DCNL: Disclosure Control of Natural Language Information to Enable Secure and Enjoyable E-Communications

Haruno Kataoka[1], Natsuki Watanabe[2], Keiko Mizutani[2], and Hiroshi Yoshiura[2]

*[1] NTT Information Sharing Platform Laboratories, NTT Corporation,
3-9-11, Midori-cho Musashino-Shi, Tokyo 180-8585 Japan
[2] Graduate school of Electro-Communications, University of Electro-Communications,
1-5-1 Chofugaoka, Chofu, Tokyo, 182-8585 Japan
yoshiura@hc.uec.ac.jp*

Abstract

Natural language communications using social networking and blogging services can result in the undesired revelation of private information. Existing disclosure control is tedious and error-prone because the user must set the disclosure level manually and must reconsider the level every time a new text is to be uploaded. This can lead to the revelation of private information or reduced enjoyment of the communication due to either disclosing too much text or hiding text that is meant to be shared. To solve these problems, we are developing a new disclosure control mechanism called DCNL or disclosure control of natural language information. DCNL automatically checks texts uploaded to social networking services or blog pages, detects words that might reveal private information, and warns the user about them. The granularity of DCNL is not the text but the words in the text. Consequently, it is not tiresome for the user and balances the protection of privacy with the enjoyment of communications. DCNL checks not only words that directly represent private information but also those that indirectly suggest it. Combinations of words are also checked. Analysis of the co-occurrence between words and reachability analysis with a search engine are used to infer what words imply what information.

Keywords: *SNS, social networking service, privacy, disclosure control.*

1. Introduction

E-communications such as through social networking services (SNSs) and blogs are becoming more and more popular. Large SNSs have more than 100 million users and the numbers of users are still increasing. While these e-communication media have been penetrating more and more into society, even newer types of e-communication media have appeared. For example, micro blogging services are successfully attracting young users.

Although e-communication media make communications more effective and enjoyable, they can also lead to the revelation of private and confidential information [1, 2, 3, 4, 5], the abuse of users, the posting of pornographic and violent content, crimes [6], and suicides [7].

Among these problems, the revelation of private information has been analyzed in several studies [1, 2, 3, 4, 5], which have shown that the revelation can occur in various parts of SNS pages such as the user profile, text in the blog part, photos, and videos. Revelation from the user profile can be prevented by setting the disclosure levels properly because the types of private information in the user profile are limited and relatively static [3]. In contrast, revelation from the other parts is much more difficult to prevent because, in these parts, new information is regularly uploaded, and it is difficult to predict what information will be uploaded in the future. The user should therefore not use the default disclosure settings but should consider the settings every time something new is uploaded. However, this is a tiresome and errorprone task. Moreover, existing disclosure control is typically all-or-nothing; i.e., the whole text is disclosed or hidden, leading to either possibly revealing private information or impairing the enjoyment of SNS communication.

In this paper we describe a new disclosure control mechanism called DCNL or disclosure control for natural language information, which we are implementing. DCNL analyzes the texts in the blog part of a given SNS page and automatically detects words that might reveal private information about the page owner. Its disclosure control is not all-or-nothing but word-by-word, thus balancing the protection of privacy with the enjoyment of communication. It detects not only direct mention of private information but also its indirect suggestion and suggestion by combinations of words.

The difficulty of this detection is that semantic analysis for natural language sentences is not yet established, so problematic words must be detected without being able to analyze their meanings reliably. DCNL uses analysis of word co-occurrence and reachability analysis using a search engine to infer what words imply what information and uses the results to complement the incompleteness of semantic analysis.

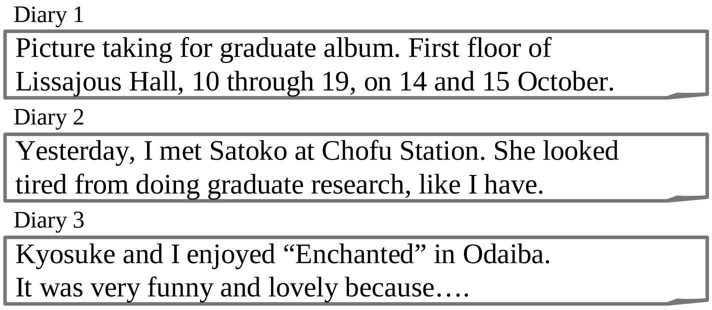
Section 2 of this paper presents the analysis of example sentences to clarify the requirements for DCNL. Section 3 describes the DCNL system design and algorithms. Section 4 reports the simulation of DCNL operations, and Section 5 concludes with a summary of the key points and a mention of future work.

