Privacy Computing: Concept, Computing Framework and Future Development Trends

Fenghua Li, *Member, IEEE,* Hui Li *, *Member, IEEE,* Ben Niu, *Member, IEEE,* and Jinjun Chen, *Senior Member, IEEE*

Abstract

With the rapid development of information technology and the continuous evolution of personalized services, huge amounts of data are accumulated by the large Internet company in the process of serving users. Moreover, dynamic data interactions increase the intentional/unintentional privacy persistence in different information systems. However, the following problems such as the short board effect of privacy information preservation among different information systems and the difficulty of tracing the source of privacy violations are becoming more and more serious. Therefore, existing privacy preserving schemes cannot provide a systematic preservation. In this paper, we pay attention to the links of information lifecycle, such as information collection, storage, processing, distribution and destruction. Then we propose the theory of privacy computing and the key technology system, including privacy computing framework, formal definition of privacy computing, four principles that should be followed in privacy computing, algorithm design criteria, evaluation of privacy preserving effect, privacy computing language and so on. Finally, we employ four application scenarios to describe the universal application of privacy computing and prospect of the future research trends. It is expected to guide the theoretical research on user's privacy preservation under open environments.

Index Terms

Privacy computing, privacy information description, privacy metric, evaluation of privacy preserving effect, privacy computing language

1 Introduction

In recent years, information technology and mobile communication technology are closely integrated and develop rapidly. The software and hardware of smart devices upgrade and evolute continuously. These have promoted the development of Internet, mobile Internet, cloud computing, big data and Internet of things. At the same time, a variety of new service models improve the quality of living greatly, for instance, e-commerce services represented by Amazon/Taobao, social network services represented by Facebook/Wechat, vehicle services represented by Uber/Didi.

However, the emergence and rapid development of new technology and new service mode lead to a normal situation in which massive user's personal information interacts across information systems, ecosystems and even national network boundaries. In each step of the whole information lifecycle, user's personal information can be inevitably retained in various information systems, such as collection, storage, processing, release (including exchange), destruction and so on. It leads to the separation of the ownership, the management and the utilization right of information, which seriously threatens users' rights to consent, to be erased/to be forgotten and to extend authorization. On the other hand, the lack of support of effective monitoring technology leads to the difficulty of tracing and forensics of privacy invasion.

Most of the existing privacy preserving schemes focus on relatively isolated application scenarios and technical points, and propose solutions to specific problems in a given application scenario. The privacy preserving scheme based on access control technology is suitable for single information system, but the privacy preserving problem in metadata storage and publishing is not solved. The privacy preserving scheme based on cryptography is also only applicable to a single information system. Although the implementation of key management with the help of trusted third parties can realize the exchange of privacy information between multi information systems, the deletion right/forgotten right and the extended authorization after the exchange are not solved. The privacy preserving scheme based on generalization, confusion, anonymity technologies fuzzes up data, making it impossible to be restored, and therefore can be applied to

- *F. Li and B. Niu are with Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China. F. Li is with School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China.*
  E-mail: lifenghua@iie.ac.cn, niuben@iie.ac.cn
- *H. Li is with State Key Laboratory of Integrated Services Networks, School of Cyber Engineering, Xidian University, Xi'an 710071, China.*
  E-mail: lihui@mail.xidian.edu.cn
- *J. Chen is with Department of Computer Science and Software Engineering, Swinburne University of Technology, Australia.*
  E-mail: jchen@swin.edu.au
- ** Corresponding Author*

many scenarios such as single anonymizing and multi-anonymizing with increased privacy preserving level. However, this kind of privacy preserving scheme reduces the utility of data, which leads to the adoption of the weaker privacy preserving schemes in the actual information system or the simultaneous storage of original data. At present, the description method and the computing model that can integrate the privacy information with the preservation demand are unavailable, and we lack the computing architecture which protects the privacy on demand in complex application scenarios, such as privacy information exchange across information system, privacy information sharing with multi service requirement and dynamic anonymizing of privacy information.

In brief, existing privacy preserving technologies can not meet the requirements of the privacy preservation in the complex information system, leading to unsolved privacy preserving problems in typical application scenarios such as e-commerce and social networking. Therefore, from the perspective of the whole lifecycle preservation on privacy information and to answer the demand of systematic privacy preservation in complex application scenarios, we put forward the privacy computing theory and the key technology system, including privacy computing framework, formal definition of privacy computing, four principles of privacy computing framework, algorithm design criteria, evaluation of privacy preserving effect, privacy computing language and so on in section 3. We look forward to the future research directions of privacy computing and unsolved problems in section 4. We conclude our paper in section 5.

2 Related Work

Existing researches on privacy preservation mainly focus on the privacy preserving techniques of information processing, the privacy measurement and evaluation.

2.1 The privacy preserving techniques of data processing

Researches on privacy preservation have been conducted at all stages including information collection, storage, process, release, destruction and so on. Meanwhile, based on access control, information confusion and cryptography technologies, plenty of privacy preserving schemes are proposed for typical scenarios such as social network, location-based services and clouding computing.

