Mathematical Duality between Anonymity and Privacy and Its Application to Law

Ken Mano†

Abstract

My colleagues and I have developed a formal verification method for the anonymity and privacy of information systems based on the idea that anonymity is hiding information about who performed a certain action and that privacy is hiding information about what action a certain person performed. In this article, I investigate the use of a formal representation of privacy in a legal context. As an example, I compare such a formal concept of privacy with a legal one. On the basis of a comparison with the *after the banquet* case, whose decision is widely used as a precedent, I point out a difference between the two concepts related to identifiability and investigate its meaning. To clarify the meaning of the difference, I use ambient intelligence as a technical context.

1. Introduction

Formal methods [1] are methods aiming at verifying the correctness of information systems in a logically rigorous manner. Such correctness is usually confirmed by putting test data into the system and checking the result. However, it is generally impossible to input all available data, so untested data could cause errors. In systems that require very high reliability such as mission-critical systems or security systems, such a possibility cannot be overlooked. Formal methods have been proposed as a way to solve this problem. By formally describing the system's correctness and by formally proving it, we can achieve a highly reliable system.

My colleagues and I are conducting research on privacy verification by using a formal method [2]. One of the characteristics of our method is the approach that we use for the formal formulation of privacy properties. Privacy is a somewhat ambiguous notion, and there are two key points in our approach regarding its formalization: (1) the use of epistemic logic and (2) the formulation of privacy as the hiding of information about what action a certain person performed or what states a certain person is in. The first point is important because security problems, including privacy, are problems concerning how much information can be known by an adversary of the system. Thus, we chose a formal framework appropriate for formalizing knowledge. The second point arose from comparison with anonymity, which is explained later.

In this article, I describe the use of a formalization method for anonymity and privacy in a legal context. Specifically, I compare formally and legally defined concepts of privacy. I do not intend to perform a thorough comparison. On the basis of a comparison with the *after the banquet* case [3], whose decision is widely used as a precedent, I point out a difference between the two concepts in terms of identifiability and investigate its legal meaning.

Identifiability is one of the requirements of a tort of privacy invasion, which would appear quite natural in a daily sense. Describing the problem in a logical formulation is expected to be useful for examining the meaning of such a *seemingly natural* requirement and for obtaining a beneficial legal consequence.

Not only should we devise the way of formulating using the logical formulation; we must devise

† NTT Communication Science Laboratories
Atsugi-shi, 243-0198 Japan

examples to clarify the difference between two concepts. For this purpose we use *ambient intelligence* as a technical context. Ambient intelligence is an ability that actively reacts to activities of people in a ubiquitous network environment by using electronic tags and sensors.

This article is organized as follows. Sections 2 and 3 explain privacy as formal and legal concepts, respectively. Section 4 compares the two concepts and points out a difference between them with respect to identifiability. Its meaning is investigated in section 5.

2. Privacy as a formal concept

There are several methods for formally formulating and verifying privacy. Here, I introduce our approach [2], which is based on anonymity research by Halpern et al. [4]. One of the characteristics of their research is that they utilize mathematical logic called epistemic logic. The expressions of epistemic logic are not only rigorous but also relevant to intuition, so epistemic logic is appropriate for legal issues.

In mathematical logic, the basic elements are expressed by symbols, and logical statements are expressed as formulae. We use the symbol $\wedge$ to express *and*, $\vee$ for *or*, $\Rightarrow$ for *implies*, and $\neg$ for *not*. The logical statement "If A or B holds, then C holds and D does not hold." is expressed by the following mathematical logic formula: $(A \vee B) \Rightarrow (C \wedge \neg D)$.

