Anonymity revisited: the degree of the knowledge transfer

Jong-Hyeon Lee

University of Cambridge, Computer Laboratory
Pembroke Street, Cambridge CB2 3QG

Abstract. Anonymity has been regarded mainly as anonymity of the person who did something. We believe that the anonymity of the subject may not be a primary concern and the more important part would be the fact that the subject did some deed. We focus on the relationship of the subject and the deed. Another argument we raise here is the need of quantitative measures for the anonymity. Anonymity has been dealt only as a quality, but we need to quantise the quality to measure and compare anonymising mechanisms. We provide a metric, we call it anonymity metric, in order to measure the anonymity of mechanisms.

As a typical example for anonymising mechanisms, we consider electronic voting schemes. Secret ballot definitely needs not to expose relationship between voter and his ballot. On this basis, most electronic voting schemes have been designed to keep anonymity of voters and nontraceability of ballots. We apply the anonymity metric to anonymous mechanisms in voting schemes and present metric computations.

1 Introduction

Anonymity is the state of being or remaining unknown to most other people. The concept is mostly regarded as anonymity of a person not as that of the deed that had been done by the person. For example, Alice buys a product with digital cash from an on-line shop. Anonymity of this case is two-folded. One is anonymity of the identity of hers, and the other is that of the deed. In this case, the anonymity of the identity is not so important since she is known to a group of people anyway such as family, friends, or neighbourhoods. More important points are the content of the purchase and the fact that she bought something somwhere. The content of the message can be protected by an encrypted channel but the anonymity of the deed itself is another question; the deed is not protected by the channel. The target of anonymity is the relationship of the subject and his deed.

Anonymity has some similarity with confidentiality in view of keeping something secret, but from the definition of anonymity, it is clear that anonymity is different from confidentiality; anonymity usually does not have a precise boundary of a group of people who know something (whatever it is either object or subject), but confidentiality does. For example, a person votes "yes" for an Act, the deed is spied by his neighbour, and the neighbour transferred the deed to his neighbours. This makes a transfer chain. It is not clear to the voter how many people knows the deed and who they are; his deed is anonymous to the public but never anonymous to this group of people in the chain. The size of the channel is not precise. On the contrary, confidentiality assumes a precise boundary of the group. It is clear that who can read an encrypted message. The group of people who do not know an object, we call *range of anonymity*, is an important factor for anonymity. When we say anonymity of an object, we should state the range of anonymity.

What makes the difference between anonymity and confidentiality? One is the chain of knowledge transfer. The chain defines the range of anonymity. The more people are involved in the chain, the less anonymity we can get. Publication has a complementary aspect to confidentiality in terms of anonymity; publication is the state of being known to the public; the longer the knowledge transfer chain is, the faster publication we can get. So anonymity presents the knowledge degree between publication and confidentiality; if an event is less anonymous, it is closer to publication rather than confidentiality, and if it is thoroughly anonymous to everybody but the person who made the event, it is confidential. We identify confidentiality by perfect anonymity and publication zero anonymity. We need to incorporate a measure to represent this degree, we call *knowleadgeability* κ. The way how to control the knowledgeability is the whole point of anonymity management. The knowledge transfer chain also provides traceability since knowledge transfer path is the trace. As far as the chain is available, we can trace back something (subject, object, or event itself) along with the chain. Conventionally anonymity is not a quantitative concept but a qualitative one, however we can quantise it with respect to the chain and see quantitative properties such as the speed of knowleadgeability decay, depth and structure of the chain, etc.

Another difference is the "object" of these properties. The object of confidentiality can be principals or data used by principals; i.e., we want to hide principals themselves or data itself secretly. That of anonymity is mainly the trace or the relationship between a principal and his data; for example, we know who is involved in an election, how many ballots are made, and what is on a voting slip after the election, but we do not want to reveal who votes for whom/what. Although anonymity for an object is still sensible, anonymity is mainly a matter of a deed: what somebody does. This is the whole point of anonymity to make it hard to trace a deed back. In this respect, we consider only the anonymity of deeds in the paper.