Access control technology protects the privacy information via making accessing strategies to ensure that only the authorized subject can access the data resource. Recent years, based on access control technology, multiple privacy preserving techniques have been presented. Scherzer et al. [1] proposed a high-assurance smart card privacy preserving scheme with mandatory access controls (MAC) [2], [3]. And Slamanig et al. [4] proposed a privacy preserving framework for outsourced data storage based on discretionary access control (DAC) [5], [6]. In order to improve the effectiveness of authority management, Ferraiolo et al. [7] presented role-based access control (RBAC). In RBAC, a user will be mapped to a specific role to obtain corresponding accessing authority, which simplified the authority management greatly in complicated scenarios. Anourd et al. [8] designed a privacy preserving data mining scheme with RBAC for multiple scenarios. In 2016, Li et al. [9] proposed a novel cyberspace-oriented access control model, which can effectively avoid security problems caused by the separation of data ownership and management rights and secondary/multiple forwarding of information by comprehensively considering vital factors, such as the access requesting entity, general time and state, access point, device, networks, resource, network interactive graph and chain of resource transmission. In addition, the attribute based encryption [10], [11] forms the identity of the user into a series of attributes, and the attribute information is embedded in the process of encryption and decryption so that the public key cryptosystem has the ability of fine-grained access control. Shao et al. [12] achieved fine-grained access control with attribute-based encryption, and protected user's location privacy in location-based services.

Information confusion technology protects the original data with generalization, anonymity, or confusion, which achieves the inability of attackers to obtain the useful information via the modified data. Anonymous technologies, such as k-anonymity [13], [14], l-diversity [15], [16] and t-closeness [17], [18], achieve the privacy preservation via masking the original data into a cloaking spatial. Differential privacy [19], [20] is widely considered as a privacy preserving technology since it does not require background knowledge of attackers. Aiming at the issue of similarity attacks, Dewri et al. [21] proposed an anonymous algorithm via applying the differential privacy technology in location-related data, which is able to maximum the effectiveness of differential privacy. However, differential privacy needs to add a lot of randomization to query results, and the utility drastically decreases with increasing privacy preservation requirements [22].

Cryptograph technology protects users' privacy information via encryption techniques and trapdoor functions. In order to protect the privacy data in cloud computing, the concept of homomorphic encryption [23] was first proposed by Rivest et al. With homomorphic encryption, Zhu et al. [24] proposed a privacy preserving spatial query framework for location-based services. In 1999, based on composite residuosity, Paillier [25] designed an additive homomorphic encryption algorithm, which is widely used in multiple scenarios. For smart grid, Lu et al. [26] proposed a privacy preserving data aggregation scheme with paillier cryptosystem, which can protect the users' sensitive information and resist various attacks. In 2009, Gentry [27] successfully constructed the fully homomorphic encryption (FHE) algorithm based on ideal lattice [28], which achieves the additive and multiplicative homomorphic simultaneously. However, the efficiency of FHE is far from practical in real world, even though there are many modified schemes [29], [30], [31] proposed in recent years. In order to improve the efficiency, Zhu et al. [32] proposed efficient and privacy preserving POI query [33] scheme with a lightweight cosine similarity computing protocol for location-based services, the proposed scheme is highly efficient and can protect users' query and location information simultaneously.

The above-mentioned privacy preserving schemes are concrete algorithms which mainly focus on partial dataset of specific scenarios. As a result, they lack the algorithm framework for dynamic dataset of specific scenarios and further lack the universal algorithm framework for dynamic dataset of multiple scenarios. Moreover, aiming at multimedia data, it is required to combine multiple algorithms to achieve privacy preserving. The mature schemes in this area are insufficient. Finally, superimposing different privacy preserving algorithms on each other to obtain better preservation quality is also need to be further researched.

2.2 The privacy measurement and evaluation

Specific researches now are focusing on the field of information theory and applications. Oya et al. [34] proposed a scheme through using the conditional entropy and the mutual information as complementary privacy metrics. Ma et al. [35] proposed a privacy metric for time-series data to quantify the amount of released information obtained by attackers. Cuff et al. [36] used mutual information to describe the information obtained by attackers from observing data, and they measured the decrease of uncertainty of the original data. Jorgensen et al. [37] combined controllable character of ϵ in differential privacy, and generated noise calibrated to $Lap(\frac{\Delta f}{\epsilon})$ based on the privacy demands of user, when ϵ gets smaller, the added noise gets more, then the intensity of privacy protection is higher. Asoodeh et al. [38] depicted the risk of privacy leakage with mutual information. They calculated the decrease of the uncertainty of privacy information in original data during the data releasing. Furthermore, the researches on application fields mainly aim at social network, location-based service, cloud computing and so on.

In the field of social network, aiming at the webpages searching, Getvais et al. [39] proposed a privacy preserving scheme based on obfuscating technique and quantified the users' privacy. Considering the different searching behaviors of users with various intentions, they designed a commonly used tool to measure the privacy preserving scheme based on obstructing technique. Aiming at spatio-temporal connection, Cao et al. [40] used calculation to analyze the data and quantified the potential risks under differential privacy technique through formal description of privacy.

