Universal ($\exists S \forall A$, Coron et al. [4])	Cannot be proven [Theorem 2]	Secure [Theorem 3]

Note: By "cannot be proven", we mean that it is impossible to prove that the construction is secure. We prove impossibility rather than give an attack; we do not mean that the construction is insecure.

VIL-VOL concrete functions. They are called by the common name mask generation functions (MGFs). The majority of existing MGFs follow the counter-based design and have been standardized by ANSI (American National Standards Institute), IEEE (Institute of Electrical and Electronics Engineers), and ISO/IEC (International Organization for Standardization, International Electrotechnical Commission). For example, the algorithm MGF1 [13], [17]–[19] takes a hash function H: $\{0, 1\}^* \longrightarrow \{0, 1\}^n$, computes upon input x the string

$$H\,(x\|\langle 0\rangle_{32}) \,\|\, H\,(x\|\langle 1\rangle_{32}) \,\|\, H\,(x\|\langle 2\rangle_{32}) \;\|\; \cdots,$$

and truncates this string to the leftmost l bits, where $\langle i\rangle_\alpha$ denotes an α-bit representation of integer i and l denotes the requested length. The main motivation behind the current work is to provide a formal security analysis for this type of construction.

The same types of algorithms are often called key derivation functions (KDFs), mostly when they take secret inputs. These are standardized in SP800-108 by NIST (National Institute of Standards and Technology) [20]. The security of KDFs is formally treated in [21]. We analyze the security of the Double-Pipeline Iteration Mode specified in SP800-108 as an MGF that takes only public inputs.

1.2 Our results

We take the systematic approach proposed by Coron et al. [4] and apply the indifferentiability framework [5] to our study of MGFs. That is, we consider two MGF constructions whose security is analyzed under the condition that an ideal hash function (a VIL/FOL random oracle) H: $\{0, 1\}^* \longrightarrow \{0, 1\}^n$ is given. Using this basic strategy, we obtain the following results, which are summarized in **Table 1**:

- Local vs. universal. We point out that in the literature there are two different versions of indifferentiability notions.
- Analysis of counter-based MGFs. We obtain two impossibility results for the counter-based MGF1. The first result says that MGF1 cannot be proven to be indifferentiable from the ideal MGF in the sense that there exist no *natural* simulators. The second result says that MGF1 *itself* cannot be proven to be insecure in the sense that there exists no *strong* adversary.
- Analysis of chained MGFs. We analyze the security of the Double-Pipeline Iteration Mode specified in NIST SP800-108, which can be shown to be indifferentiable from an ideal MGF. We provide concrete security bounds for the Double-Pipeline Iteration Mode. Unlike the case of domain extension, the security of this scheme degrades only *linearly* with the number of oracle queries[*1].

1.3 Organization

Section 2 defines our notation and provides other preliminaries. Section 3 reviews the notion of indifferentiability, identifies a class of natural simulators, and defines an MGF. Section 4 defines the counter-based MGF and analyzes its security. Section 5 defines the Double-Pipeline Iteration Mode and analyzes its security. Section 6 concludes with a brief summary and some concluding remarks about future work.

2. Preliminaries

2.1 Basic notation

$\{0, 1\}^m$ denotes the set of bit strings whose length is equal to $m > 0$. $\{0, 1\}^0$ denotes the set consisting of only the null string ε. $\{0, 1\}^*$ denotes the set of finite bit strings.

$|x|$ denotes the bit length of a string $x \in \{0, 1\}^*$.

$[x]^m$ represents the leftmost m bits of a string

*1 In the case of domain extension, a collision in the chaining values immediately leads to insecurity, which implies that the degradation is quadratic in query complexity.

$x \in \{0, 1\}^*$. $[x]_m$ represents the rightmost m bits.

Given two strings x and y, we let $x \| y$ be the concatenation of x and y.

$\lceil m \rceil$ indicates the smallest integer greater than or equal to an integer m.

We write $\mathbf{N}$ for the set of positive integers and write $\mathbf{Z}_{\geq 0}$ for the set of nonnegative integers.

By writing $x \in_U X$, we mean that x is an element chosen uniformly at random from the set X.

2.2 Security parameters and length encoding

A security parameter is a positive integer $\kappa \in \mathbf{N}$. It is customary to write $1^\kappa \in \{0, 1\}^*$ instead of $\kappa \in \mathbf{N}$ to emphasize the fact that κ is a security parameter. Whenever possible, we omit the security parameter κ and make it implicit in our statements.

2.3 Oracle machines

Throughout the paper, the computation model is fixed. Specifically, we regard any probabilistic algorithm as a (probabilistic) Turing machine. We consider an oracle machine, which is a Turing machine given access to an oracle. Interaction with an oracle is done via the machine's communication tape, and a reply from an oracle is given immediately, i.e., the time for interaction is 1 (unit time) irrespective of the query length, the reply length, and the oracle's behavior. Note, however, that the machine consumes the time taken to write its query onto the communication tape. Moreover, if the machine wants to read partially or wholly the reply written on the tape, the corresponding amount of time is consumed.

We write $A^{\mathcal{O}}$ to indicate the fact that a Turing machine A interacts with an oracle $\mathcal{O}$. We also let $A^{\mathcal{O}}$ denote the output value returned by A after its interaction with $\mathcal{O}$. We can always replace $\mathcal{O}$ with any other machine B that has a compatible interface, in which case we write A^B. We write $A^{\mathcal{O}_1, \mathcal{O}_2, \cdots}$ when A has access to multiple oracles.

2.4 Modes and distinguishers

A mode is a deterministic algorithm M that takes as its input a security parameter 1^κ and a finite string $x \in X$, where domain X is a subset of $\{0, 1\}^*$, and computes as its output a finite string $y \in \{0, 1\}^*$. A mode M has access to an oracle $\mathcal{H}$, and the interface between M and $\mathcal{H}$ depends on the security parameter κ. In other words, we can consider a family of oracles $\{\mathcal{O}_\kappa\}_\kappa$, from among which an appropriate oracle is chosen by M according to the value κ. Succinctly, we can write $y \leftarrow M^{\mathcal{O}_\kappa}(1^\kappa, x)$. Obviously, the algorithm $M^{\mathcal{H}}$ may not be deterministic if $\mathcal{H}$ is not, even though the mode M itself must be deterministic.

A distinguisher is a probabilistic algorithm D that takes as its input a security parameter 1^κ and outputs a bit $b \in \{0, 1\}$. A distinguisher D is given access to multiple oracles, and one of them is frequently mode M. In such a setting, we say that "the distinguisher D attacks the mode M." Succinctly written, $b \leftarrow D^{M^{\mathcal{O}_\kappa}(1^\kappa, \cdot), \cdots}(1^\kappa)$. Note that the same security parameter κ is used for both D and M.

2.5 Time and query complexities

Generally speaking, we may want to restrict the capacity of an oracle machine in terms of its time complexity and query complexity. In the present work, however, we treat only query complexity because an oracle machine's running time is irrelevant to the context of our security analysis[*2]. The query complexity is measured in terms of two quantities q_A and l_A for a given oracle machine A, where q_A represents the limit on the total number of queries that machine A can send to its oracles and l_A represents the limit on the maximum length of each query or reply.

A construction F is said to be *tractable* if its bounds q_F and l_F are polynomials in the following three variables: security parameter κ, input length $|x|$, and output length $|y|$. A distinguisher D is said to be *efficient* if its bounds q_D and l_D are polynomials in the security parameter κ. A simulator S is said to be *efficient* if its bounds q_S and l_S, as well as the size $|\sigma'|$ of updated state σ', are polynomials in the following four variables: security parameter κ, input state length $|\sigma|$, input length $|x|$, and output length $|y|$.

3. Indifferentiability framework and security of the MGF

In this section, we revisit the notion of indifferentiability. There are two points that we would like to clarify: (1) the definition of a simulator and (2) the order of quantifiers with respect to the simulator. Now, we define the security of the MGF.

3.1 Simulator division and connector extraction

In order to define indifferentiability, we need to introduce a simulator. A simulator S is a probabilistic algorithm that takes as its input a security parameter 1^κ, current state information $\sigma \in \Sigma$ (where the set Σ of state information is a subset of $\{0, 1\}^*$), and an input value $x \in X$ (where the domain X is a subset of $\{0, 1\}^*$)

*2 In our analysis, the source of randomness always involves random oracles, and we never deal with computational assumptions.

and computes as its output a pair of updated state information $\sigma' \in \Sigma$ and a finite string $y \in \{0, 1\}^*$. For convenience, we assume that the empty string ε is in the set Σ. Simulator S is always an oracle machine having access to some oracle M. Succinctly, we can write

$$(\sigma', y) \leftarrow S^M(1^\kappa, \sigma, x).$$

S's goal is to mimic some oracle $\mathcal{O}_\kappa$: $X \longrightarrow \{0, 1\}^*$ that is expected to return $y \in \{0, 1\}^*$ in response to the query $x \in X$.