2 Anonymity metric

We introduce a measure to quantise anonymity. The anonymity level called knowledgeability κ is defined between zero and one. The perfect anonymity, the same level as confidentiality, represents $\kappa = 0$, and the the perfect knowledge, the same level as publication, means $\kappa = 1$.

The knowledgeability κ is decided by the way how to control the knowledge transfer channel. What are the factors that affect anonymity of a deed? They are the number n of principals who know or handle an object or its trace, and the way τ how to transfer the knowledge; that is, $\kappa = f(n, \tau)$ where τ is a function to provide some level of anonymity and represents degree of anonymising factor with range of $0 \leq \tau \leq 1$. Since one can only reveal the knowledge that he received, τ cannot be larger than 1. Even though one of principals expose all the knowledge he has obtained, the degree of anonymity which has obtained before can be kept unless the protocol itself is ill-constructed. In this respect, anonymising process is one-way; no one can obtain more knowledge than he received and the degree of the knowledge only decays.

Basic model we can think is a model with two principals, data sender and receiver, i.e., $n = 2$. There are two possibilities in this model: perfect anonymity or perfect knowledge, that is,

$$\kappa = f(n, \tau) = \begin{cases} 0 & \text{if } \tau \text{ provides a secure channel;} \\ 1 & \text{otherwise.} \end{cases} \tag{1}$$

This is the case that the deed between principals is either totally secret or open to everybody. If there is an apparatus that enables both principals to communicate without revealing themselve, the deed can be totally anonymous. Otherwise, the deed can be known publicly. No intermediate degree of anonymity is allowed.

For $n \geq 3$, we can put some functions between sender and receiver in order to increase anonymity of the deed done by principals. If all principals between sender and receiver function in the same way, we formulate $\kappa = \tau^{n-2}$. If each of them behaves in a different way one another, then we say $\kappa = \tau_1 \tau_2 \cdots \tau_{n-2}$. where τ_i is the anonymising factor of the i-th principal between sender and receiver. When multiple principals transfer some degrees of knowledge to a principal, the maximum of all the knowledge transfer is the knowledgeability of the receiver; i.e., when τ_i is the degree of knowledge transfer of i-th principal, $\kappa = \max_i \tau_i$. Obviously, the transferred knowledge is determined by the receiver's viewpoint not by the sender's.

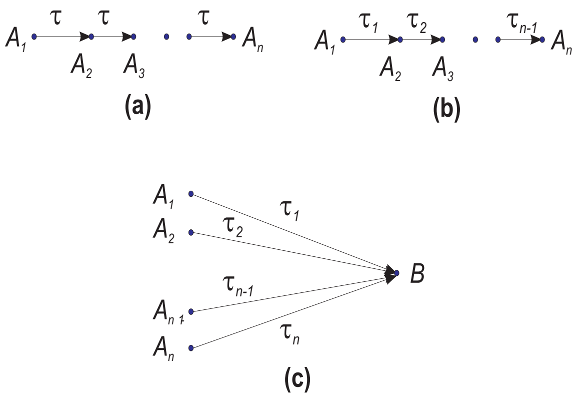

Fig. 1. (a) When all principals between sender and receiver do the same anonymising function, $\kappa = \tau^{n-2}$; (b) when each principal A_i does a different degree of transfer τ_i, $\kappa = \tau_1\tau_2\cdots\tau_{n-1}$; (c) when more than one principal transfer aknowledge to a principal, the maximum degree of transferred knowledge among them is κ.

The function τ depends on the construction of principals involved in anonymising. It shows how fast or slow the knowledge decays. Theoretically the knowledge transfer chain can obtain zero knowledge transfer, if there is at least one principal to transfer zero knowledge to the next.