2. Example Analysis and Requirements

As mentioned above, existing disclosure control requires a user to predefine disclosure control rules for the texts expected to be uploaded or to manually set the disclosure level for each text. We illustrate the problems with these methods and clarify the requirements for our DCNL by using example sentences taken from an SNS pages.

2.1. Unsafe Expressions

Consider the three diaries in Figure 1[1], which were actually entered onto an SNS page by a female student at our university (UEC).



Diary 1

Picture taking for graduate album. First floor of Lissajous Hall, 10 through 19, on 14 and 15 October.

Diary 2

Yesterday, I met Satoko at Chofu Station. She looked tired from doing graduate research, like I have.

Diary 3

Kyosuke and I enjoyed “Enchanted” in Odaiba.
It was very funny and lovely because….

Figure 1. Examples of unsafe expressions

[1] The sentences were actually written in Japanese; they were translated into English for this paper.

Although the phrase “Lissajous Hall” in the first diary seems harmless enough, it actually is harmful. A Google search (in Japanese) on this phrase found five mentions of her school in the first ten retrievals. She thus unintentionally revealed her affiliation, which she wanted to hide due to the risk of stalking victimization. She could have avoided this inadvertent disclosure by predefining a disclosure control rule, but it is difficult to anticipate all (or most) of the words that might reveal private information. She also could have manually set the disclosure level of this text so that only her friends could read it, but she might overlook the danger of the phrase “ Lissajous Hall” because this phrase does not directly represent her affiliation. Both predefining many rules for the various possible cases and setting disclosure levels manually for each text would likely be troublesome enough to cause a user to stop using the service, even if doing either were possible.

The combination of “Chofu Station” and “graduate research” in the second diary also reveals her affiliation because “graduate research” implies she is a university student, and the only university around “Chofu Station” is UEC. Moreover, it reveals that she is a graduate student. Thus, even though each phrase alone is relatively harmless, their combination is not. Writing rules or setting disclosure levels for all such combinations would be even more troublesome.

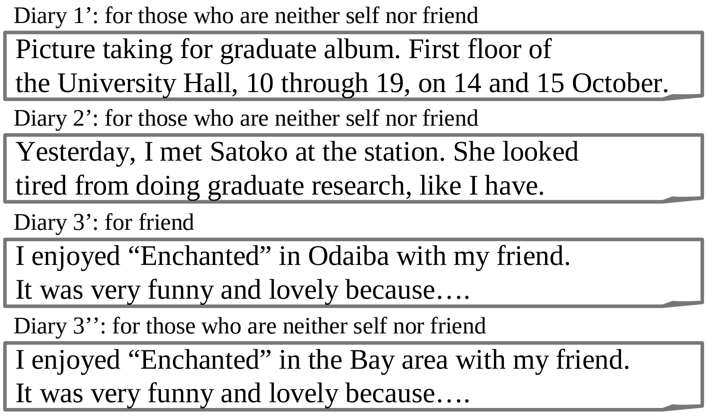
The use of “Odaiba,” the name of an entertainment area, in the third diary implies she has a boyfriend because Odaiba is a popular place for couples. This illustrates that the attributes of objects represented by a phrase need to be considered.

In addition, this diary reveals the name of her boyfriend. This example as well illustrates the difficulty of preparing rules or setting levels for disclosure control.

In addition to these problems, disclosure control is typically all-or-nothing, i.e., whether all the sentences in the target text are disclosed or hidden. It thus leads to either unsafe disclosure or no communication.

2.2. Desirable Transformation

The sentences above reveal private information but could be disclosed after being transformed as shown in Figure 2.



Diary 1’: for those who are neither self nor friend

Picture taking for graduate album. First floor of
the University Hall, 10 through 19, on 14 and 15 October.

Diary 2’: for those who are neither self nor friend

Yesterday, I met Satoko at the station. She looked
tired from doing graduate research, like I have.

Diary 3’: for friend

I enjoyed “Enchanted” in Odaiba with my friend.
It was very funny and lovely because….