We introduce a connector C, which is a dummy functionality whose purpose is merely to connect simulator S to an oracle machine D. A connector C is a stateful machine; that is, it has an internal memory that can store current state information $\sigma \in \Sigma$. The state σ is initially set to the empty string ε. Connector C works as follows. Upon receiving an oracle query $x \in X$ from distinguisher D, connector C forwards (σ, x) to simulator S and lets S compute $(\sigma', y) \leftarrow S^M(1^\kappa, \sigma, x)$. Connector C receives the output (σ', y) from S, updates its own state information from σ to σ', and returns the value y to D.

Consider a distinguisher $D^{\mathcal{O}_\kappa}$ interacting with an oracle $\mathcal{O}_\kappa$: $X \longrightarrow \{0, 1\}^*$. We can replace the oracle $\mathcal{O}_\kappa$ with the machine C^S and hence obtain D^{C^S}. Since the connector C does nothing but provide a trivial interface, we write (with abuse of notation) D^S instead of D^{C^S}.

3.2 Definition of indifferentiability: local vs. universal

There are two different versions of the indifferentiability notion. The setting for the notion of indifferentiability is as follows. Let D be an adversary. D's goal is to distinguish between the real world and the ideal world. In either world, D has access to two oracles. In the real world, we define efficient construction F having access to oracle ϕ to be indifferentiable from oracle Φ as follows. Consider a polynomial-time simulator S having access to oracle Φ and trying to simulate ϕ. Simulator S has complete knowledge of F. Consider a polynomial-time adversary D that has access to two oracles and is expected to output a bit at the end of each game execution. D has complete knowledge of not only F but also S. The notion of indifferentiability for F (together with S and D) is given by the following two different games: in the real game, D is given access to two oracles F and ϕ, while in the ideal game, D is given access to two oracles ϕ and S. We define the advantage $\mathrm{Adv}^{\mathrm{indiff}}_{F,S}(D)$ of adversary D as

$$\mathrm{Adv}^{\mathrm{indiff}}_{F,S}(D) = \left| \Pr[D^{F^\phi, \Phi} = 1] - \Pr[D^{\Phi, S^\phi} = 1] \right|,$$

where the probability is taken over the coin tosses ϕ and Φ.

Definition 1 (Local: Maurer et al. [5]). *Let F be an efficient construction. We say that F is indifferentiable from the random oracle (in the sense of Maurer et al.'s definition) if for any polynomial-time adversary D there exists an efficient simulator S and a negligible function $\epsilon(\kappa)$ such that the inequality $\mathrm{Adv}^{\mathrm{indiff}}_{F,S}(D) \leq \epsilon$ holds.*

Definition 2 (Universal: Coron et al. [4]). *Let F be an efficient construction. We say that F is indifferentiable from the random oracle (in the sense of Coron et al.'s definition) if there exists an efficient simulator S such that for any polynomial-time adversary D there exists a negligible function $\epsilon(\kappa)$ satisfying the inequality $\mathrm{Adv}^{\mathrm{indiff}}_{F,S}(D) \leq \epsilon$.*

To avoid confusion, we give specific names to these two notions: we say that an efficient construction F *is locally indifferentiable if* it is indifferentiable in the former sense and *universally indifferentiable* if it is indifferentiable in the latter sense. Clearly, universal indifferentiability implies local indifferentiability.

Remark 1. It seems that the two definitions arise from the difference in purpose. The main purpose of the former definition is to discuss security under the system compositions, and the definition indeed gives a necessary and sufficient condition for composability. On the other hand, the purpose of the latter is to measure *how good* a construction is, because the existence of a universal simulator shows that it is indeed a good construction, with the simulator being the inverse construction.

Remark 2. We emphasize that adversaries D and simulators S are merely algorithms (Turing machines). Hence, a simulator S is not allowed to *observe* the queries/replies made in the interaction between D and Φ[*3]. Moreover, note that D is not allowed to observe the running time of oracles with which it interacts because any oracle interaction takes exactly unit time.

*3 In fact, when D is interacting with the Φ-oracle, simulator S is not even invoked.

3.3 Definition of MGF functionality/security

To formalize the functionality of MGFs, we define MGFs and their corresponding random oracle (i.e., ideal MGF function), and we define hash functions and their corresponding random oracle (i.e., ideal hash function).

We start by giving a definition of an MGF. Intuitively, an MGF is defined as a concrete function that takes as its input a seed x together with the requested length l and returns a string of l bits. An ideal MGF is simply a monolithic random function having such an interface.

Definition 3 (MGF). *An MGF is a VIL-VOL function* F: $\{0, 1\}^* \times \{1\}^* \longrightarrow \{0, 1\}^*$ *satisfying the following two properties:*

1. Length: For all $x \in \{0, 1\}^* X$ *and* $l \in \mathbf{Z}_{\geq 0}$*, we have*

$$|F(x, 1^l)| = l.$$

2. Prefix: For all l; $l' \in \mathbf{Z}_{\geq 0}$ *such that* $l \leq l'$*, we have*

$$F(x, 1^l) = [F(x, 1^{l'})]^l.$$

The MGF random oracle $\mathcal{O}^{\mathrm{mgf}}$ *is a function chosen uniformly at random from the set of MGFs.*

The notion of MGFs corresponds to the original definition of VIL/VOL random oracles by Bellare and Rogaway [6]; an MGF specifies the length of outputs. The notion of MGFs is also compatible with the interface of the MGF1, which has been widely standardized [13], [17]–[19].

Below, we give one of several possible definitions of a hash function.

Definition 4 (Hash function). *A hash function is a VIL/FOL function* H: $\{0, 1\}^* \longrightarrow \{0, 1\}^n$*, where* $n \in \mathbf{N}$.

The random oracle $\mathcal{O}_n^{\mathrm{hash}}$ *is a function chosen uniformly at random from the set of hash functions with n-bit outputs.*

We say that efficient construction F of an MGF using random oracle $\phi = \mathcal{O}_n^{\mathrm{hash}}$ is secure if it is indifferentiable from MGF random oracle $\Phi = \mathcal{O}^{\mathrm{mgf}}$.

4. Analysis of counter-based MGFs

First, we define the counter-based MGF F. Then, we show that F cannot be proven to be indifferentiable from $\mathcal{O}^{\mathrm{mgf}}$ in the sense that there exists no unaffected simulator. On the other hand, we also show that F cannot be proven to be insecure in the sense that there exists no unaffected adversary.

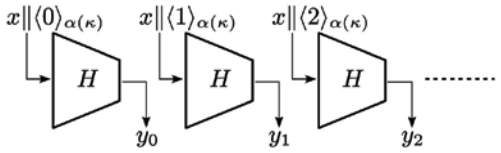


Fig. 1. Description of the counter-based MGF.

4.1 Description of the counter-based MGF

The counter-based MGF F: $\{0, 1\}^* \times \mathbf{N} \longrightarrow \{0, 1\}^*$ uses a hash function H: $\{0, 1\}^* \longrightarrow \{0, 1\}^n$. Here, the output length n is a polynomial function of the security parameter κ. The description of F is as follows:

1. Receive an input $(x; l) \in \{0, 1\}^* \times \mathbf{N}$.
2. Set $t = \lceil l/n \rceil$ and $r = l - n(t - 1)$.
3. Compute $y_i = \mathrm{H}(x \| \langle i \rangle_{\alpha(\kappa)})$ for $i = 0, \ldots, t - 1$.
4. Output $F(x, l) = y_0 \| \cdots \| y_{t-2} \| [y_{t-1}]^r$.

In the above, $\langle i \rangle_{\alpha(\kappa)}$ denotes an $\alpha(\kappa)$-bit representation of integer i, where $\alpha(\kappa)$ is a polynomial in κ. The counter-based MGF is illustrated in **Fig. 1**.

4.2 Proof that the counter-based MGF is locally indifferentiable

We prove that the counter-based MGF M defined above is *locally* indifferentiable. Intuitively, the proof goes as follows. Given an efficient distinguisher D, there exist polynomials $q_D(\kappa)$ and $l_D(\kappa)$. Using these polynomials, we can construct an efficient simulator S that sets the advantage of D to zero.

Theorem 1. *The counter-based MGF* M *is locally indifferentiable from an ideal MGF* M.

Proof. Let D be an efficient distinguisher attacking the counter-based MGF F. We show that there exists an efficient simulator S that makes the advantage function of D equal to 0.

Let $q_D(\kappa)$ and $l_D(\kappa)$ be polynomial functions restricting the capacity of D. Using these polynomials, we construct a simulator $S^{\Phi}(1^{\kappa}, \sigma, x)$ as follows.

1. Receive an $\mathcal{O}_n^{\mathrm{hash}}$-query $X \in \{0, 1\}^*$ from adversary D.
2. If (X, Y) is already in state σ, then return (σ, Y) to adversary D.
3. If $X = x \| \langle i \rangle_{\alpha(\kappa)}$ and $i \cdot n(\kappa) \leq l_D(\kappa)$, then compute $Y = [\Phi(x, (i+1) \cdot n(\kappa))]_{n(\kappa)}$ by asking Φ oracle and obtain updated state σ' by adding (X, Y) to σ and return (σ', Y) to adversary D.

4. If $X = x\|\langle i\rangle_{\alpha(\kappa)}$ and $i \cdot n(\kappa) > l_D(\kappa)$, then choose a random string $Y \in_U \{0, 1\}^{n(\kappa)}$ and obtain updated state σ' by adding (X, Y) to σ and return (σ', Y) to adversary D.