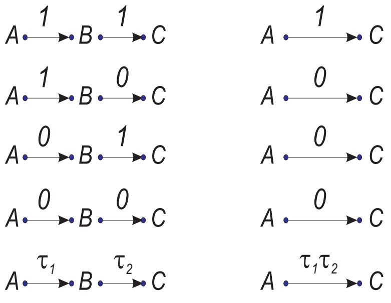

Fig. 2. If there is an all-or-none knowledge transfer chain, that is τ has a value either 0 or 1, it shows the same behaviour as that of the logical AND. In practice, κ has a value between 0 and 1 by multiplying τ_1 and τ_2 as shown in the last row.

The knowledgeability κ is mainly dependent on the factor τ, the degree of knowledge transfer. The next step we should follow is the construction of τ and provision of factors affecting the knowledge transfer.

3 Criteria for knowledge transfer

How do we measure the degree of knowledge transfer? For each principal, the degree of knowledge transfer τ should be defined to obtain the knowledgeability κ of the mechanism. We list a series of criteria to affect knowledge transfer and provide a rough measure for each criteria. These criteria are mutually independent and we may assume that each of them is described as a random variable with the exponential distribution $EXP(1)$. That is, each criteria Y can be represented by

$$g_Y(y) = \begin{cases} e^{-y} & y \geq 0; \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Assumption of trust ($A; \alpha$) If a trust relationship between principals to obtain anonymity is assumed, so is the possibility of principals' forgery. It makes the mechanism weaker and the anonymity of a step where such a forgery happens can be blown off. We define $\alpha = 1$ for steps assuming such trust. If there is a mechanism to prevent principals from the forgery, we define $\alpha = 0$. If there is only a detection mechanism, $\alpha = 0.5$.

Computational vs. Information-theoretic ($B; \beta$) To provide the protection for subjects' anonymity, some anonymising mechanisms relies on some computational complexity, we call computational anonymity. For example, Sako and Killian's mechanism [14] relies on the difficulty of computing discrete logarithms both for the secrecy of mixes' private keys and for the content of the ballots. To provide an information-theoretic anonymity, it is generally regarded that private channels between users and anonymising principals are required. We define $\beta = 0$ when the mechanism provides information-theoretic anonymity, and $\beta = 0.5$ when it provides computational anonymity. If there is no such mechanism, $\beta = 0$.

Collusion vulnerability ($C : \gamma$) Collusion of a group of principals can make some communication traced back. If a mechanism assumes the collusion vulnerability, we rate $\gamma = 1$. When there is a perfect anti-collusion mechanism, we define $\gamma = 0$. If there is an anti-collusion mechanism with some threashold, $\gamma = 0.5$.

Freshness ($D : \delta$) If it is possible for some principals to use stored information in order to trace some deed, the mechanism has high vulnerability. A mechanism making previous information or history obsolete is needed. We define $\delta = 0$ if there is such a mechanism. Otherwise, $\delta = 1$.

- **Order/delay vulnerability ($E : \epsilon$)** If an anonymising principal produces the result in the same order as the input or in a predictable time delay, attackers can guess which one is the result of which. If there is no mechanism to protect this, it increases the knowledgeability. ϵ is defined by 0 when there is such a mechanism against order and delay vulnerability, 1 when there is no such mechanims.
- **Key-coded items manipulation ($F : \zeta$)** Extracting information from a set of non-critical information to trace a data-subject is a strong threat in database security and inference attacks. For example, some countermeasures against such inference atatcks have been studies in medical privacy protection [8]. In this respect, anonymity is also affected by the method to manipulate key-coded items; the key-coded item is an item bridging non-critical items to reveal critical items. We define $\zeta = 0$ if there is a mechanism to preventing such inference attacks. Otherwise, $\zeta = 1$.

The degree of knowledge transfer τ is a function of above factors and we define τ by

$$\begin{aligned} \tau &= 1 - g_A(\alpha)g_B(\beta)g_C(\gamma)g_D(\delta)g_E(\epsilon)g_F(\zeta) \\ &= 1 - e^{-(\alpha+\beta+\gamma+\delta+\epsilon+\zeta)}, \end{aligned} \tag{3}$$

when none of random variable is zero, since the random variables are independent.