Diary 3’’: for those who are neither self nor friend

I enjoyed “Enchanted” in the Bay area with my friend.
It was very funny and lovely because….

Figure 2. Desired transformations

The transformation of the first diary is straightforward omission of the problematic phrase, which is not difficult once it has been identified. The transformation of the second is trickier: the two problematic phrases must first be identified, and then the best one to eliminate must be determined. In the third diary, her boyfriend's name should be omitted, even for friends, because revealing it could damage their friendship. "Odaiba" should be omitted for those other than self and friend because she wants to hide not only the name but also the existence of her boyfriend from them. However, simply omitting it would make the sentence dull and unlively. Its replacement with the ambiguous "the Bay area" results in a sentence that is less revealing but still lively.

As shown in these analyses, the original or the transformed sentence should be disclosed depending on whether the reader is herself or not. Furthermore, different transformations are desired depending on the reader class. One possible solution to this problem is to write different sentences for each class of reader. This, however, would be tiresome and would reduce the enjoyment of the SNS.

2.3. Requirements for DCNL

From these examples, we can derive requirements for a method that would enable safe e-communication.

- Before sentences in the communications are disclosed, they should be checked automatically.
- The granularity of disclosure control should not be the whole text but the words in the text. Thus, any word that could reveal private information should be detected.
- In the detection, not only direct mentions of private information but also indirect mentions should be taken into account. Not only each word but also their various combinations should be taken into account.
- The detected phrases should be either shown to the user so that he/she can modify them or transformed so that they are no longer revealing.
- The burden imposed on the user should not be large. For example, the user should not have to define many detection rules or to modify many sentences.

3. DCNL Design

3.1. System Structure

The system structure is shown in Figure 3. DCNL comprises the four shaded components. When the user uploads text to the blog section of the SNS, the main process reads and sends it to a natural language analyzer, which recognizes the words in the sentences. The recognized words are sent to a suggestion analyzer, which estimates whether the recognized words or their combinations imply private information represented by sensitive phrases. The suggestion analysis is based on a suggestion matrix generated by using word co-occurrence analysis and reachability analysis with a search engine. When it receives the result from the suggestion analyzer, the main process judges whether the words or their combinations directly represent, indirectly suggest, or do not suggest sensitive information. The result of suggestion analysis is stored in knowledge of sensitive phrases for future use. The words that are judged to reveal sensitive information are shown to the user so that he/she can modify them, or they are automatically transformed. When a reader accesses the text, he/she is authenticated, and his/her class is identified. On the basis of the reader's class, DCNL sends the modified sentences instead of the original ones. The current implementation does not include the automatic transformation of detected phrases; they are simply shown to the user as a warning. The knowledge of sensitive phrases consists of sensitive phrases and suggestive relations between phrases.