We see that S is an efficient simulator because we have $t_S = O(q_D l_D)$, $q_S = q_D$, and $l_S = l_D + n$. We also observe that S perfectly mimics the oracle $\mathcal{O}_n^{\mathrm{hash}}$ in a way consistent with the oracle $\mathcal{O}^{\mathrm{mgf}}$ (up to length l_D). Hence, we can set $\epsilon(\kappa) = 0$.

4.3 Proof that the counter-based MGF is not universally indifferentiable

Now we prove that the counter-based MGF F is not universally indifferentiable from an ideal MGF Φ. Intuitively, we argue that any single simulator S will fail when a distinguisher D starts by raising a query $x\|\langle i\rangle_{\alpha(\kappa)}$ for some huge i. In such a case, S is forced to decide whether or not to send a query $(x, n(i + 1)$ to its Φ oracle. If the value i is within the resource bounds of D, then S should certainly send such a query (and return the *consistent* value). On the other hand, if i is beyond the resource bounds of D, then S should simply ignore making such a query (and return a random string). However, S cannot make such a decision intelligently because it is not allowed to have any information about D.

Theorem 2. *The counter-based MGF F is not universally indifferentiable from an ideal MGF* Φ.

Proof. Suppose, on the contrary, that there exists a single simulator S that works against any efficient distinguisher. We show that this leads to a contradiction.

Let κ be the security parameter. Throughout the proof, we set the seed x to be a one-bit string "0," i.e., $x = 0$.

First, we define two types of events, Query_i and Reply_i, for $i \in \mathbf{Z}_{\geq 0}$. Every distinguisher that we construct in the current proof sends a query of the form $x\|\langle i\rangle_{\alpha(\kappa)}$, for some $i \in \mathbf{Z}_{\geq 0}$, to its H oracle at the beginning of each game execution. Let Query_i denote this event. After the event Query_i, the simulator S is given a pair $(\varepsilon, x\|\langle i\rangle_{\alpha(\kappa)})$ (ε being the null state) and is required to return updated state σ' and a string $y \in \{0, 1\}^{n(\kappa)}$. Let Reply_i denote this event, i.e., the event that $(\sigma', y) \leftarrow S^{\Phi}(1^{\kappa}, \varepsilon, x\|\langle i\rangle_{\alpha(\kappa)})$ is computed and returned to the intermediary I.

Next, we define probabilities $p_i(\kappa)$ for $i \in \mathbf{Z}_{\geq 0}$. Let Tune_i denote the event, which occurs between Query_i and Reply_i, of simulator S sending a query $(x, 1^l)$ to its Φ oracle for some $l \geq n(\kappa) \cdot i + 1$. Put $p_i(\kappa) = \Pr[\mathrm{Tune}_i]$. Since we fix the seed x, the probability $p_i(\kappa)$ is well-defined for each pair of an integer $i \in \mathbf{Z}_{\geq 0}$ and a security parameter $\kappa \in \mathbf{N}$. Note that the probability $p_i(\kappa)$ does not depend on the description of distinguishers and is defined over the coins of S and Φ (S may send some other queries to its Φ oracle in the interval).

We define a function $\mathbf{j}: \mathbf{N} \longrightarrow \mathbf{Z}_{\geq 0} \cup \{\infty\}$ as follows. For security parameter $\kappa \in \mathbf{N}$, let $\mathbf{j}(\kappa)$ be the smallest index i such that $p_i(\kappa) \leq 1/3$ (This fraction can be any constant strictly larger than 0 and strictly smaller than 1/2). If no such index exists, then we define $\mathbf{j}(\kappa)$ as the special symbol ∞, which is defined to be larger than any $i \in \mathbf{Z}_{\geq 0}$. Again, note that $\mathbf{j}$ is determined as soon as we fix the simulator S; the description of $\mathbf{j}$ is independent of distinguishers.

We show that $\mathbf{j}$ cannot be bounded by a polynomial function. Suppose, on the contrary, that there exists some polynomial function $f(\kappa)$ and an integer $N_1 \in \mathbf{N}$ such that for all security parameters $\kappa > N_1$ the inequality $\mathbf{j}(\kappa) < f(\kappa)$ holds. If such a polynomial function f exists, then it implies that there also exists a distinguisher $D_f^{\Phi, S}(1^{\kappa})$ as follows.

1. Choose a random index $i \in_U \{0, 1 \ldots, f(\kappa) - 1\}$,
2. Send a query $x\|\langle i\rangle_{\alpha(\kappa)}$ to its S oracle and receive a string $y \in \{0, 1\}^{n(\kappa)}$,
3. Send a query $(x, 1^{n(\kappa) \cdot (i + 1)})$ to its Φ oracle and receive a string $y \in \{0, 1\}^*$,
4. $y' \leftarrow [y]_{n(\kappa)}$,
5. If $y = y'$, return 1; otherwise, return 0.

Observe that the distinguisher D_f makes exactly two queries, each being at most $n(\kappa) \cdot f(\kappa)$ bits. Therefore, D_f is an efficient distinguisher.

We show that this is in direct contradiction to the requirement that the success probability of the distinguisher D_f be negligible. To see this, let us compute the advantage $\mathrm{Adv}^{\mathrm{mgf}}(D_f{}^{(1^{\kappa})})$ for sufficiently large security parameters $\kappa > N_1$. If D_f interacts with the pair $(F, \phi_{n(\kappa)})$, we can easily verify that D_f outputs 1 with probability 1. On the other hand, if D_f interacts with the pair (Φ, S), we claim that the probability of $D_f(1^{\kappa})$ returning 1 is at most $1 - f(\kappa)^{-1} \cdot (2/3 - 2^{-n(\kappa)})$. To see this, let One denote the event $D_f^{\Phi, S} = 1$ and Hit denote the event $i = \mathbf{j}(\kappa)$ in line 1 of the description of distinguisher $D_f^{\Phi, S}(1^{\kappa})$. We have

$$\begin{aligned}\Pr[\mathrm{One}] &= \Pr[\mathrm{Hit} \wedge \mathrm{One}] + \Pr[\overline{\mathrm{Hit}} \wedge \mathrm{One}] \\ &= \Pr[\mathrm{Hit}] \cdot \Pr[\mathrm{One} \mid \mathrm{Hit}] + \Pr[\overline{\mathrm{Hit}}] \cdot \Pr[\mathrm{One} \mid \overline{\mathrm{Hit}}] \\ &\leq \frac{1}{f(\kappa)} \cdot \Pr[\mathrm{One} \mid \mathrm{Hit}] + \frac{f(\kappa) - 1}{f(\kappa)} \cdot 1,\end{aligned}$$

and we also have

$$\begin{aligned}\Pr[\text{One} \mid \text{Hit}] &= \Pr[\text{Tune}_i \wedge \text{One} \mid \text{Hit}] + \Pr[\overline{\text{Tune}_i} \wedge \text{One} \mid \text{Hit}] \\ &= \Pr[\text{Tune}_i \mid \text{Hit}] \cdot \Pr[\text{One} \mid \text{Hit} \wedge \text{Tune}_i] \\ &\quad + \Pr[\overline{\text{Tune}_i} \mid \text{Hit}] \cdot \Pr[\text{One} \mid \text{Hit} \wedge \overline{\text{Tune}_i}] \\ &\leq \frac{1}{3} \cdot 1 + 1 \cdot \frac{1}{2^{n(\kappa)}},\end{aligned}$$

where the variable i denotes the value selected in line 1 of the description of the distinguisher $D_f^{\Phi, S}(1^\kappa)$. Hence, we get

$$\Pr\left[D_f^{\Phi, S}=1\right] \leq \frac{1}{f(\kappa)}\left(\frac{1}{3}+\frac{1}{2^{n(\kappa)}}\right)+\frac{f(\kappa)-1}{f(\kappa)}=1-\frac{1}{f(\kappa)}\left(\frac{2}{3}-\frac{1}{2^{n(\kappa)}}\right),$$

and we also get

$$\mathrm{Adv}_{F, S}^{\mathrm{mgf}}(D_f(1^\kappa)) \geq 1-1+\frac{1}{f(\kappa)}\left(\frac{2}{3}-\frac{1}{2^{n(\kappa)}}\right)=\frac{1}{f(\kappa)}\left(\frac{2}{3}-\frac{1}{2^{n(\kappa)}}\right) \geq \frac{1}{6f(\kappa)},$$

which is clearly not a negligible function.

Thus, we have shown that function $\mathbf{j}(\kappa)$ is not bounded by any polynomial function. We show that this also leads to a contradiction, creating another type of distinguisher.

To construct the distinguisher, we first identify a polynomial $g(\kappa)$ as follows. Consider an initial query $x\|\langle i\rangle_{\alpha(\kappa)}$. This leads to an input $(1^\kappa, \sigma, x\|\langle i\rangle_{\alpha(\kappa)})$ to the simulator S, where $\sigma = \varepsilon$ and $x = 0$. Hence, the length of such an input is $\kappa + 0 + 1 + \alpha(\kappa)$, which is a polynomial in κ. Since the bound l_S is a polynomial in the input length, we can regard l_S as a polynomial in κ, which we define as $g(\kappa) = l_S(\kappa + 1 + \alpha(\kappa))$. Now that we have identified a polynomial function $g(\kappa)$, we construct the distinguisher $D_g^{\Phi, S}(1^\kappa)$ as follows.