The degree of knowledge transfer represents the vulnerability of the anonymising principal involved. The larger the transferred knowledge is, the higher the vulnerability we get. It may be very rough but can be a measure to compare anonymising mechanisms.

There is a point to consider. The number of people who use the mechanism is also a critical factor; if there is only one person using the anonymising mechanism, it is clear that there is no anonymity however perfect methods are used in the anonymiser. Regardless of the performance of the anonymising mechanism, the result will be very poor. Obviously, the larger the domain is, the more anonymity we get. Since this factor is not an internal factor of the mechanism, we simply assume that we have enough number of users of the mechanism to get reasonable result.

4 Electronic voting

Since many voting scheme assumes secret ballot, anonymity of ballots has been an important issue; i.e., the ballot should not be traced back to the voter. In order to obtain such anonymity, two types of approaches have been tried [13]: one is using anonymous channels or mixers, the other is relying on number theoretic techniques. The former is good at efficiency and the latter secrecy. Depending on anonymising structure, the knowledgeability κ of the former is evaluated which is higher than 0, because mixers increase anonymity but never reach the perfect anonymity. Usually the schemes using anonymous channels or mixes rely on computational difficulties: factorisation [3] or discrete logarithm [14]. The recovery of the private key for the mixes reveals all ballots posted to the first mix.

The schemes using number theoretic techniques have a similar aspect. For example, the schemes presented by Benaloh *et al.* rely on r-th residuosity assumption [1, 2, 4]. When the public modulus is factorised, each ballot can be decrypted. It also relies on computational assumptions. Furthermore, this type of schemes does not consider the trace of voting, but the secrecy of the ballot. The anonymity of ballot relies entirely on the secrecy of it.

Let us consider anonymous channels. Anonymous channel is a multi-party protocol that changes the anonymity level of the original input. Each principal inputs a secret and he obtains a result with some anonymity at the end of the protocol. For example, a voter marks a ballot and put it into an anonymous channel. At the end of the channel, it is hard to trace the ballot back to the voter. It decays the knowledge of the voter-ballot relationship through the channel.

The mix-type channels are one of typical anonymous channels. Originally this idea is presented by Chaum [3] and has some descendants [7, 10, 14]. Mix is a shuffling machine agent and a certain number of mixes construct a network. Its anonymity is bounded by computational infeasibility. It requires a public channel to broadcast such as bulletin board or newsgroup. Attacks on this type of channels can be found in [9–12]

Let us evaluate the knowledgeability of Chaum's mix-net [3]. It has a series of redundant principals, called mixes, doing the same anonymisation with a different key for each principal. Each of them has the same degree of knowledge transfer. Let us calculate the degree τ. The mechanism assumes trust between principals, uses RSA encryption, uses a nonce for sealing each transfer, has a mechanism to shuffle all the inputs. It is vulnerable for the collusion of principals. Shuffling procedure is provided for re-ordering. Mix-net anonymising process implies the process to avoid inference attacks We have $\alpha = 1$, $\beta = 0.5$, $\gamma = 1$, $\delta = 0$, $\epsilon = 0$ and $\zeta = 0$. We then have

$$\tau = 1 - e^{-(\alpha+\beta+\gamma+\delta+\epsilon+\zeta)} = 1 - e^{-2.5} = 0.865 \qquad (4)$$

Hence there is 86.5% of vulnerability for each anonymising principal that he can transfers the knowledge of the information he has received. If there are n mixes, κ will be 0.865^n. When there are 10 mixes, they decreases the vulnerability to 23.5%.

Cramer *et al.*'s secret election scheme [5, 6] provides information theoretic anonymity. Although the influence of a forgery is minimised, some level of trust is assumed and there is also a possibility of collusion between principals. It has a mechanism for freshness. A role of bulletin board is shuffling received messages. The bulletin board is an apparatus against inference attacks. We then have $\alpha = 1$, $\beta = 0.5$, $\gamma = 1$, $\delta = 0$, $\epsilon = 0$, $\zeta = 0$, and

$$\tau = 1 - e^{-2.5} = 0.865.$$

Hence the vulnerability for each anonymising principal is 86.5%.