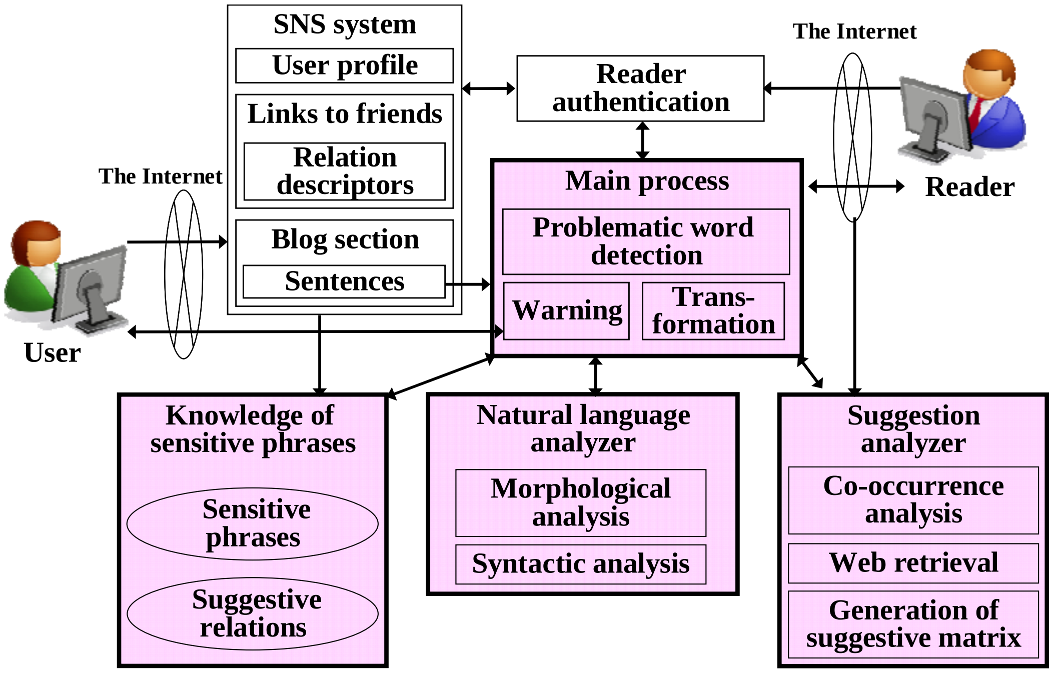


Figure 3. System structure of DCNL

3.2. Knowledge of Sensitive Phrases and Suggestive Relations

We extend the notion of "phrase" by defining it as a sequence of words with an arbitrary length. This is an extension of the linguistic definition: "a phrase is a sequence of words that are continuous in the sentence and are grammatically structured." Our definition ignores these conditions. In our definition, a phrase can represent a single word as well as any combination of words, such as the combination of "Chofu," "station," "graduate," and "research," which are a problematic combination in example diary 2. Thus, a phrase can be the level of granularity of any DCNL operation.

A sensitive phrase is a phrase that the user wants to hide because it reveals sensitive information. The sensitive phrases part in Fig. 3 is a collection of them. They are prepared beforehand, i.e., before the user starts to use the SNS, and are expanded over time. The initial set of sensitive phrases contains phrases from the user's profile, which are typically entered by the user when joining an SNS system. Some SNSs link user pages to friends' pages. Each of these links is labeled with a user-defined relation descriptor, which describes the relation (classmate, boyfriend, etc.) between the user and the friend. DCNL can thus collect sensitive phrases from these descriptors. For the third example diary, it would know that Kyosuke represents a boyfriend and is thus a sensitive phrase.

A suggestive relation shows which phrase suggests a sensitive phrase and how strongly it suggests it. The suggestive relations part in Fig. 3 is a collection of them. Because of our extended notion of a phrase, a suggestion by a combination of multiple words is naturally represented as a suggestive relation between a phrase and a sensitive phrase. The strength of the suggestive relations is normalized in the range 0.0–1.0. The strength of a suggestive relation between the same phrases is 1.0. The direct mention of a sensitive phrase is thus naturally represented. The suggestive relations are calculated for each phrase in the given text. Once calculated, they are recorded in the knowledge of sensitive phrases and used for the same phrases in other texts.

Suggestive relations are calculated by generating a suggestion matrix using the results of phrase co-occurrence analysis and reachability analysis with a search engine. In the second example diary, “Chofu” suggests the sensitive phrase “University of Electro-Communications (UEC)” because these two phrases have a strong co-occurrence relationship. The strength of the suggestion is the degree of the co-occurrence. In the first example diary, “Lissajous Hall” is recognized as suggesting “UEC” because five mentions of UEC were found in the first ten retrievals using Google API with the keywords “Lissajous Hall.” In this case, the strength of the suggestion is the degree of reachability, which is calculated using the Web retrieval result.

3.3. Suggestion Matrix and Generation Algorithm

3.3.1. Suggestion Matrix: The suggestive relations are represented by a suggestion matrix. As shown in Figure 4, each row of the matrix corresponds to a phrase and each column corresponds to a sensitive phrase. The number of rows is N!, where N is the number of words in the target text because we consider every phrase (i.e., every sequence of words) in the text. The number of columns is M, where M is the number of phrases sensitive for the user. Elements S(i, j) represent the strength with which the i-th phrase suggests the j-th sensitive phrase.

Sensitive phrases **Phrases**	1	2	…	j	…	M
1	S(1, 1)	S(1, 2)		S(1, j)		S(1, M)
2	S(2, 1)	S(2, 2)		S(2, j)		S(2, M)
⋮						
i	S(i, 1)	S(i, 2)		S(i, j)		S(i, M)
⋮						
N!	S(N!, 1)	S(N!, 2)		S(N!, j)		S(N!, M)

Figure 4. Structure of suggestion matrix

3.3.2. Generation Algorithm: The algorithm for generating the suggestion matrix uses cooccurrence and reachability analyses. Co-occurrence analysis is used to obtain the (normalized) degree of co-occurrence between the i-th phrase and the j-th sensitive phrase, C(i, j), for $1 \leq i \leq N!$ and $1 \leq j \leq M$. Reachability analysis is used to obtain the normalized degree of reachability from the i-th phrase to the j-th sensitive phrase, R(i, j). Element S(i, j) is the larger of C(i, j) and R(i, j).