1. $i \leftarrow \min(g(\kappa) + 1, 2^{\alpha(\kappa)} - 1)$
2. Send a query $x\|\langle i\rangle_{\alpha(\kappa)}$ to its S oracle and discard whatever is received,
3. Return 1.

Next, we find a security parameter κ_1 for which running D_g with S leads to a contradiction. Observe that there exists some integer $N_0 \in \mathbf{N}$ such that for all $\kappa > N_0$, the inequality $g(\kappa) + 1 < 2^{\alpha(\kappa)} - 1$ holds because the left-hand side is a polynomial in κ whereas the right-hand side is an exponential function of κ. Now recall that $\mathbf{j}(\kappa)$ is not bounded by any polynomial function, which implies that there exists some integer $\kappa_1 > N_0$ such that $g(\kappa_1) + 1 < \mathbf{j}(\kappa_1)$. Here, it is important to note that we have $P_{g(\kappa_1)+1}(\kappa_1) > 1/3$ from the definition of $\mathbf{j}$.

Finally, by setting the security parameter κ to κ_1, we find a contradiction in the simulator's bound l_S when running D_g. The distinguisher D_g sends its S oracle a query $x\|\langle g(\kappa_1) + 1\rangle_{\alpha(\kappa_1)}$, which forces S with probability of more than 1/3 to send a query (x, l) to its Φ oracle for some $l \geq n(\kappa_1) \cdot (g(\kappa_1) + 1)$. Then, we have

$$g(\kappa_1) = l_S(\kappa_1 + 1 + \alpha(\kappa_1)) \geq l \geq n(\kappa_1) \cdot (g(\kappa_1) + 1),$$

which is a contradiction.

5. Analysis of chained MGFs

Our results for the counter-based MGF raise the question of whether there exists an MGF construction that can be proven to be universally indifferentiable from an ideal MGF. In this section, we present one such construction: the chained MGF. As an example of the chained MGF, we describe the Double-Pipeline Iteration Mode specified in NIST SP800-108 [20]. We then prove that the Double-Pipeline Iteration Mode is universally indifferentiable from an ideal MGF.

5.1 Description of the Double-Pipeline Iteration Mode

The Double-Pipeline Iteration Mode specified in NIST SP800-108 [20] F: $\{0, 1\}^* \times \mathrm{N} \longrightarrow \{0, 1\}^*$ uses a hash function H: $\{0, 1\}^* \longrightarrow \{0, 1\}^n$. Here, the output length $n = n(\kappa)$ is a polynomial function of the security parameter κ such that the inequality $n(\kappa) > \kappa$ holds for all $\kappa \in \mathbf{N}$. The description of F is as follows:

1. Receive an input $(x, l) \in \{0, 1\}^* \times \mathrm{N}$.
2. Set $t = \lceil l/n \rceil$, $r = l - n(t - 1)$ and $\upsilon_0 = x \in \{0, 1\}^*$.
3. Compute $\upsilon_i = H(\upsilon_{i-1})$ and $y_i = H(\upsilon_i \| \langle i\rangle_{\alpha(\kappa)} \| x)$ for $i = 1, \ldots, t$.
4. Output $F(x, l) = y_1 \| \cdots \| y_{t-1} \| [y_t]^r$.

In the above, $\langle i\rangle_{\alpha(\kappa)}$ denotes an $\alpha(\kappa)$-bit representation of integer i, where $\alpha(\kappa)$ is a polynomial in κ. The Double-Pipeline Iteration Mode is illustrated in **Fig. 2**.

Since the resource bound of the Double-Pipeline Iteration Mode F is $O(|x|l)$ and the output length is $O(l)$, the Double-Pipeline Iteration Mode F is efficient.

5.2 Proof that the Double-Pipeline Iteration Mode is universally indifferentiable

The Double-Pipeline Iteration Mode F is indifferentiable from MGF random oracle $\mathcal{O}^{\mathrm{mgf}}$.

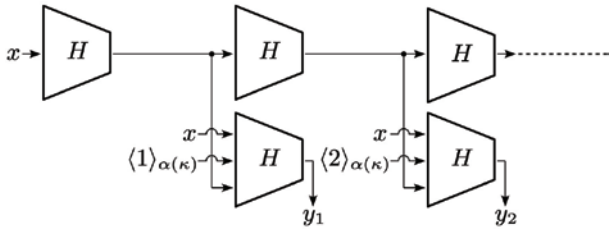


Fig. 2. Description of the Double-Pipeline Iteration Mode.

Theorem 3. *The Double-Pipeline Iteration Mode* $F^{\mathcal{O}_n^{\text{hash}}}$ *is universally indifferentiable from oracle* $\mathcal{O}^{\text{mgf}}$ *in the sense that there exists a universal simulator S having bounds* $t_S = O(q_D^2)$, $q_S = q_D$, $l_S = q_D n$, *and* $\epsilon = q_D/2^n$.

Proof. We construct a simulator S that has access to oracle $\mathcal{O}^{\text{mgf}}$ and that tries to simulate oracle $\mathcal{O}_n^{\text{hash}}$. S works as follows:

1. Receive an $\mathcal{O}_n^{\text{hash}}$-query $X \in \{0, 1\}^*$ from adversary D.
2. If the query $X \in \{0, 1\}^*$ is stored, return the stored answer to adversary D.
3. If $X = x$, return a random string $v_1 \in_U \{0, 1\}^n$ to D and store $(v_0 = x, v_1, 1, \text{“chained”})$.
4. If $X = v_i$ and $(v_{i-1}, v_i, i, \text{“chained”})$ is stored, return a random string $v_{i+1} \in_U \{0, 1\}^m$ to D and store $(v_i, v_{i+1}, i + 1, \text{“chained”})$.
5. If $X = v_i \,\|\langle i\rangle_{\alpha(\kappa)}\|x$ and $(v_{i-1}, v_i, i, \text{“chained”})$ is stored, return $y_i = [\mathcal{O}^{\text{mgf}}(x, i \cdot n)]_n \in \{0, 1\}^n$ to D and store $((v_i, i, x), y_i, i, \text{“chained”})$.
6. Otherwise, return a random string $Y \in_U \{0, 1\}^n$ to D and store $(X, Y, \text{“junk”})$.

Now we argue that simulator S is a polynomial-time adversary. To see this, let t_D, q_D, l_D be polynomial functions such that $D(\kappa) \in D(t_D, q_D, l_D)$. Since S makes $\mathcal{O}^{\text{mgf}}$-oracle queries only if distinguisher D makes a *chained* query, the number of queries sent by S to $\mathcal{O}^{\text{mgf}}$-oracle is at most q_D, and each query is of length at most $q_D n$ bits. Hence, we have $q_S = q_D$, $l_S = q_D n$. S needs to search at most q_D stored queries at most q_D times. Hence, we have $t_S = O(q_D^2)$.

S perfectly simulates oracle $\mathcal{O}_n^{\text{hash}}$ in a way consistent with the construction F except in the case that distinguisher D asks $X = v_i \,\|\langle i\rangle_{\alpha(\kappa)}\|x$ with the *correct* v_i before asking $X = v_{i-1}$. Hence, we have $\epsilon(\kappa) = q_D/2^n$.

However, the Double-Pipeline Iteration Mode outputs $n/2^n$ bits per hash function computation, so it is less efficient than a counter-based MGF, which outputs n bits per hash function computation.

6. Conclusion

We have shown that the counter-based MGF cannot be proven to be naturally indifferentiable from the ideal MGF. As a solution to this problem, we have shown that a chained MGF is proven to be indifferentiable from the ideal MGF. However, the chained MGF is less efficient than the counter-based MGF because it outputs fewer bits per invocation and operates in a non-parallelizable manner. It might be worth performing a more detailed study of this security/performance tradeoff because the current work opens up other possibilities for new MGF constructions that are indifferentiable from the ideal MGF and at the same time more efficient (or more secure) than the chained MGF.

References

[1] M. Bellare and P. Rogaway, “Random oracles are practical: A paradigm for designing efficient protocols,” Proc. of the 1st ACM Conference on Computer and Communications Security, pp. 62–73, Fairfax, VA, USA, 1993.

[2] M. Bellare and P. Rogaway, “Optimal asymmetric encryption,” In Alfredo De Santis, editor, EUROCRYPT 1994, Lecture Notes in Computer Science, Vol. 950, pp. 92–111, Heidelberg, 1995, Springer.

[3] M. Bellare and P. Rogaway, “The exact security of digital signatures—How to sign with RSA and Rabin,” In Ueli M. Maurer, editor, EUROCRYPT 1996, Lecture Notes in Computer Science, Vol. 1070, pp. 399–416, Heidelberg, 1996, Springer.

[4] J. -S. Coron, Y. Dodis, C. Malinaud, and P. Puniya, “Merkle-Damgård revisited: How to construct a hash function. In Victor Shoup, editor, CRYPTO 2005, Lecture Notes in Computer Science, Vol. 3621, pp. 430–448, Heidelberg, 2005, Springer.