In the field of location-based service, based on identifying the attacking model and adversaries' background knowledge, Shokri et al. [41] used information entropy to describe the precision, certainty and validity for measuring the effectiveness of privacy preservation. Meanwhile, based on Bayesian Stackelberg model of game theory [42], the user in this model acts as a leader, and the attacker acts as a follower, to form the game theory model. Kiekintveld et al. proposed a framework [43] to find the optimal privacy mechanism that is able to resist the strongest inference attack.

In the field of cloud computing, as a service-oriented privacy preserving framework, SAFE [44] implemented secure coordination for cross-neighbour interaction between protocol and itself in cloud computing. Based on game theory and differential privacy, Wu et al. [45] quantified the game elements that have been come down by the users. They also implemented users' privacy measurement by analyzing a single dataset. The work [46] used definition of differentiation to quantify the level of privacy of participating users, and then to implement accurate incentive mechanism.

Most above-mentioned schemes lack unified definition of privacy concept. Moreover, the privacy metric varies dynamically with the subject which receives information, the size of the data quantity and the scenarios, but nowadays the dynamic privacy metric method is lacked. Finally, the dissemination of information is cross information system, but the above schemes is short of consistency among different information systems and formalized description method for dynamic privacy quantity. Therefore, they are far from satisfying the dynamic requirements of privacy preservation of cross-platform privacy information exchange, extended authorization and so on.

In summary, existing privacy preserving technologies and privacy measurement methods are fragmented, and they also lack formalized description method for auditing of privacy information and constraint conditions. However, the scheme, which integrates the privacy preserving with tracking and forensic of privacy infringement, has not been considered yet. Meanwhile, it is hard to construct a uniform information system covering all stages of information collection, storage, process, release, destruction and so on.

3 Definition and Framework of Privacy Computing

3.1 Concepts of Privacy and Privacy Computing

3.1.1 Privacy Right and Privacy Information

The legal definition of privacy concentrates on protecting individual's rights given by laws, including the requirement that personal information, activities and spaces can not be published, interfered or intruded illegally. It emphasizes privacy's independence of public interests and group interests, including personal information that a person does not want others to know, personal affairs that a person does not want others to touch and personal area that a person does not want others to invade. The essence of legal definition is actually privacy rights.

This paper focuses on full lifecycle preservation of privacy information. Specifically, privacy information includes personal information that a person does not want others to know or is inappropriate for others to know, or that a person wants to be disseminated within an approved circle of personnel in the way he/she agrees with. Privacy information can be used to deduce the user's profile, which may impact his/her daily life and normal work.

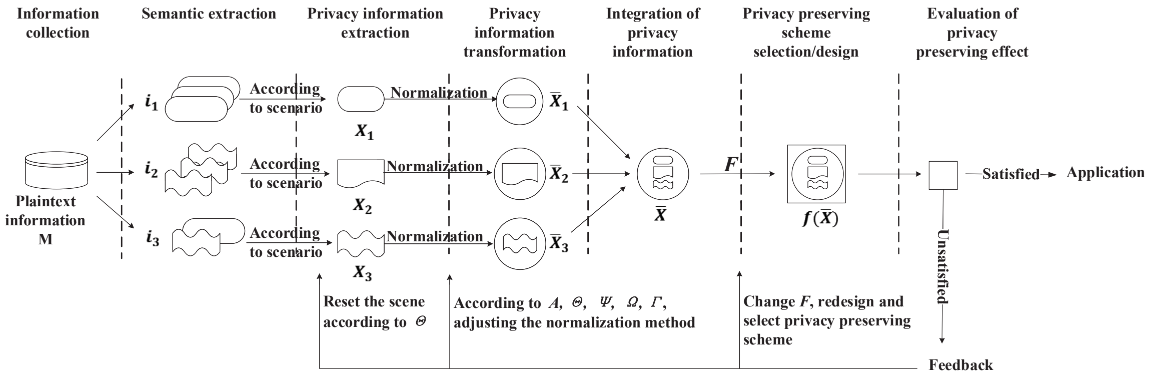

Information collection
Semantic extraction
Privacy information extraction
Privacy information transformation
Integration of privacy information
Privacy preserving scheme selection/design
Evaluation of privacy preserving effect
Plaintext information M
i_1
i_2
i_3
According to scenario
According to scenario
According to scenario
X_1
X_2
X_3
Normalization
Normalization
Normalization
$\bar{X}_1$
$\bar{X}_2$
$\bar{X}_3$
$\bar{X}$
F
$f(\bar{X})$
Satisfied
Application
Unsatisfied
Feedback
Reset the scene according to Θ
According to A, Θ, Ψ, Ω, Γ, adjusting the normalization method
Change F, redesign and select privacy preserving scheme

Fig. 1. Privacy computing framework

Academically speaking, privacy information is closely related to spatiotemporal scenario and the cognitive ability of the subject. It shows dynamic perceptual results. Different from the definition of privacy in law, we mainly define and describe privacy information technically so that it can support research on various technical aspects such as semantic understanding of privacy, extraction, design of privacy preserving algorithms, evaluation of privacy preserving effect and etc.