The original purpose of mathematical logic was to express the truth of logical statements, but there are some extensions for making it easy to express human epistemic activities such as "i knows that ….", "j thinks it is possible that ….", etc. Such an extension is called *epistemic logic*. In epistemic logic, K_i is an operator expressing i's knowledge (K is the initial letter of *know*). For instance, if *Beautiful*(*rose*) is a formula expressing "A rose is beautiful.", then K_i(*Beautiful*(*rose*)) means that "i knows that a rose is beautiful." By combining the knowledge operator with negation $\neg$, we can express possibility. For instance, if *Delicious*(*grasshopper*) expresses "a grasshopper is delicious", then $\neg K_i \neg$(*Delicious*(*grasshopper*)) means that "i does not know that a grasshopper is not delicious", in other words, "i thinks that it is possible that a grasshopper is delicious." Thus, using the initial letter of *possibility*, $\neg K_i \neg$ is abbreviated to P_i.

Halpern et al. proposed a general method for formulating anonymity properties of information systems by using epistemic logic [4]. For example, they call an anonymity property defined by the following epistemic formula *anonymity up to* I_A.

$$\theta(i, a) \Rightarrow \bigwedge P_j[\theta(i', a)]$$

Here, $\theta(i, a)$ is an abstract expression indicating that i has performed action a (or that i is in state a). In other words, θ expresses a linkage between people and actions (states). We regard anonymity properties as how well the information about *who performed a certain action* is hidden. This makes the method general, in the sense that it is independent of specific applications. Moreover, I_A is a subset of participants called the *anonymity set*, and j is an observer of the system. Thus, the meaning of the whole formula is "if i has performed a, then j thinks that a could have been performed by anybody in I_A." The demand for anonymity can vary from system to system, so some systems can require more complicated forms of anonymity. Even in such a case, the expressiveness of epistemic logic enables a flexible description of the required anonymity. By describing such an anonymity formula and choosing an appropriate interpretation of θ and I_A, we can define the specific anonymity required for a specific system.

On the basis of the above research, we proposed a method for formulating and verifying privacy properties formally [2]. In this method, we consider privacy properties that can be expressed as a relation between people and their actions. In other words, we regard privacy properties as how well the information about *what action a certain person performed* is hidden. Then, privacy can be formalized in a similar way to the formalization of anonymity. That is, we can define various privacy properties in terms of knowledge about $\theta(i, a)$, and as with anonymity, an epistemic-logic-based approach provides a good way of specifying them.

For example, we define a property called privacy up to A_I, whereby, from j's point of view, i could have performed any action A_I:

$$\theta(i, a) \Rightarrow \bigwedge_{a' \in A_I} P_j[\theta(i, a')]$$

Note that the above formulae for privacy and anonymity are symmetric[*1]. However, in general, there are no logical dependencies such as one holds if the other holds, or one holds if the other does not hold. They are logically independent of each other.

*1 This symmetry can be thought of as a kind of duality.

3. Privacy as a legal concept

In this section I introduce privacy as a legal concept in Japan. Here, I concentrate on the right of privacy in private law. There are other related legal concepts, e.g., the right to control personal information in public law or the Personal Information Protection Law, but they are beyond the scope of this article.

When we consider the right of privacy in private law, the most important decision is that of the *after the banquet* case. This decision recognized the legal guarantee and/or right that private life not be disclosed without reason so that personalities are mutually respected and private individuals are protected against unjustifiable interference. In its decision, the court presented the following three tort requirements of privacy right violation for the disclosed matter: (1) it is true, or can be taken to be true in intimate life, (2) it is offensive to a reasonable person from his or her viewpoint, and (3) it is not of concern to public.

For (1), there is an additional requirement, which is not explicitly stated in the above three requirements: the said person must be identifiable by the disclosed matter. This is essential, e.g., in the case of invasion of privacy by a novel based on real people and incidents.

In the following, I use the above requirements as the definition of privacy as a legal concept and call it *legal privacy*. That is, legal privacy is protected, or satisfied, when any of the above three conditions does not hold. In the following, I investigate the difference between legal and formal privacy.

First, a trivial difference is that legal privacy has the explicit conditions "offensive" and "not of concern to the public", while formal privacy does not, or at least such conditions are implicit. We regard this difference as trivial since it originates from the basic idea of formal privacy formulation, where we concentrate on privacy properties that can be expressed as a linkage between persons and actions. However, a specific privacy definition suitable for a specific case can be obtained by fixing the interpretations of I_A, A_I, and θ of formal privacy. The conditions of legal privacy should be reflected in the interpretation. On the other hand, for legal privacy, whether or not the matter is offensive and whether or not it is of concern to the public are judged according to the individual situation. There seems to be rather a natural correspondence between them.