[5] U. M. Maurer, R. Renner, and C. Holenstein, “Indifferentiability, impossibility results on reductions, and applications to the random oracle methodology,” In Moni Naor, editor, TCC 2004, Lecture Notes in Computer Science, Vol. 2951, pp. 21–39, Heidelberg, 2004, Springer.

[6] M. Bellare and T. Ristenpart, “Multi-property-preserving Hash Domain Extension and the EMD Transform,” In Xuejia Lai and Kefei Chen, editors, ASI-ACRYPT 2006, Lecture Notes in Computer Science, Vol. 4284, pp. 299–314, Heidelberg, 2006, Springer.

[7] D. Chang, S. Lee, M. Nandi, and M. Yung, “Indifferentiable Security Analysis of Popular Hash Functions with Prefix-free Padding,” In Xuejia Lai and Kefei Chen, editors, ASIACRYPT 2006, Lecture Notes in Computer Science, Vol. 4284, pp. 283–298, Heidelberg, 2006, Springer.

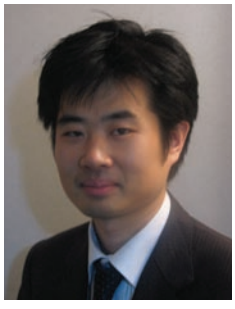
[8] M. Bellare and T. Ristenpart, “Hash functions in the dedicated-key setting: Design choices and MPP transforms,” In Lars Arge, Christian Cachin, Tomasz Jurdzinski, and Andrzej Tarlecki, editors, ICALP 2007, Lecture Notes in Computer Science, Vol. 4596, pp. 399–410, Heidelberg, 2007, Springer.

[9] S. Hirose, J. H. Park, and A. Yun, “A simple variant of the Merkle-Damgård scheme with a permutation,” In Kaoru Kurosawa, editor, ASIACRYPT 2007, Lecture Notes in Computer Science, Vol. 4833, pp. 113–129, Heidelberg, 2007, Springer.

[10] U. M. Maurer and S. Tessaro, “Domain extension of public random functions: Beyond the birthday barrier,” In Alfred Menezes, editor,

CRYPTO 2007, Lecture Notes in Computer Science, Vol. 4622, pp. 187–204, Heidelberg, 2007, Springer.
[11] D. Chang and M. Nandi, "Improved Indifferentiability Security Analysis of ChopMD Hash Function," In Kaisa Nyberg, editor, FSE 2008, Lecture Notes in Computer Science, Vol. 5086, pp. 429–443, Heidelberg, 2008, Springer.
[12] ISO/IEC, Geneva. ISO/IEC 9796-3 Information technology—Security techniques—Digital signature schemes giving message recovery—Part 3: Discrete logarithm based mechanisms, 2006.
[13] IEEE Computer Society, New York, "IEEE 1363.1 Standard Specifications For Public-Key Cryptography," 2000.
[14] L. A. Pintsov and S. A. Vanstone, "Postal revenue collection in the digital age," In Yair Frankel, editor, Financial Cryptography 2000, Lecture Notes in Computer Science, Vol. 1962, pp. 105–120, Heidelberg, 2001, Springer.
[15] M. Abe and T. Okamoto, "A Signature Scheme with Message Recovery as Secure as Discrete Logarithm," In Kwok-Yan Lam, Eiji Okamoto, and Chaoping Xing, editors, ASIACRYPT 1999, Lecture Notes in Computer Science, Vol. 1716, pp. 378–389, Heidelberg, 1999, Springer.
[16] M. Abe, T. Okamoto, and K. Suzuki, "Message Recovery Signature Schemes from Sigma-protocols," NTT Technical Review, Vol. 6, No. 1, 2008.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200801sp2.html
[17] ANSI, New York, "ANSI X9.44 Draft D2," 2002.
[18] RSA Security, Bedford, "PKCS#1 v2.1," 2002.
[19] ISO/IEC, Geneva, "ISO/IEC 18033-2 Information technology—Security techniques—Encryption algorithms—Part 2: Asymmetric ciphers," 2006.
[20] NIST, "NIST SP800-108," 2009.
[21] H. Krawczyk, "Cryptographic Extraction and Key Derivation: The HKDF Scheme," CRYPTO 2010, pp. 631-648, Tal Rabin, editor, Lecture Notes in Computer Science Vol. 6223, Heidelberg, 2010, Springer.

Koutarou Suzuki
Senior Research Scientist, Information Security Project, NTT Secure Platform Laboratories.
He received the B.S., M.S., and Ph.D. degrees from the University of Tokyo in 1994, 1996, and 1999, respectively. He joined NTT in 1999. He has been engaged in research on public key cryptography, especially on cryptographic protocols and digital signatures.



Kan Yasuda
Senior Research Scientist, Information Security Project, NTT Secure Platform Laboratories.
He received the Ph.D. degree in mathematical sciences from the University of Tokyo in 2003. He has been working for NTT since 2004.

Revision of the Common Patent Guidelines for ITU/ISO/IEC

Isamu Yoshimatsu

Abstract

The "Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC" were revised by a working group in ITU and published through the web pages of the organizations after their approval (ITU-T: International Telecommunication Union, Telecommunication Standardization Sector; ITU-R: International Telecommunication Union, Radiocommunication Sector; ISO: International Organization for Standardization; IEC: International Electrotechnical Commission). The revised guidelines clarified the following two points.

(1) The submitted patent declaration form shall be irrevocable and only superseded by another form containing a more favorable licensing commitment from an implementer's perspective.

(2) The owner of the declared patents shall make reasonable efforts to notify an assignee or transferee of the existence of such a license undertaking if the owner assigns or transfers them. In addition, if the owner specifically identified patents in the declaration, the owner shall have the assignee or transferee agree to be bound by the same licensing commitment.

1. Introduction

A patent is a right granted by a government that confers upon the creator of an invention the sole right to make, use, and sell that invention for a set period of time [1]. The right's owner may assert this right against anyone exercising the technologies without the owner's license. Therefore, the concept of patent right confronts that of standards, which mean technologies published and available to anyone since if a standard includes patents that are essential for its implementation and the patent owner asserts his or her rights, the standard will not be publically available.

To avoid this contradictory situation, if the owner declares in writing (called a patent declaration) to license essential patents for implementing a Standard according to the procedures in the "Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC", then such a standard will be approved and published formally by a standards developing organization (SDO) such as ITU-T (International Telecommunication Union, Telecommunication Standardization Sector), ITU-R (International Telecommunication Union, Radiocommunication Sector), ISO (International Organization for Standardization), or IEC (International Electrotechnical Commission).

2. Basic concept of patent declaration

The form of patent declaration specified in the Guidelines shall be made by the owner of the patents and submitted to the bureau of the SDO pursuant to the procedures in the Guidelines. The contents in the form are as follows.

(1) The name of the owner and the address of the department to contact for a license

(2) The name and the formal number of the specification of a Standard to which the declared patents refers.

(3) The license policy of the patents. It shall be selected as one among the three options below:

(i) Granting a free-of-charge license on a non-discriminatory basis and under reasonable terms: Option 1.

(ii) Granting a license on a non-discriminatory basis and under reasonable terms: Option 2.

(iii) Unwilling to grant a license in accordance with the provisions of either (i) or (ii) above: Option 3.

(4) Information about the patents such as application number, the titles of the patents, etc.

The revised Guidelines make clear that the owner might classify different claims of the patents as different options on the declaration form. All of the items of information regarding the patent declaration above may be found on the web pages of the SDOs concerned [2].

3. Submission of the patent declaration form

The Chairman of the working group in which draft specifications of a Standard are developed will ask, if appropriate, whether anyone has knowledge of essential patents and will request anyone who believes that they hold essential patents to submit the patent declaration form. If a form with Option 3 selected is submitted, the draft will be changed in order not to include such patents; otherwise, it would be necessary to give up development activities for making the Standard.

Someone who finds or knows of essential patents owned by another person or persons may also submit the form after filling in the information for them. If a member of the working group finds someone who seems to have essential patents, he or she reports the name and address of the patent owner to the SDO's bureau. The bureau requests the owner to submit the patent declaration if the owner recognizes that the patents are essential for implementing the Standard.

If the owner of essential patents submits a patent declaration form with Option 3 selected, the owner might in future file a patent infringement lawsuit against the implementer of the Standard. Therefore, it is necessary to continuously watch whether or not Option 3 is selected in patent declarations. It is also necessary to investigate carefully whether patents are essential or not since the decision for essentiality depends on the patent owner's judgment.

It is dangerous to believe that there are no problems concerning patents related to a Standard merely because the Standard's specifications have already been approved and published by an SDO. We need to recognize that a Standard may become unimplementable in order to avoid patent infringement even after it has already been published and spread throughout the world.

4. Retraction or resubmission of a patent declaration

In the revised Guidelines, a submitted patent declaration form is irrevocable and only superseded by another form containing a more favorable licensing commitment from an implementer's perspective such as follows.

(1) a change in commitment from Option 2 to Option 1 or

(2) a change in commitment from Option 3 to either Option 1 or Option 2

The revised Guidelines also make clear that a form containing no patent information shall mean that the licensing policy applies to any essential patents even if they are found after the form's submission. If the form's submitter does not provide any patent information, such as the patent's application number, in the form in order to submit it as soon as possible, the submitter must consider carefully whether the submitter can accept that the licensing commitment in the form shall apply to any essential patents that might be granted in the future.