3.1.2 Privacy Computing

In general, Privacy Computing refers to a computing theory and methodology, which can support the describing, measuring, evaluating and integrating operations on privacy information during the processing of video, audio, image, graph, numeric values and behavior information flow in pervasive network. It has a set of symbolized and formulized privacy computing theories, algorithms and application technologies with quantitative assessment standards and support on the integration of multiple information systems.

Privacy computing includes all computing operations by information owners, collectors, publishers and users during the entire lifecycle of privacy information from data generation, sensing, publishing, dissemination, storage, processing, usage and destruction. It supports privacy preservation with massive number of users, high concurrency and high efficiency. In ubiquitous network, privacy computing provides an important theoretical foundation for privacy preservation.

From the perspective of full lifecycle privacy preservation, we build a framework of privacy computing. Based on constructed six tuples of privacy information for any format of plaintext information M, we implement privacy computing through 5 steps, including privacy information extraction, scenario abstraction, privacy operation selection, privacy preserving scheme selection or design and evaluation of privacy preserving effect. Its process is depicted in figure 1:

(1) According to the format and semantics of plaintext information M, extract privacy information X and obtain privacy information vector $\boldsymbol{I}$.

(2) According to the type and semantics of each i_k in $\boldsymbol{I}$, define and abstract application scenario.

(3) According to privacy operations supported by each i_k, select and generate dissemination control operation set.

(4) According to application requirements, select or design proper privacy preserving scheme. If there are available and capable schemes, they can be selected directly. Otherwise, design new schemes.

(5) According to relevant assessment criteria, assess the preservation effectiveness of selected privacy preserving scheme.

If the evaluation result of privacy preservation does not meet expected requirements, the feedback mechanism will be executed. It consists of three situations: 1) if the application scenario is mis-abstracted, it should be re-abstracted iteratively; 2) if the application scenario is abstracted properly but privacy operation is selected improperly, privacy operation should be re-organized; 3) if the application scenario and privacy operation are selected correctly, privacy preserving scheme should be adjusted or improved to eventually achieve a satisfactory effectiveness of privacy preservation.

3.2 Formalization of Privacy Computing

In this section, we first define the privacy information X and its six basic components, related axioms, theorems and assumptions, which provide the foundation to describe the other part of privacy computing. It is noted that the extraction methods of privacy information vector of any information M are outside the scope of this paper, which is subject to domain specific extraction conditions. Moreover, the quantification of information privacy content is also outside the scope of this paper, which is the task of programmer or modeler of information systems.

Definition 1: Privacy information X consists of six components $\langle \boldsymbol{I}, \boldsymbol{A}, \Gamma, \Omega, \Theta, \Psi \rangle$, which are privacy information vector, privacy attribute vector, location information set, audit control information set, constraint condition set and dissemination control operation set respectively.

Definition 2: Privacy information vector $\boldsymbol{I} = (I_{ID}, i_1, i_2, \cdots, i_k, \cdots, i_n)$, where i_k $(1 \leq k \leq n)$ is a privacy information element. Each i_k represents semantically informational and indivisible atomic information. The information types include text, audio, video, images, etc., and these types' combination. The semantic characteristics include words, phrases, tone of voice, pitch of tone, phoneme, sound, frame, pixels, color, etc., and their combination. It is used to represent atomic information that is semantically informative, indivisible and mutually disjoint in information M. I_{ID} is the unique identifier of privacy information vector that is independent of privacy information elements. For example, in the text "U_1 and U_2 went to Loc to drink beer", the privacy information vector is $I = (I_{ID}, i_1, i_2, i_3, i_4, i_5, i_6, i_7)$ = (I_{ID}, U_1, and, U_2, went to, Loc, to drink, beer). In this case, $n = 7$.

Axiom 1: Under a natural language and its grammar rules, and in the granularity of words, phrases and slangs, the number of elements of privacy information vector $\boldsymbol{I}$ is bounded.

Property 1: Privacy information vector conforms the first normal form (1NF) and the second normal form (2NF).

Privacy information component i_k is defined as the smallest granularity that can not been divided further, which is called atomic property. First normal form (1NF) is a property of a relation in a relational database. A relation R is in first normal form if and only if the domain of each attribute contains only atomic values, and each attribute can only have a single value from that domain. Under this definition, i_k conforms 1NF. Meanwhile, privacy information vector $\boldsymbol{I}$ has unique identification I_{ID} as primary key. Other non-prime attributes are all dependent on this primary key. A relation R is in 2NF if $R \in$ 1NF and every non-prime attribute of the relation is dependent on the unique primary key. Therefore, i_k conforms 2NF.

Definition 3: Constraint condition set is denoted by $\Theta = \{\theta_1, \theta_2, \cdots, \theta_n\}$, where θ_k $(1 \leq k \leq n)$ is a constraint condition vector corresponding to privacy information component i_k. θ_k is to describe permissions for an entity to access i_k in different scenarios, such as who, at what time, using what devices, by what means access and use the privacy attribute component i_k, and the duration of usage of privacy information vector. Only entities which satisfies constraint condition vector θ_k can access privacy information component i_k. An entity can be an owner, a receiver or a publisher of the information, etc.