3.3.3. Calculating Co-Occurrence and Reachability Degrees: The degree of cooccurrence between words A and B is defined by the following equations, where the numbers of pages are those in the corpus [8].

$$C(A,B) = \frac{\text{Number of pages in the corpus that include } A \text{ and } B}{\text{Number of pages in the corpus that include } A \text{ or } B}\,. \qquad (1)$$

The corpus used here is the complete collection of Web pages retrieved by a search engine. Thus, the numerator in equation (1) is the number of Web pages retrieved by the search engine under the condition that words A and B are included. The denominator is similarly calculated. The co-occurrence degree between Chofu and UEC and that between Cambridge and UEC are

$$C(Chofu,UEC) = \frac{14000}{1960000} = 7.14E^{-3} \text{ and} \tag{2}$$

$$C(Chofu,Cambridge) = \frac{19200}{111000000} = 1.73E^{-4}\,. \tag{3}$$

These degrees show that “Chofu,” the location of UEC, implies UEC much more strongly than “Cambridge,” which is not. The degree of reachability is calculated using the Web retrieval results for the target phrase as the search expression, that is, on the basis of the number of retrievals containing the sensitive phrase and their position in the search results.

3.4. Detection of Problematic Words and Warning

The algorithm for detecting words that reveal private information is run when the user uploads text. The algorithm first generates suggestion matrix S for the text. Each row of S corresponds to the suggestive relations between a phrase in the text and sensitive phrases (which are or are not included in the text). If DCNL has encountered the same phrase in a previous text and thus has stored the suggestive relation for the phrase, it simply retrieves the stored relation. If not, it calculates the suggestive relation from scratch. Figure 6 shows the suggestion matrix for the first example diary.

If S(i,j) ≤ T for 1 ≤ i ≤ N! and 1 ≤ j ≤ M, where T is the decision threshold, then do nothing. Else omit words in the text so that S’(i,j) ≤ T for 1 ≤ i ≤ N’ and 1 ≤ j ≤ M, where S’ is a submatrix of S made by the omission and N’ is the number of rows of S’. Note that if word W is omitted, all the phrases that contain W and the corresponding rows are deleted from the suggestion matrix. The strategy for omission is as follows.

- If S(k, l) > T and the k-th phrase consists of only one word, omit this word.
- The larger the S(k, l), the more preferred the omission of the k-th phrase. This means that one of the words contained in the k-th phrase is preferably omitted.
- The greater the number of matrix elements that are larger than T and that would be deleted by the omission of a word, the more preferable the omission of this word.

The set of words that have been omitted are the problematic words, and they are shown to the user.

4. Simulation

The operation of DCNL was simulated using the example diaries in Figures 1 and 2. For the first sentence in Figure 1, the suggestion matrix shown in Figure 5 was generated. Detection threshold T was set to, for example, 1.0E-01. The algorithm for detecting problematic words found that {Lissajous, Hall} suggests a sensitive phrase {UEC} with a degree greater than T. Because {Lissajous} suggests {UEC} more strongly than {Hall} does, the algorithm omits “Lissajous” from the text and shows this phrase to the user. The user repairs the diary, changing it, for example, to Diary 1’ in Figure 2.

Sensitive phrases Phrases	UEC	Kyosuke	・・・
picture	4.93E-04	3.17E-03	
taking	4.78E-04	3.09E-04	
for	3.40E-03	8.41E-04	
graduate	2.18E-03	3.35E-04	
album	1.83E-04	6.09E-05	
・・・			
Lissajous	3.02E-03	2.69E-04	
Hall	1.44E-03	1.70E-04	
・・・			
picture taking	5.52E-04	9.48E-04	
graduate album	6.10E-04	3.42E-04	
Lissajous Hall	1.29E-02	2.03E-03	
・・・			
for graduate album	1.03E-04	1.32E-03	
・・・			
Picture taking for graduate album. First floor of Lissajous Hall, 10 through 19, on 14 and 15 October.	0	0	

Figure 5. Suggestion matrix for example diary 1

For Diary 2, the algorithm works the same as for the first example. Because of the extended notion of a phrase, all combinations of words are listed in the rows of the suggestion matrix, and the combinatorial suggestion of “Chofu,” “station,” “ graduate,” and “research” is naturally identified. For Diary 3, the system uses two different values for threshold T. The value used for Diary 3’ is larger than that used for Diary 3’’; i.e., more disclosure is allowed for a friend.