5. Assignment or transfer of declared patents

If the essential patents declared in the form are assigned or transferred, whether or not the license commitment is also assigned or transferred, the revised Guidelines make clear that the owner shall make reasonable efforts to notify such an assignee or transferee of the existence of such a license undertaking when the owner assigns or transfers its declared patents. In addition, if the owner specifically identified patents, the owner shall have the assignee or transferee agree to be bound by the same licensing commitment as the owner for the same patent.

The patent declaration shall not be understood as a legal agreement. Therefore, if the assignee or transferee refuses the license offer from implementers or asserts that they are patent infringers with disregard to the patent declaration of the former owner, such an assignee or transferee is not assumed to be violating a licensing commitment in the patent declaration form.

However, in a courtroom trial, such an assertion may be assumed to breach the trust of implementers who believed that they would be able to obtain a patent license according to the patent declaration. In one such case, a favorable judgment was passed on the defendant who implemented the Standard [3]. Once the revised Guidelines are adopted, they may lead to

favorable judgments for defendants since assignee or transferees should be bound by the same licensing commitment according to the procedures in the Guidelines.

6. Conclusion

Standards and patents have different purposes, but following the procedures in the Guidelines, including patent declaration form submission, should enable one to avoid being sued for patent infringement. Following the revision of the Guidelines, I expect that there will be requests to make clearer the definition of reasonable terms, give specific examples, and also make clearer that the application of injunction relief shall be restricted to patents declared as essential patents. I hope to continue to be involved in efforts to revise the Guidelines from the viewpoint of both patent holders who are licensors and implementers who are licensees in order that all terms and conditions for a license can be agreed easily.

References

[1] The Free Dictionary. http://www.thefreedictionary.com/patent
[2] See for example, http://www.itu.int/ipr/IPRSearch.aspx?iprtype=PS for ITU.
[3] See for example in U.S. Federal Trade Commission, In the Matter of Negotiated Data Solutions LLC., File No. 051-0094 (Sep. 23, 2008).

Isamu Yoshimatsu
Senior Manager, Licensing Group, NTT Intellectual Property Center.

He received the B.E. and M.E. degrees in advanced organic chemistry from Kyushu University, Fukuoka, in 1984 and 1986, respectively. He joined NTT Electrical Communication Laboratories, Tokyo, in 1986 and studied lithium rechargeable batteries until 1993. He moved to the Licensing Group of the Intellectual Property Center in 1993. He has been a member of the Intellectual Property Rights Committee in the Telecommunication Technology Committee since 2002.

Report on NTT Communication Science Laboratories Open House 2012

Kaname Kasahara, Seiichiro Tani, Keisuke Kinoshita, Ryoko Mugitani, and Takashi Hattori

Abstract

Open House 2012 was held at NTT Communication Science Laboratories in Keihanna Science City, Kyoto. Over 1000 people visited the facility on June 7 and 8 to enjoy six lectures and twenty-six exhibits of the labs' latest research efforts.

1. Overview

At NTT Communication Science Laboratories (NTT CS Labs), we are studying aspects of both human science and information science in order to devise communication technologies based on a deeper understanding of humans and information. In studies relating to future communication environments, intelligent computing, and the quality of human life, all of our researchers are continuously promoting the creation of innovative technologies that will revolutionize telecommunications. The labs are located in Kansai Science City (Seika-cho, Kyoto) and Atsugi City, Kanagawa.

NTT Communication Science Laboratories Open House has been held annually with the aim of introducing the results of the labs' basic research and innovative leading-edge research to not only NTT Group employees but also visitors from companies, universities, and research facilities engaged in research and development, business, and education. This year, the event was held at NTT Keihanna Building (**Photo 1**) in Kansai on the afternoon of June 7 and all day on June 8, 2012; there were 1070 visitors. This article reports on the event's research talks and exhibits.

Photo 1. Event site (NTT Keihanna Building).

2. Keynote speech

The open house started with a speech by the Director of NTT CS Labs, Naonori Ueda, entitled, "Communication science for big data era" (**Photo 2**).

In recent years, social media have been widely used, and information terminals have become highly compact and more advanced. These developments have accelerated the *information explosion* on the Internet more than ever. We are approaching the *big data era*, when telecommunication systems and services will deeply analyze a wide variety of data and provide the analysis results to systems in the real world, which will in turn lead to highly efficient

Photo 2. Naonori Ueda, Director of CS Labs, giving the keynote speech.

Photo 3. Research talk by Dr. Naoyuki Hironaka.

social systems.

To create such systems and services, telecommunication technologies should be not only highly sophisticated but also safe and secure for their users, which will lead to an enriched and relaxed information and communications technology (ICT) society. Therefore, it is necessary to study communication science, which includes not only telecommunications but also human science and social science. Accordingly, Dr. Ueda introduced the social trends of big data, involving machine-learning technologies that are expected to analyze data powerfully, and the related research conducted at NTT CS Labs.

3. Research talks

Four research talks were given, highlighting recent significant research results and high-profile research themes.

- "Neuroscience of liking and wanting—Exploring biological foundations of human emotion in animal behavior—" Naoyuki Hironaka, Human and Information Science Laboratory
- "Preservation of digital contents—Standardization activities and best practices for digital content preservation being discussed at ISO/IEC and other organizations—" Noboru Harada, Moriya Research Laboratory
- "Random number generation from light—Fast physical random number generation using chaos in semiconductor lasers—" Kazuyuki Yoshimura, Media Information Science Laboratory
- "Real world revealed through sensor networks—Technologies for collecting, interpreting, and presenting information from massive and heterogeneous sensor nodes—" Yoshiyuki Suyama, Innovative Communication Laboratory

Each presentation introduced some of the latest research results, including some background and an overview of the research. All of the talks were very well received by the many participants.

In "Neuroscience of liking and wanting," state-of-the-art studies on the emotion of liking and wanting, based on the development of neuroscience in recent years, were introduced (**Photo 3**). In "Real world revealed through sensor networks," the latest research developments derived from activities of the s-room Project were introduced from the viewpoint of collecting, interpreting, and presenting data from sensors (**Photo 4**).

4. Research exhibits

The open house featured 26 exhibits displaying the latest research results. These were classified into three categories—information science, interface science, and life science—from the viewpoint of how the results will influence our society of the future. Each exhibit had a booth and used techniques such as slides on a large-screen monitor or hands-on demonstrations, with researchers explaining the latest results directly to visitors (**Photo 5**). The research exhibits are summarized below.

Photo 4. Research talk by Dr. Yoshiyuki Suyama.

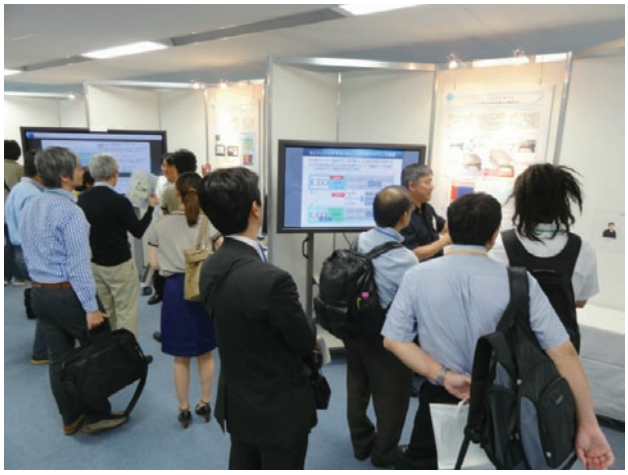


Photo 5. Research exhibits.

Information science

- Unmeasurable light produces common secrets
  —Secret key distribution using correlated randomness in lasers—
- Computer guarantees ultimate security of cryptosystems
  —Security proof using formal methods—
- Does the law protect your privacy?
  —Mathematical formulation of privacy and its applications to law—
- Adaptive learning from similar examples
  —Robust semi-supervised learning and its application to NLP—
- This is the essence of your data, isn't it?
  —Extracting hidden structure of data for deeper data mining—
- Quickly finding similar objects to a query
  —Fast similarity search based on a neighborhood-graph index—
- Live TV search
  —Realtime media search using incremental feature database—
- Observing the noisy world
  —Collective sensing, coding, and large deviation properties—
- Stereoscopic camera system for accurate color and shape reproduction
  —Stereoscopic 6-band video system—

Interface science

- Massive sensor networks collect data you want to know
  —Dynamic configuration of sensor networks—
- Word order is critical for translation quality
  —English <-> Japanese translation by Japanized English—
- Dynamic displays provide more information about conversations
  —Conversation space by physical representation of head motions—
- Toward media spaces that reminisce
  —Reflecting on past activities on t-Room and its applications—
- Analyzing your singing style
  —Singing style extraction based on singing voice F0 model—
- Who spoke when and what?
  —Progress in scene analysis for multi-speaker conversation—
- Clearly distinguishing your voice from ambient noise
  —Speech enhancement using temporal, spatial, and spectral cues—
- Listening and understanding conversations
  —Advanced techniques for spontaneous speech recognition—
- When sound alters vision
  —Distortion of visual space and time by audio-visual integration—
- Reaching now, looking later
  —Implicit eye-hand coordination—
- Seeing materials from image cues
  —Adaptive strategy of human visual system—

Life science

- Why do children suddenly begin to learn words?