Definition 4: Privacy attribute vector $\boldsymbol{A} = (a_1, a_2, \cdots, a_k, \cdots, a_n, a_{n+1}, \cdots, a_{n+m})$, a_k denotes privacy attribute component and is used to measure the degree of privacy information preservation. In practical applications, different privacy information components are able to form weighted dynamic combinations of different scenarios. These combinations will produce new privacy information. However, based on the atomicity of privacy information components, we represent the privacy information preservation degree of different combination of i_k with privacy attribute component. When $1 \leq k \leq n$, there is a one-to-one correspondence between a_k and i_k; when $n < k \leq m$, a_k represents the privacy information preservation degree of two or more privacy information components' combination.

We set $a_k \in [0, 1]$, and $a_k = 0$ represents that privacy information component i_k has the highest degree of preservation. Under this condition, information M has no sharing, that is to say, there is no possibility of any leakage, meaning that the information is protected to the highest degree. In that case, the mutual information between the protected privacy information and the original privacy information is 0. For example, in cryptography-based privacy preserving methods, $a_k = 0$ means secret key has lost and the information could not be reversed; In the cases of applying noise injection, anonymization or other irreversible techniques, $a_k = 0$ represents that the degree of distortion of the data has led to a complete irrelevancy between the processed information and the initial information. $a_k = 1$ represents that i_k is not protected and could be published freely without any limit. The other values between 0 and 1 represent different degrees of privacy information preservation. The lower the value is, the higher the degree of privacy information preservation is.

The privacy preserving quantitative operation function is denoted by σ, and it can be a manually labelled function, a weighting function, etc. Since different types of privacy information i_k correspond to different kinds of operation functions, the resulting privacy attribute components are also different, expressed by $a_k = \sigma_k(i_k, \theta_k)$ $(1 \leq k \leq n)$. For any combination of privacy information components $i_1, i_2, \cdots, i_n$, we denote it as $i_{n+j} = i_{k_1} \vee i_{k_2} \vee \cdots \vee i_{k_s}$, where $\vee$ stands for the combination operation of privacy information components. Given privacy preserving quantitative operation function σ and privacy attribute component a_{n+j}, we have $a_{n+j} = \sigma(i_{n+j}, \theta_{k_1}, \theta_{k_2}, \cdots, \theta_{k_s})$ $(1 \leq k_1 < \cdots < k_s \leq n)$. Privacy attribute vector $\boldsymbol{A} = (a_1, a_2, \cdots, a_k, \cdots, a_n, a_{n+1}, \cdots, a_m)$ is generated by privacy information components $i_1, i_2, \cdots, i_n$ and their combination vectors $i_{n+1}, i_{n+2}, \cdots, i_m$. The relationship between privacy information vector and privacy attribute vector can be denoted by $\boldsymbol{A} = \sigma(\boldsymbol{I}, \Theta)$. As quantitative operation and constraints go hand by hand, the results of quantitative operation vary with different scenarios and entities.

Theorem 1: For a specific privacy information vector $\boldsymbol{I} = (i_1, i_2, \cdots, i_n)$, if the number of its components is bounded, the dimensionality of its corresponding privacy attribute vector $\boldsymbol{A} = (a_1, a_2, \cdots, a_n, a_{n+1}, \cdots, a_m)$ is bounded. When each binary or multiple combination of the components of $\boldsymbol{I}$ only corresponds to one privacy attribute component, the number of privacy attribute components m satisfies $m \leq 2^n - 1$.

Proof: According to definition/axiom 1, given privacy information vector $\boldsymbol{I}$, its dimension is limited and denoted by n. According to the definition of privacy attribute vector, its privacy attribute components corresponds to privacy information components and their combination vectors, thus the size of privacy attribute vector is limited. When each combination of privacy information components corresponds with one privacy attribute, the maximum size of privacy attribute vectors is the number of all the combinations of privacy information components, including 2 to n-size combination, denoting as $C_n^1 + C_n^2 + \cdots + C_n^n = 2^n - 1$, hence the inequality $n + m \leq 2^n - 1$ holds.

Definition 5: Location information set $\Gamma = (\gamma_1, \gamma_2, \cdots, \gamma_n)$, γ_k denotes the location information vector, which stands for the location information and attribute information of i_k in information M. Using γ_k, privacy information component i_k could be quickly positioned. Meanwhile, the location information describes i_k's specific location in M. For example: page numbers, chapters, paragraphs, serial numbers, coordinates, frame numbers, time periods, audio tracks, layers, pixels, etc. In a text file, location information mainly includes the page number, sections, paragraphs, serial number, etc. Its attribute information mainly includes the font, font size, thickness, italic, underline, strikeout, superscript, subscript, style, line spacing, etc. Attribute information includes font, size, font-weight, line spacing, pixel, color, brightness, tone, intonation and etc.

Definition 6: Audit control information set $\Omega = (\omega_1, \omega_2, \cdots, \omega_n)$, ω_k denotes a specific audit control information vector during the propagation process of i_k. It is to record subjective and objective information, such as information owner, information sender, information receiver, information sending device, information receiving device, information transmission pattern, information transmission channel as well as operations performed on them during the transfer process. The operations including copy, paste, cut, forward, modify, delete and etc. If the privacy information is revealed, the source of the leakage point could be tracked.