As mentioned at the beginning of this article, it is not my objective to thoroughly compare formal and legal privacy. In the following, I concentrate on the difference as regards identifiability: the identifiability of the said person is a requirement in the definition of legal privacy, while there is no corresponding condition in the definition of formal privacy. Unlike in the trivial difference mentioned above, there is a reason for the lack of this condition in formal privacy. The negation of identifiability is valid as a necessary condition for anonymity. Actually, both properties concern hiding information about who performed a certain action. However, formal privacy hides the information about what action a certain person performed. They are generally independent, and there must be no relation between them such as one being a necessary condition for the other.

From a legal standpoint, the most plausible interpretation of this difference is that the definition of formal privacy is insufficient. However, this article poses the following question in order to examine the legal meaning of the difference: Is there any social situation or technical context where formal privacy is more suitable than legal privacy? This question can be generalized to whether or not identifiability is necessary. In the following, I call this the *identifiability problem*.

4. Identifiability problem

In this section, I investigate the problem posed in the previous section by using ambient intelligence [5] as the technical context.

4.1 Ambient intelligence

The ambient in ambient intelligence means the environment of a ubiquitous society, where electronic tags, biometric sensors, and networks connecting them are widespread. The intelligence is of course analogous to artificial intelligence and means the ability to actively react to human activities by processing information gathered by the ambient intelligence using natural language processing, knowledge and data processing, and other means [6].

If ambient intelligence is achieved, it will be convenient for users in various ways. On the other hand, some potential problems have been pointed out. One of the biggest criticisms is the potential loss of privacy. For example, let us consider the following case. Person A uses an ambient intelligence system at home. The system takes care of his family's health by using biometric technology and a healthcare service site on the Internet. One day, B visits A, and the ambient intelligence system measures B's biometric information and sends it outside the system. B could

potentially file a lawsuit against A for invasion of privacy.

There are two important points in this case. One is that the root of the problem lies in the incentive to actively circulate the obtained information, which is assumed to lead to improved convenience for individuals and society. This can lead to private information being communicated without foresight. This is not restricted to ambient intelligence; for instance, the privacy issues related to Google Street View can be thought of as a consequence of such an incentive. The other point is that this is not only a technical problem but also a social problem and in particular a legal problem.

Such potential problems with ambient intelligence have been extensively studied [7]. For instance, the aim of PRIAM [8], a research project in INRIA [9], is transversal and multidisciplinary research through the exchange of ideas between lawyers and experts in information and communications technology. One of the specific objectives of PRIAM is to develop a methodology for privacy policy specifications, and a formal method is expected to provide a promising approach for handling both technical and social problems.

4.2 Two cases

To investigate the difference mentioned in section 3, let us consider the following two cases.

(1) Biometric information

In ambient intelligence, biometric information is assumed to be recorded and exchanged in various situations. Concerning the handling of such information, there have been reports and discussions in relation to the Personal Information Protection Law. Among the arguments, those related to whether or not some information corresponds to personal information under the Personal Information Protection Law are closely connected with our problem. Shimpo [10] pointed out the importance of the existence of referable information for distinguishing individuals. Murakami [11] pointed out the importance of the owner of such referable information. A similar precise argument is needed for invasion of privacy.

Here, let us consider an extreme case to make the problem clear. Sato stated that a symbolic issue of privacy is that the analysis of a complete human genome has become possible [12]. Genome information can be regarded as the ultimate personal information. Now let us assume that your complete genome information is disclosed without reason. However, only the genome information itself is disclosed, and no information about your identity such as your name and address is provided. Is this an invasion of your privacy? A key point here is the possibility of identifying someone solely from their complete genome information.

Actually, the owner of a given genome is uniquely determined with very high probability, except in the case of an identical twin. However, it is generally difficult to determine a genome's owner by using only the genome information since that person could be anywhere in the world. So if we conclude that the identification of the said person is impossible, the disclosure of genome information would not in itself be an invasion of privacy.