Digital cash mechanisms share many similarity with electronic voting schemes in view of anonymity [15]. The anonymity metric can be applied to digital cash mechanisms.

5 Concluding Remarks

We revisited anonymity and focused on the anonymity of the deed rather than that of the identity. Our emphasis is that we need to understand the anonymity as the degree of knowledge transfer. We also tried to quantise this qualitative property by defining some measures: knowledgeability and degree of knowledge transfer. As a result, we presented a metric to measure degrees of anonymity of a mechanism, called anonymity metric, and evaluated the anonymity degree of some voting schemes with this metric.

Currently, the metric is rough and evaluates only global features of mechanisms. Especially factors for knowledge transfer, i.e., α, β, γ, δ, ϵ, and ζ are defined in a loose manner. Provision of a more precise and refined metric will be a further study.

References

1. J. C. Benaloh and D. Tuinstra. Receipt-free secret-ballot elections. In *the 26th Annual ACM Symposium on Theory of Computing*, pages 544–553. Association for Computing Machinery, 1994.
2. J. C. Benaloh and M. Yung. Distributing the power of a government to enhance the privacy of voters. In *the 5th Annual ACM Symposium on the Principles of Distributed Computing*, pages 52–62. Association for Computing Machinery, 1986.

3. D. L. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 24(2):84–88, 1981.
4. J. D. Cohen and M. J. Fisher. A robust verifiable cryptographically secure election scheme. In *the 26th Annual Symposium on Foundations of Computer Science*, pages 372–382. IEEE, 1985.
5. R. J. F. Cramer, M. Franklin, B. Schoenmakers, and M. Yung. Multi-authority secret-ballot elections with linear work. In *the Proceedings of Advances in Cryptology - Eurocrypt '96*, pages 72–83. Springer-Verlag, 1996.
6. R. J. F. Cramer, R. Gennaro, and B. Schoenmakers. A secure and optimally efficient multi-authority election scheme. In *the Proceedings of Advances in Cryptology - Eurocrypt '97*, pages 103–118. Springer-Verlag, 1997.
7. A. Fujioka, T. Okamoto, and K. Ohta. Practical secret voting scheme for large scale elections. In *the Proceedings of Advances in Cryptology - Auscrypt '92*, pages 244–251. Springer-Verlag, 1992.
8. W. W. Lowrance. Privacy and health research. Report to the U. S. secretary of health and human services, U. S. Department of Health and Human Services, May 1997.
9. M. Michels and P. Horster. Cryptanalysis of a voting scheme. In *Communications and Multimedia Security II*, pages 53–59, Essen, 1996.
10. C. Park, K. Itoh, and K. Kurosawa. Efficient anonymous channel and all/nothing election scheme. In *the Proceedings of Advances in Cryptology - Eurocrypt '93*, pages 248–259. Springer-Verlag, 1993.
11. B. Pfitzmann. Breaking an efficient anonymous channel. In *the Proceedings of Advances in Cryptology - Eurocrypt '94*, pages 332–340. Springer-Verlag, 1994.
12. B. Pfitzmann and A. Pfitzmann. How to break the direct RSA-implementation of mixes. In *the Proceedings of Advances in Cryptology - Eurocrypt '89*, pages 373–381. Springer-Verlag, 1990.
13. K. Sako and J. Killian. Secure voting using partially compatible homomorphism. In *the Proceedings of Advances in Cryptology - CRYPTO '94*, pages 411–424. Springer-Verlag, 1994.
14. K. Sako and J. Killian. Receipt-free mix-type voting scheme. In *the Proceedings of Advances in Cryptology - Eurocrypt '95*, pages 393–403. Springer-Verlag, 1995.
15. D. R. Simon. Anonymous communication and anonymous cash. In *the Proceedings of Advances in Cryptology - CRYPTO '96*, pages 61–73. Springer-Verlag, 1996.