Definition 7: Dissemination control operation set $\Psi = (\psi_1, \psi_2, \cdots, \psi_m)$, ψ_k denotes dissemination control operation vector. It is to describe the operations which can be performed on i_k and their combinations, such as copy, paste, cut, forward, modify, delete and etc. Meanwhile, these operations will not break the atomicity of $\boldsymbol{I}$. We have $\psi_l = judg(a_l, \theta_l)$, where constraint condition vector $\theta_l = \theta_{k_1} \vee \theta_{k_2} \vee \cdots \vee \theta_{k_s} (n+1 \leq l \leq m)$, and $judg$ is the operation judgement function including artificial markers, weighting function and etc.

Axiom 2: During the cross information system exchange process, if both of the two information control sides that extend authorization can not perform the exchange completely and effectively, it must cause the leakage of privacy information.

Assumption 1: Privacy computing can be defined as finite atomic operations. The other operations are combinations of these finite atomic operations.

Assumption 2: Privacy computing is established under the condition where the number of privacy information components is finite.

3.3 Four Principles for Privacy Computing

The four principles of privacy computing:

(1) Atomicity: The privacy information components are independent of each other and can be divided to minimum granularity and can not be divided further.

(2) Consistency: For the same privacy data, various privacy preserving algorithms all aim to make all the components of privacy attribute vector A approach 0. Even though they have different preserving degrees, they have similar aims.

(3) Sequence: In a privacy preserving algorithm, different orders of some operations may lead to different preservation effectiveness.

(4) Reversibility: Some privacy preserving algorithms can be recovered reversibly, such as encryption-based ones by decryption. However, some others are irreversible on privacy information processing.

3.4 The Characterization Elements of Privacy Computing

Definition 8: Privacy computing spans four factors (X, F, C, Q), where X represents privacy information (the detail illustrated in **Definition 1**). F represents for the privacy computation operation set, C represents for the cost of privacy preservation, and Q represents for the effectiveness of privacy preservation.

Definition 9: Privacy computation operation set $F = \{f_1, f_2, \cdots\}$, F is the set of atomic operations such as modular addition, modular multiplication, modular exponentiation, insert, delete and other operations on X. A privacy preserving algorithm is composed of multiple elements in the collection of privacy operations, and each element can be used for multiple times.

Privacy perception, privacy preservation, privacy analysis, exchange of privacy information and second transmission, integration of privacy information, update of privacy information and etc., which are defined as the specific operations consisting of several atomic operational combination.

Axiom 3: After privacy operations on information M, the change of privacy information vector from $\boldsymbol{I}$ to $\boldsymbol{I'}$ is triggered, and then further change the privacy attribute vector from $\boldsymbol{A}$ to $\boldsymbol{A'}$. The number and value of component a'_i will also be changed. In brief, when $\boldsymbol{I}$ conducts privacy operation f, $\boldsymbol{I'} = f(\boldsymbol{I})$ is achieved, and $\boldsymbol{A} \neq \boldsymbol{A'}$ where $\boldsymbol{A} = \sigma(\boldsymbol{I})$ and $\boldsymbol{A'} = \sigma(\boldsymbol{I'}) = \sigma(f(\boldsymbol{I}))$.

Definition 10: Cost of privacy preservation C represents the quantification of various resources required for achieving certain level of privacy preservation on information M, including computation, storage, network transmission cost and computational complexity. Each privacy information component i_k corresponds to the cost of privacy preservation C_k. The parameter C_k is related to privacy information component i_k, constraint condition vector θ_k and privacy computing operation vector f_k. It can be described as follows:

$$C_k = c_k(i_k, \theta_k, f_k). \tag{1}$$

Each i_k may have different types of information. For instance, a Word file contains characters and images, even audios. Hence, the corresponding function c_k of parameter i_k has different expressions, depending on the type of information. The parameter C can be described by the vector $\{c_k\}$ $(1 \leq k \leq m)$.

Definition 11: The effectiveness of privacy preservation Q represents the level of privacy preservation on information M, namely, the difference between privacy metric before and after privacy preservation. Generally, we need to consider privacy information vectors of information M, information access entities (including information owners, information receivers, information publishers, participants of information creation and transfer process), constraint conditions, privacy computing operation and other elements. In previous sections, we have introduced privacy metric, namely, the expression of privacy attribute component, $a_k = \sigma(i_k, \theta)$, where function σ has contained the vector of privacy computing operation. The definition of constraint conditions also covers the factors of information access entities. Therefore, the effectiveness of privacy preservation Q_k corresponding to privacy information component i_k can be expressed as follows:

$$\begin{aligned} Q_k &= \Delta\sigma_k(i_k, \theta_k) \\ &= \sigma_{k(before)}(i_k, \theta_k) - \sigma_{k(after)}(i_k, \theta_k), \end{aligned}$$

where $\sigma_{k(before)}$ is a privacy metric function before privacy preservation and $\sigma_{k(after)}$ is a privacy metric function after privacy preservation.