There are various possible interpretations of this problem. For instance, it is possible to think that it is not a problem to disclose genome information about an unnamed person. However, today's biotechnology, which is enabling personalized medicine and genomic medicine, seems to be leading to a situation in which such a pastoral attitude is not allowed. For instance, genome medicine could be used to produce material that is harmful only to the genome's owner.

(2) Surveillance of a bank doorway

Let us consider the following situation. A publicly accessible surveillance camera is placed in front of a bank door. For privacy protection, the camera's image output is pixelated by realtime image processing so that people in the images cannot be identified by their faces. However, other parts of their bodies remain clear, and we can recognize features such as shirt color and the extent to which a bag is bulging. Moreover, a thermographic camera and a microwave-based motion sensor are also installed near the door and they are connected to the ambient intelligence and are publicly accessible.

Information about clothing is insufficient to identify a person, but sufficient to allow us to guess that *the person coming out now is the same person who went in three minutes ago*. Many public organizations in Japan are setting up infrared thermographic cameras in preparation for a pandemic of a new type of influenza. Automatic sliding doors often have a type of approach sensor called 2.4-GHz microwave Doppler radar. The same type of radar has been used in research aimed at remote sensing of the human pulse [13].

Thus, on the basis of the above knowledge, suppose that a third party observes that the bag of a man just coming out of the bank is bulging more than at the time he entered, that his pulse rate is elevated, and that his body temperature is lower. This third party

could presume that the man has withdrawn a large amount of money. Is this an invasion of the man's privacy?

What I want to stress here is the following. We are considering situations where the incentive to publish and circulate obtained information is dominant, that various types of information about a person are published since they do not identify that person, and thus that such information can easily be accumulated. Then, even if identification is impossible, some information can still be known, e.g., that a man in a red shirt will pass a certain place carrying a lot of money, and such knowledge can easily be used for malicious purposes. If we suppose, in the first case, that the disclosed genome information is known to be that of a rich person, the situation is similar to this case.

5. Self-information condition

It seems likely that a reasonable person would feel that the cases described in the previous section are invasions of privacy. However, it cannot be so according to the precedent that identifiability is required. Moreover, the disclosed information does not correspond to personal information under the Personal Information Protection Law since it is insufficient to distinguish individuals, so the cases do not conflict with the law.

The idea of requiring identifiability dates back to early legal arguments on the right to privacy. For instance, in the four types of invasion of privacy in the classic research by Prosser [14], the second type called *public disclosure of private facts* is closely related to the right recognized in the decision in the *after the banquet* case. The protected interest of this type is considered to be reputation[*2]. When the protected interest is reputation, identifiability is required.

Should we devise a new interest or right in order to avoid the situations highlighted in the previous section? The more widely a person's right to privacy is recognized, the more restricted the rights of others to know become. Thus, we should not devise new rights thoughtlessly. In fact, I think that it is possible to regard the above situations as invasions of the recognized interest that private life not be disclosed without reason.

In many privacy torts, including the *after the banquet* case, the problem was damage to reputation, that is, detriment that accrues from a person being blamed behind his or her back. Therefore, identifiability of the person to be blamed became a requirement. However, the cases in the previous section show that, in future technical contexts, private lives could be invaded for no reason, and serious detriment could be caused in a completely different way from blaming. It is obviously impossible to reduce this problem to that of the right not to be blamed behind one's back. Actually, one of the main criticisms of the Prosser's classification is that it constitutes such a reductionistic attitude to the right of privacy [15].

So, instead of reducing the problem of privacy to a legal guarantee of reputation or of not being criticized, let us return to the decision in the *after the banquet* case, which recognizes the legal guarantee and/or right for one's private life not to be disclosed without reason in order that personalities be mutually respected and that private individuals be protected against unjustifiable interference. If we consider the right itself as the interest to be protected, the genome information and surveillance information in the cases described in the previous section should not be disclosed without reason.