Photo 6. Visitor experiencing a video conference with dynamically moving displays.

Photo 7. Visitors watching research presentation on infant's vocabulary spurt onset.

—Unveiling the myth of vocabulary spurt by analyzing longitudinal data—

- Appropriate words for children at particular ages
  —Searching for appropriate contents for toddlers—
- Speaking plays tricks on hearing
  —Close link between articulation and speech perception—
- Rats can better themselves by observing others
  —Neural basis of adaptive social behaviors—
- Exploring brain mechanisms for selective listening
  —Psychophysics, modeling, and functional brain measurements—
- Tactile sensation categories based on mimetic words
  —Tactile textures and their phonetic representations—

The exhibit "Dynamic displays provide more information about conversations" demonstrated a remote video conference where the head movement of each member at the remote location was reconstructed as the movement of a display located at the person's corresponding seat in the main location; the projected video of the remote person's head and shoulders gave many visitors the vivid feel of a real communication environment, as if all the conference participants were in the main location (**Photo 6**). The exhibit entitled "Why do children suddenly begin to learn words?" showed the underlying mechanism of an infant's vocabulary spurt onset, the phenomenon in which the child's vocabulary-learning speed suddenly become faster at an age of 1.5 years; the study was based on an analysis of personal longitudinal data, and many visitors showed great interest in this mechanism (**Photo 7**).

Photo 8. Associate Professor Akihiro Kitada of the University of Tokyo giving an invited talk.

5. Invited talk

This year's event also featured an invited talk by Akihiro Kitada, Associate Professor of the University of Tokyo. He spoke on the topic of "Connection, the social, and the political: The mode of communication

and the sociality of the youth" (**Photo 8**). From the viewpoint of social science, he gave an explanation of analyzing the mode of young people's communication while introducing movies about youth. Although it was a sophisticated presentation, the audience laughed and enjoyed the talk in a friendly mood.

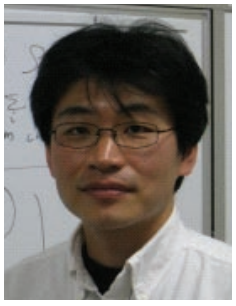
6. Concluding remarks

Just as last year, many visitors came to the NTT CS Labs open house and engaged in lively discussions on the research talks and exhibits and provided many valuable opinions about the presented results. In closing, we would like to offer our sincere thanks to all of the visitors and participants who attended this event.

References

[1] Open House website (in Japanese). http://www.kecl.ntt.co.jp/openhouse/2012/
[2] Open House poster. http://www.kecl.ntt.co.jp/openhouse/2012/oh2012_poster_en.pdf

Kaname Kasahara
Senior Research Scientist, Linguistic Intelligence Research Group, NTT Communication Science Laboratories.
He received the B.S. degree in chemistry from Keio University in 1989, the M.S. degree in physics from Tokyo Institute of Technology in 1991, and the Ph.D. degree in media and governance from Keio University in 2004. He joined NTT in 1991, studying word meaning in artificial intelligence and natural language processing. He was also a visiting researcher at the Center for the Study of Language and Information, Stanford University, USA, from 1998 to 1999.



Seiichiro Tani
Senior Research Scientist, Computing Theory Research Group, NTT Communication Science Laboratories.
He received the B.E. degree in information science from Kyoto University in 1993 and the M.S. and Ph.D. degrees in computer science from the University of Tokyo in 1995 and 2006, respectively. He joined NTT in 1995. He was a researcher at the Quantum Computation and Information Project, ERATO/SORST, Japan Science and Technology Agency (JST) from 2004 to 2009, and a visiting researcher at the Institute for Quantum Computing, University of Waterloo, Canada, from 2010 to 2011. His current research interests include distributed computing and complexity theory in the classical as well as quantum settings.

Keisuke Kinoshita
Researcher, NTT Communication Science Laboratories.
He received the M.E. and Ph.D. degrees from Sophia University, Tokyo, in 2003 and 2010, respectively. He is currently engaged in research on speech and audio signal processing. He received the 2006 IEICE Paper Award, the 2009 ASJ Outstanding Technical Development Prize, and the 2011 ASJ Awaya Prize. He is a member of IEEE, the Acoustical Society of Japan (ASJ), and the Institute of Electronics, Information and Communication Engineers (IEICE).

Ryoko Mugitani
Senior Research Scientist, Sensory Resonance Research Group, Human Information Science Laboratory, NTT Communication Science Laboratories.
She received the B.A. degree in education in 1999, the M.A. degree in health science in 2001, and the Ph.D. degree in arts and sciences in 2004 from the University of Tokyo. She joined NTT CS Labs in 2004. She is currently engaged in research on speech perception/production development in infants and children.

Takashi Hattori
Research Scientist, Learning and Intelligent Systems Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.
He received the B.E. degree in mechanical engineering and the M.S. degree in informatics from Kyoto University in 2002 and 2004, respectively. He joined NTT CS Labs in 2004. He is currently interested in fast approximate similarity search for large-scale and various media.

External Awards

Achievement Award

Winners: Atsushi Fukuda[†1], Hiroshi Okazaki[†2], and Shoichi Narahashi[†2]

†1 Radio Access Network Development Department, NTT DOCOMO, INC.

†2 Research Laboratories, NTT DOCOMO, INC.

Date: May 26, 2012

Organization: IEICE

For "A leading study on the multi-band operation of power amplifiers for mobile terminals".

http://www.ieice.org/eng/awards/gyouseki_01e.html

Achievement Award

Winners: Seizo Onoe[†1], Toshio Miki[†2], and Hiroshi Nakamura[†3]

†1 R&D Strategy Department, NTT DOCOMO, INC.

†2 Product Department, NTT DOCOMO, INC.

†3 Core Network Development Department, NTT DOCOMO, INC.

Date: May 26, 2012

Organization: IEICE

For "Implementation of LTE".

http://www.ieice.org/eng/awards/gyouseki_04e.html

Best Paper Award

Winners: Yasushi Ikei[†1], Koji Abe[†1], Koichi Hirota[†2], and Tomohiro Amemiya[†3]

†1 Tokyo Metropolitan University

†2 The University of Tokyo

†3 NTT Communication Science Laboratories

Date: Sept. 16, 2012

Organization: 18th International Conference on Virtual Systems and Multimedia (VSMM 2012)

For "A Multisensory VR System Exploring the Ultra-Reality".

Published as: Y. Ikei, K. Abe, K. Hirota, and T. Amemiya, "A Multisensory VR System Exploring the Ultra-Reality," Proc. of the 18th International Conference on Virtual Systems and Multimedia (VSMM 2012), Milan, Italy, 2012.

Papers Published in Technical Journals and Conference Proceedings

Conductive Polymer Combined Silk Fiber Bundle for Bioelectrical Signal Recording

S. Tsukada, H. Nakashima, and K. Torimitsu

PloS ONE, Public Library of Science, Vol. 7, No. 4, p. e33689, 2012.

Electrode materials for recording biomedical signals, such as electrocardiography (ECG), electroencephalography (EEG) and evoked potentials data, are expected to be soft, hydrophilic and electroconductive to minimize the stress imposed on living tissue, especially during long-term monitoring. We have developed and characterized string-shaped electrodes made from conductive polymer with silk fiber bundles (thread), which offer a new biocompatible stress-free interface with living tissue in both wet and dry conditions.

An electroconductive polyelectrolyte, poly(3,4-ethylenedioxythiophene)-poly(styrenesulfonate) (PEDOT-PSS) was electrochemically combined with silk thread made from natural *Bombyx mori*. The polymer composite 280 μm thread exhibited a conductivity of 0.00117 S/cm (which corresponds to a DC resistance of 2.62 MΩ/cm). The addition of glycerol to the PEDOT-PSS silk thread improved the conductivity to 0.102 S/cm (20.6 kΩ/cm). The wettability of PEDOT-PSS was controlled with glycerol, which improved its durability in water and washing cycles. The glycerol-treated PEDOT-PSS silk thread showed a tensile strength of 1000 cN in both wet and dry states. Without using any electrolytes, pastes or solutions, the thread directly collects electrical signals from living tissue and transmits them through metal cables. ECG, EEG, and sensory evoked potential (SEP) signals were recorded from experimental animals by using this thread placed on the skin. PEDOT-PSS silk glycerol composite thread offers a new class of biocompatible electrodes in the field of biomedical and health promotion that does not induce stress in the subjects.

Vapor Phase Polymerization of EDOT from Submicrometer Scale Oxidant Patterned by Dip-pen Nanolithography

C. D. O'Connell, M. J. Higgins, H. Nakashima, S. E. Moulton, and G. G. Wallace

Langmuir, American Chemical Society, Vol. 28, No. 1, pp. 9953–9960, 2012.