Definition 12: Profit and loss ratio of privacy disclosure L represents the ratio between profit and loss after privacy disclosure. The relationship between L, cost of privacy preservation C and effectiveness of privacy preservation Q can be described as follows:

$$L_k = l_k(C_k, Q_k). \tag{2}$$

The core idea of privacy computing model is to describe the relationships among the four factors of privacy computing and the profit and loss ratio of privacy disclosure L.

3.5 Evaluation Method of Privacy Preservation

Definition 13: Privacy preserving algorithm or scheme f is the combined operation f_i on the elements in privacy computing operation set F. After the combined operation f on privacy information vector $\boldsymbol{I}$, each component in corresponding privacy attribute vector $\boldsymbol{A}$ approaches 0. In brief, for vector $\boldsymbol{I}$, $\boldsymbol{A}$ where $\boldsymbol{A} = \sigma(\boldsymbol{I})$, if $f \in F^k$, $\boldsymbol{I'} = f(\boldsymbol{I})$, $\boldsymbol{A'} = \sigma(\boldsymbol{I'})$, $s.t. ||\boldsymbol{A'}|| < ||\boldsymbol{A}||$, then f is called a privacy preserving algorithm, where $|| \cdot ||$ represents a kind of measuring method of vector $\boldsymbol{A}$, such as L_2 norm.

Definition 14: Evaluation of privacy preserving effect means the evaluation on privacy attribute vector of the new privacy information vector $\boldsymbol{I'}$ after different privacy preserving operations f on $\boldsymbol{I}$. In brief, the closer to 0 the value of $\sigma(f(\boldsymbol{I}))$ is, the better the effectiveness of privacy preserving algorithm is.

Axiom 4: The effectiveness of privacy preservation is measurable.

The effectiveness evaluation mainly includes the utility of privacy information after preservation, the irreversibility of privacy preservation and the reversibility in controlled environments. The utility of privacy information refers to the impacts the new information has on the information system function or performance after the execution of privacy preserving algorithm. The irreversibility of privacy preservation means that any third party or attacker can not deduce the original privacy information from privacy preserving algorithms and obtained information. In a controlled environment, the reversibility means that third parties can restore the whole of information based on partially known information. As such, this paper generalizes the evaluation of privacy preserving effect into five indicators.

3.5.1 Reversibility

Reversibility refers to the ability of restoring privacy information after the execution of privacy preserving algorithm. Specifically, reversibility is the ability of attacker or the third party to deduce privacy information component i_k from the observed privacy information component i'_k. If i_k can be inferred accurately then it is reversible, otherwise it is not.

For example, when some data needs to be published, we first assess the attack resistance ability of the selected privacy preserving scheme under different attacks. Then, based on the data after the execution of privacy preserving algorithm, we compute privacy attribute vector. Further, we figure out the restoring degree of unauthorized information and authorized information under different attacks.

Conjecture 1: If privacy preserving policies do not match with each other, then a reversible privacy preserving algorithm may lead to privacy leakage after the privacy information is disseminated across different trustable domains.

3.5.2 Extended Controllability

Extended controllability refers to the matching degree between the receiver's effectiveness of privacy preservation and the sender's requirements of privacy preservation during the cross information system exchange process. Specifically, it means the dissimilarity between the privacy attribute component a_k in the information system Sys_1 and the privacy attribute component a'_k in the information system Sys_2 when privacy information X is transferred from Sys_1 to Sys_2. In brief, for any values of parameter k, $a_k = a'_k$ in different information systems means that the extended control is well maintained. Otherwise, the extension of authorization is deviated. For instance, user *Alice*, *Bob*, *Charles* are friends. *Alice* publishes a privacy information in WeChat and sets up a sharing list which allows *Bob* to access this information but prohibits *Charles* is prohibited. However, user *Bob* transfers this information to Weibo without any access restrictions. In this situation, *Charles* could see that information, and *Alice*'s access privilege on the same information in Weibo and in WeChat does not match with each other.

3.5.3 Deviation

Deviation refers to the dissimilarity between privacy information component i_k and the observed privacy information component i'_k by attackers or third parties after the execution of privacy preserving algorithms. For instance, in location privacy, the physical distance between mobile user's real location (m, n) and the processed location $(m'$,$n')$ obtained by location-based privacy preserving schemes can be calculated as $\sqrt{(m-m')^2+(n-n')^2}$.

3.5.4 Complexity

Complexity refers to the required cost of performing privacy preserving algorithm, which is similar to cost of privacy preservation C. For example, a user uses a handheld terminal to execute a 2048-bit RSA encryption algorithm. The calculation resource cost of this process is greater than that of executing AES algorithm once.

3.5.5 Information Loss

Information loss refers to the loss of information utility after the information is processed by irreversible privacy preserving algorithm, such as information confusion or information obfuscation.

For location privacy, if a mobile user submits his real location to server without k-anonymity process, he can receive accurate service information. If he employs k-anonymity to process locations, he will receive coarse-grained service information and the proportion of unavailable results is increased. This results in certain loss of information availability.