Now, recognizing the problem in the previous section as a legal one, what solution is possible? In the following, I consider how to solve this problem by making a minimal modification to the interpretation of the decision in the *after the banquet* case. Specifically, I consider how to relax the identifiability requirement.

First of all, it is nonsense to simply eliminate the identifiability requirement. If we did that, we would lose any condition that requires the disclosed matter to concern the said person, and then disclosure of matters related to other people could be an invasion of the said person's privacy. Thus, let us instead consider adding a requirement to avoid such a silly situation, that is, a requirement that the disclosed matter is a true fact about the person. In this article, I call this requirement *self-information*. The difference between identifiability and self-information is that the former concerns whether or not others can know who the person is from the disclosed matter, while the latter concerns whether or not the matter is about the person regardless of how it is seen by the others.

To fulfill self-information in the case of genome information disclosure, it is enough to prove that the disclosed genome information is that of the said person by using a DNA (deoxyribonucleic acid) test. With surveillance information, let us suppose the court recognizes that, even in a public area, there is no implicit consent to allow the disclosure of an

*2 There are criticisms concerning the idea; I refer to these later.

unnecessarily detailed image, body temperature reading, or pulse rate of a person. Then it is sufficient to prove that, for example, the sensor was operating when the person was there. Then we can prove these cases to be illegal regardless of identifiability.

However, it is impossible simply to substitute self-information for identifiability. As is mentioned in the first requirement, the disclosed matter is true, or *can be taken to be true* in intimate life. That is, the matter does not have to be true, whereas self-information must be true.

So such a simple replacement would not be a minimal modification since it would narrow the range of privacy protection. Therefore, we propose to interpret the disjunctive sentence as actually expressing a disjunction of two requirements as follows:

- "it is true in intimate life" expresses self-information and
- "it can be taken to be true in intimate life" expresses identifiability.

In the following, I call the legal privacies with the former and latter requirements *legal privacy based on self-information* and *legal privacy based on identifiability*, respectively.

Interestingly, this relaxation of identifiability is closely related to formal privacy. To clarify the relationship, let us ignore the conditions "offensive" and "not of concern to the public" that are abstracted from formal privacy. Then legal privacy is something concerning an identifiable and disclosed matter. Let us consider an invasion of privacy where a private matter "i performed action a" of i is true. If we formally interpret identifiability as the negation of anonymity up to I_A and disclosure as the negation of the conclusion part of privacy up to A_I, then legal privacy is interpreted as being expressed by the following formula.

$$\theta(i, a) \wedge \left\{ \neg \bigwedge_{i' \in I_A} P_j[\theta(i', a)] \right\} \wedge \left\{ \neg \bigwedge_{a' \in A_I} P_j[\theta(i, a')] \right\}$$

In this interpretation, formal versions are weaker than legal versions. For instance, identifiability implies failure of anonymity but the reverse does not hold in general. Nevertheless, the above formula is too restrictive, or too strong, as a formula expressing the invasion of formal privacy[*3]. So, to weaken the above formula, let us omit the second condition $\neg \bigwedge_{i' \in I_A} P_j[\theta(i', a)]$. The resulting formula

$$\theta(i, a) \wedge \left\{ \neg \bigwedge_{a' \in A_I} P_j[\theta(i', a)] \right\}$$

intuitively means that the matter is self-information $\theta(i, a)$ and is disclosed ($\neg \bigwedge_{a' \in A_I} P_j[\theta(i, a')]$), which corresponds to legal privacy based on self-information. And this formula is also acceptable as a formula expressing an invasion of privacy in a formal sense, since it is itself the negation of privacy up to A_I.

6. Concluding remarks

I compared formal privacy with legal privacy and investigated their difference and its meaning by using ambient intelligence as a technical context. This comparison yielded a problem with identifiability, and to resolve it I introduced the notion of self-information. Moreover, I showed that there is a close relationship between legal privacy based on self-information and formal privacy.

There are various issues related to self-information. For instance:

- What kind of self-information is the target of the protection? What is offensive self-information?
- How good a match is sufficient to judge that a certain matter is self-information? It is easy to change just the appearance of digital data.
- How about the case where the matter is not disclosed but exploited at someone's discretion?