Some of the most exciting recent advances in conducting polymer synthesis have centered around the method of vapor phase polymerization (VPP) of thin films. However, it is not known whether the VPP process can proceed using significantly reduced volumes of oxidant and therefore be implemented as part of a nanolithography approach. Here, we present a strategy for submicrometer-scale patterning of the conducting polymer poly(3,4-ethylenedioxythiophene) (PEDOT) via in situ VPP. Attolitre (10^{-18} L) volumes of oxidant "ink"

are controllably deposited using dip-pen nanolithography (DPN). DPN patterning of the oxidant ink is facilitated by the incorporation of an amphiphilic block copolymer thickener, an additive that also assists with stabilization of the oxidant. When exposed to EDOT monomer in a VPP chamber, each deposited feature localizes the synthesis of conducting PEDOT structures of several micrometers down to 250 nm in width. PEDOT patterns are characterized by atomic force microscopy (AFM), conductive AFM, two-probe electrical measurement, and micro-Raman spectroscopy, evidencing in situ vapor phase synthesis of conducting polymer at a scale (picogram) which is much smaller than that previously reported. Although the process of VPP on this scale was achieved, we highlight some of the challenges that need to be overcome to make this approach feasible in an applied setting.

Security Enhancements by OR-Proof in Identity-Based Identification

A. Fujioka, T. Saito, and K. Xagawa

Applied Cryptography and Network Security, Lecture Notes in Computer Science, Vol. 7341, pp. 135–152, 2012.

We investigate three security enhancement transformations, based on the well-known OR-proof technique, in identity-based identification (IBI) protocols and show a required condition of the underlying IBI protocols. The transformations can convert an IBI protocol, which satisfies a property similar to the Σ-protocol and is secure against impersonation under passive attacks, into one secure against impersonation under concurrent attacks in both adaptive and weak selective identity attack models. In addition, we argue that enhancing the security in the static identity attack model with two of the transformations seems to be difficult; however, we prove that the third one can convert an IBI protocol, which satisfies another property, in the model.

Optimal entanglement manipulation via coherent-state transmission

K. Azuma and G. Kato

Phys. Rev. A, Vol. 85, No. 6, 060303(R), 2012.

We derive an optimal bound for arbitrary entanglement manipulation based on the transmission of a pulse in coherent states over a lossy channel followed by local operations and unlimited classical communication (LOCC). This stands on a theorem to reduce LOCC via a local unital qubit channel to local filtering. We also present an optimal protocol based on beam splitters and a quantum nondemolition (QND) measurement on photons. Even if we replace the QND measurement with photon detectors, the protocol can achieve near-optimal performance, outperforming known entanglement generation schemes.

Sufficient Condition for Ephemeral Key-Leakage Resilient Tripartite Key Exchange

A. Fujioka, M. Manulis, K. Suzuki, and B. Ustaoğlu

Proc. of the 17th Australasian Conference, ACISP 2012, Wollongong, NSW, Australia, 2012.

Tripartite Key Exchange (3KE) represents today the only known class of group key exchange protocols in which computation of unauthenticated session keys requires only one round and proceeds with minimal computation and communication overhead. The first one-round authenticated 3KE version that preserved the unique efficiency properties of the original protocol and strengthened its security towards resilience against leakage of ephemeral secrets was proposed recently by Manulis, Suzuki, and Ustaoğlu.

In this work we explore sufficient conditions for building such protocols. We define a set of *admissible polynomials* and show how their construction generically implies 3KE protocols with the desired security and efficiency properties. Our result generalizes the previous 3KE protocol and gives rise to many new authenticated constructions, all of which enjoy forward secrecy and resilience to ephemeral key-leakage under the Gap Bilinear Diffie-Hellman assumption in the random oracle model.

Grammar Error Correction Using Pseudo-Error Sentences and Domain Adaptation

K. Imamura, K. Saito, K. Sadamitsu, and H. Nishikawa

Proc. of the 50th Annual Meeting of the Association for Computational Linguistics, pp. 388–392, Jeju, Korea, 2012.

This paper presents grammar error correction for Japanese particles that uses discriminative sequence conversion, which corrects erroneous particles by substitution, insertion, and deletion. The error correction task is hindered by the difficulty of collecting large error corpora. We tackle this problem by using pseudoerror sentences generated automatically. Furthermore, we apply domain adaptation, the pseudo-error sentences are from the source domain, and the real-error sentences are from the target domain. Experiments show that stable improvement is achieved by using domain adaptation.

Meta-envy-free Cake-cutting and Pie-cutting Protocols

Y. Manabe and T. Okamoto

Journal of Information Processing, Vol. 20, No. 3, pp. 686–693, 2012.

This paper discusses cake-cutting protocols when the cake is a heterogeneous good, represented by an interval on the real line. We propose a new desirable property, the meta-envy-freeness of cake-cutting, which has not been formally considered before. Meta-envy-free means there is no envy in role assignments; that is, no party wants to exchange his/her role in the protocol with that of any other party. If there is envy in role assignments, the protocol cannot actually be executed because there is no settlement of which party plays which role in the protocol. A similar definition, envy-freeness, is widely discussed. Envy-free means that no player wants to exchange his/her part of the cake with that of any other player. Though envy-freeness was considered to be one of the most important desirable properties, it does not prevent envy about role assignment in the protocols. We define meta-envy-freeness to formalize this kind of envy. We propose that simultaneously achieving meta-envy-freeness and envy-freeness is desirable in cake-cutting. We show that current envy-free cake-cutting protocols do not satisfy meta-envy-freeness. Formerly proposed properties such as strong envy-free, exact, and equitable do not directly consider this type of envy and these properties are very difficult to realize. This paper then shows cake-cutting protocols for two- and three-party cases that simultaneously achieve envy-freeness and meta-envy-freeness. Finally, we show meta-envy-free pie-cutting protocols.

Low-complexity PARCOR coder designed for entropy coding of prediction residuals

Y. Kamamoto, T. Moriya, and N. Harada

Acoust. Sci. & Tech. Vol. 33, No. 4, 2012.

The low-complexity PARCOR quantization method, which is used for ITU-T G.711.0 (lossless compression of G.711), is described.

Laser Sharing between Transmitter and Receiver in Optical FDM-PON Access System Based on Optical Heterodyne Detection

S. Narikawa and N. Sakurai

IEICE Trans. on Communications, Vol. J95-B, No. 7, pp. 800–808, 2012 (in Japanese).

A lot of recent research has been devoted to the wavelength division multiplexing passive optical network (WDM-PON). We studied and evaluated a coherent FDM-PON access system which uses optical heterodyne detection for the receivers. Optical heterodyne detection can increase the PON branch number and enhance the transmission distance; however, it requires multiple lasers in the optical network unit (ONU), so a cost-effective ONU architecture is needed. To resolve this issue, we propose sharing the directly modulated laser with upstream signals for transmitter and receiver. We experimentally evaluated its effect.

Separability and Commonality of Auditory and Visual Bistable Perception

H. M. Kondo, N. Kitagawa, M. S. Kitamura, A. Koizumi, M. Nomura, and M. Kashino

Cerebral Cortex, Oxford University Press, Vol. 22, No. 8, pp. 1915–1922, 2012.

It is unclear what neural processes induce individual differences in perceptual organization in different modalities. To examine this issue, the present study used different forms of bistable perception: auditory streaming, verbal transformations, visual plaids, and reversible figures. We performed factor analyses on the number of perceptual switches in the tasks. A 3-factor model provided a better fit to the data than the other possible models. These factors, namely the "auditory", "shape", and "motion" factors, were separable but correlated with each other. We compared the number of perceptual switches among genotype groups to identify the effects of neurotransmitter functions on the factors. We focused on polymorphisms of catechol-O-methyltransferase (COMT) Val(158)Met and serotonin 2A receptor (HTR2A)-1438G/A genes, which are involved in the modulation of dopamine and serotonin, respectively. The number of perceptual switches in auditory streaming and verbal transformations differed among COMT genotype groups, whereas that in reversible figures differed among HTR2A genotype groups. The results indicate that the auditory and shape factors reflect the functions of the dopamine and serotonin systems, respectively. Our findings suggest that the formation and selection of percepts involve neural processes in cortical and subcortical areas.

Joint estimation of confidence and error causes in speech recognition

A. Ogawa and A. Nakamura

Speech Commun., Elsevier, Vol. 54, No. 9, pp. 1014–1028, 2012.

Speech recognition errors are essentially unavoidable under the severe conditions of real fields, and so confidence estimation, which scores the reliability of a recognition result, plays a critical role in the development of speech-recognition-based real-field application systems. However, if we are to develop an application system that provides a high-quality service, in addition to achieving accurate confidence estimation, we also need to extract and exploit further supplementary information from a speech recognition engine. As a first step in this direction, in this paper, we propose a method for estimating the confidence of a recognition result while jointly detecting the causes of recognition errors based on estimating the confidence of a recognition result, while jointly detecting the causes of recognition errors, by using a discriminative model. The confidence of a recognition result and the nonexistence/existence of error causes are naturally correlated. By directly capturing these correlations between the confidence and error causes, the proposed method enhances its estimation performance for the confidence and each error cause complementarily. In the initial speech recognition experiments, the proposed method provided higher confidence estimation accuracy than a discriminative model based a state-of-the-art confidence estimation method. Moreover, the effective estimation mechanism of the proposed method was confirmed by detailed analyses.