5. Conclusion and Future Work

Private information can be revealed by natural language text entered into social networking services or onto blog pages that are written by page owners and their friends. Existing techniques for disclosure control are not effective because consideration of the disclosure level for each text to be uploaded is tedious and errorprone, and the all-or-nothing approach to disclosure impairs the enjoyment of communications. Our proposed disclosure control of natural language information automatically checks texts on SNS and blog pages and detects words that might reveal private information. Its disclosure control is not all-or-nothing but for each word. Thus, DCNL is not tiresome for users and balances the protection of privacy with the enjoyment of communication.

DCNL detects not only direct mentions of private information but also indirect suggestions and suggestions by combinations of words. This broad detection is made possible by the generation of a suggestion matrix for each text to be uploaded. Each element in the matrix represents the strength with which a word combination in the text suggests private information. The matrix is automatically calculated using word co-occurrence analysis and reachability analysis using a search engine. DCNL enables different levels of disclosure in accordance with user class. This is made possible by evaluating suggestion matrix elements with different thresholds.

References

[1] Gross, R., Acquisti, A.: Information Revelation and Privacy in Online Social Networks. In: Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society (WPES), New York, pp. 71–80 (2005)

[2] Lam, I., Chen, K., Chen, L.: Involuntary Information Leakage in Social Network Services. In: Proceedings of 2008 International Workshop on Security, Kanagawa, Japan, pp. 167–183 (2008)

[3] Lewis, K., Kaufman, J., Christakis, N.: The Taste for Privacy: An Analysis of College Student Privacy Settings in an Online Social Network. Journal of Computer-Mediated Communication 14(1), 79–100 (2008)

[4] Data Protection Working Party: Opinion 5/2009 on online social networking,

http://epic.org/privacy/socialnet/Opinion_SNS_090316_Adopted.pdf

[5] Viegas, F.: Bloggers' Expectations of Privacy and Accountability: An Initial Survey. Journal of Computer-Mediated Communication 10(3) (2005)

[6] Calvo-Armengol, A., Zenou, Y.: Social Networks and Crime Decisions: The Role Of Social Structure in Facilitating Delinquent Behavior. International Economic Review 45(3), 939–958 (2004)

[7] ABC news: Florida Teen Live-Streams His Suicide Online,

http://abcnews.go.com/Technology/MindMoodNews/story?id=6306126&page=1

[8] Weeds, J., Weir, D.: Co-occurrence retrieval: A Flexible Framework for Lexical Distributional Similarity. Computational Linguistics 31(4), 439–475 (2005)

Authors

Haruno Kataoka received the B.S. and M.S. degrees from the University of Electro-Communications, Japan, in 2006 and 2008, respectively. She joined the NTT Information Sharing Platform Laboratories, NTT Corporation in 2008, and has been engaged in research and development of Web contents collaborative system for call parties. She is a member of the Information Processing Society of Japan (IPSJ).

Natsuki Watanabe received B.S. from the University of Electro-Communications, Japan, in 2008. She is currently a master course student in the Graduate School of Human Communication, the University of Electro-Communications and has been continuing research on DCNL.

Keiko Mizutani received B.S. from the University of Electro-Communications, Japan, in 2008. She is currently a master course student in the Graduate School of Human Communication, the University of Electro-Communications and has been continuing research on detection of private information.

Hiroshi Yoshiura received his B.S. and D.Sc. from the University of Tokyo, Japan, in 1981 and 1997. He joined Hitachi, Ltd. in 1981 and until 2003 was a Senior Research Engineer in the company's Systems Development Laboratory. He is currently a Professor in the Graduate School of Human Communication, the University of Electro-Communications. He has been engaged in research on information security and copyright protection technologies and received the President's Technology Award from Hitachi in 2000, the Best Paper Award from IPSJ in 2005, the Industrial Technology Award from ISCI in 2005, and the Best Paper Award of IIHMSP-2006.