These cases are always subjects of privacy whether or not they are based on self-information. It is not apparent whether a formal method can help solve these problems.

A problem specific to self-information is the following. Although a person can prove illegality regardless of identifiability, he or she is identified by the very fact of proving in court that the matter is self-information, which could be detrimental. First, a person may want to file a lawsuit even if it enables him or her to be identified, which should be admitted by law. Next, this problem can resolved if the privacy court is anonymized.

Moreover, it would be worthwhile applying the same perspective to investigations of the Personal Information Protection Law.

*3 This is proved by formally showing the existence of a model that formally satisfies both the negation of privacy and the negation of the above formula expressing legal privacy invasion. The cases in the previous section can be regarded as similar models in the context of ambient intelligence.

References

[1] Y. Tsukada (Eds.), "Special Feature: New Trends of Formal Methods," Information Processing Society of Japan, Vol. 49, No. 5, pp. 491–543, 2008 (in Japanese).

[2] K. Mano, Y. Kawabe, H. Sakurada, and Y. Tsukada, "Role Interchange for Anonymity and Privacy of Voting," Journal of Logic and Computation, Vol. 20, No. 6, pp. 1251–1288, 2010.

[3] After the Banquet. See e.g., http://en.wikipedia.org/wiki/After_the_Banquet

[4] J. Y. Halpern and K. R. O'Neill, "Anonymity and Information Hiding in Multiagent Systems," Journal of Computer Security, Vol. 13, No. 3, pp. 483–514, 2005.

[5] Y. Tonomura and E. Maeda (Eds.), "Proposing Ambient Intelligence—New Paradigm for Information-driven Society," Maruzen, 2008 (in Japanese).

[6] W. Weber, E. H. L. Aarts, and J. Rabaey (Eds.), "Ambient Intelligence," Springer, 2005.

[7] D. Wright, S. Gutwirth, M. Friedewald, E. Vildjiounaite, and Y. Punie (Eds.), "Safeguards in a World of Ambient Intelligence," Springer, 2010.

[8] PRIAM: Privacy Issues and Ambient Intelligence. http://priam.citi.insa-lyon.fr/

[9] INRIA. http://www.inria.fr/recherches/domaines-de-recherche/actions-de-recherche-collaboratives (in French).

[10] F. Shimpo, "Use of Biometrics Based on Personal Information Protection Law Journal of Information and Media Studies, Vol. 4, No. 1, pp. 55–76, 2006 (in Japanese).

[11] Y. Murakami, "Legal Issues Related to Biometrics," The Information Network Low Review, the Information Network Law Association Japan, Vol. 4, No. 2, pp. 74–92, 2005 (in Japanese).

[12] K. Sato, "Article 13 of the Constitution of Japan and the Rights of Controlling Self-information," Lecture Meeting, the Information Network Law Association Japan, 2009 (in Japanese).

[13] S. Suzuki, T. Matsui, and S. Fujie, "Trial of Contactless Measurement of Heart-rate-variability Index during Work Using Microwave Rader," Journal of Japan Ergonomics Society, Vol. 44, Special Issue, pp. 272–273, 2008 (in Japanese).

[14] W. L. Prosser, "Privacy," California Law Review, Vol. 48, No. 3, pp. 383–423, 1960.

[15] R. Gavison, "Privacy and the Limits of Law," The Yale Law Journal, Vol. 89, No. 3, pp. 421–471, 1980.

Ken Mano

Senior Research Engineer, Innovative Communication Laboratory, NTT Communication Science Laboratories.

He received the B.E. degree in applied physics and the M.E. degree in information engineering from Nagoya University, Aichi, in 1987 and 1989, respectively. He joined NTT Basic Research Laboratories in 1989 and studied term rewriting systems and process algebra. He is currently studying formal verification of security systems. He is also interested in legal aspects of security and privacy. He is a member of the Institute of Electronics, Information and Communication Engineers, the Information Processing Society of Japan, and the Japan Society for Software Science and Technology.