3.6 Design Principles for Privacy Preserving Algorithms

Although requirements of privacy preservation for different scenarios and information categories vary greatly, some common criteria exist during the design of privacy preserving algorithms. According to the concept of privacy computing, we summarize five basic criteria for the design of privacy preserving algorithms.

Criterion 1: Pre-processing. Firstly, we need to pre-process the privacy information X to determine the data distribution character, its value range, privacy preserving sensitivity, the expected value of privacy preserving operations, social experience points, etc. For example, the expected value of privacy preserving operations can be denoted as $times = f(\boldsymbol{I}, \boldsymbol{A}, \Theta)$.

Criterion 2: Algorithmic framework. Based on the scenarios and information categories, the mathematics foundation of privacy preserving algorithm can be determined, including the procedures and their combination relationships, as well as the relationship between privacy attribute vector and privacy information vector. For example, in the scenario that irreversible operations for privacy preservation are allowed, techniques based on generalization, obfuscation, anonymity, differential privacy, etc., can be employed. Take differential privacy for example, the specific mechanism of noise addition should be determined by following the guidance of Criterion 1 and considering elements including $\boldsymbol{I}, \boldsymbol{A}, \Theta,\ C,\ Q$ and L.

Criterion 3: Design of algorithm parameter. According to requirements of privacy preservation effect and usability, the relevant parameters of privacy preserving algorithm can be determined with the consideration of criterion 1 and 2. For example, the expected times of privacy operations should be determined based on requirements of privacy preservation in differential privacy mechanism. Besides, the sensitivity and social experience value of privacy operation results should also be determined upon the query function. Then we can determine the specific distribution of noise by combining $\boldsymbol{I},\ \Theta$ under the guidance of criterion 2.

Criterion 4: Combination of algorithm. To improve the security property and algorithm performances, we combine different procedures within a particular algorithm or between similar algorithms based on application scenarios and information characteristics. Take differential privacy as an example, we achieve the flexible combination of different procedures in one algorithm by considering factors such as $\boldsymbol{I},\ \Theta$ and some composition properties of differential privacy including post-processing, sequential composition, and parallel composition properties; in case of complex requirements of privacy preservation, such as the scenario which publishes data while emphasizing statistical characteristics and anonymity simultaneously, we need to consider characteristics of different algorithms with similar mathematical mechanisms, and organically integrate such algorithms to satisfy requirements of privacy preservation during the privacy information processing. Through this way, the security property and algorithm performance can be entirely improved.

Criterion 5: Analysis of algorithm complexity and efficiency. In order to evaluate whether the selected algorithm adapts to the corresponding scenario, we need to comprehensively analyze and evaluate the implementation cost of the privacy preserving algorithm with the consideration of factors such as the number of privacy information components which needs to be protected, the value range of security parameters, time and space complexity, and the expectation of effectiveness of privacy preservation.

In the following, we explain the applicability of the above-mentioned criteria with differential privacy mechanism.

(1) Pre-processing: in differential privacy algorithms, denote the dataset as X. With X, constraint condition set Θ and propagation control operation set Ψ, the privacy information vector set $\boldsymbol{I} = i(\boldsymbol{X}, \Theta, \Psi)$ can be generated. And through analyzing the distribution characteristic of $\boldsymbol{I}$: $\Phi{=}\phi(I)$, we can determine the value range of $\boldsymbol{I}$ or the value set Ran. Then, based on the statistical query function $g(\cdot)$ which is defined over $\boldsymbol{I}$, we can determine the expected value of query numbers $t(\cdot)$ and the social experience value of query results $v(\cdot)$, and obtain the noise adding's value space or value set $S = s(\Phi, Ran, g(\cdot), t(\cdot))$, and compute the sensitivity of query function $g(\cdot)$. For a statistical function $g(\cdot)$ which is defined on the subset D of $\boldsymbol{I}$, its sensitivity can be described as follows:

$$\Delta g = \max \|g(D_1) - g(D_2)\|_p,$$

where $D_1, D_2 \subseteq \boldsymbol{I}$ are two arbitrary datasets. When the difference between D_1 and D_2 is up to one element, we call them neighboring sets. Moreover, $p \geq 1$, and $p \in \mathbb{N}$.

(2) Algorithmic framework: based on the result of pre-processing, the mathematical definition of differential privacy mechanism can be represented as follows with fully considering the cost of privacy preservation C, the effectiveness of privacy preservation Q and so on.

$$\Pr[Alg(D_1) \in S] \leq h(\cdot) \times \Pr[Alg(D_2) \in S] + \delta(\cdot),$$

$h(\cdot) = h(\lambda, \epsilon, \kappa)$ presents the extended privacy estimate, where λ is a constant number that is related to noise distribution; ϵ is related to the expected value of query numbers, and κ is related to the social experience value of query results. In addition, $\delta(\cdot) = \delta(\epsilon, \kappa)$ is the correction parameter, which is used to soften conditions to make algorithms satisfy the definition of differential privacy. Furthermore, D_1 and D_2 is a couple of neighboring sets, Alg is a randomized algorithm.

Then, the framework of differential privacy can be described as follows:

$$\begin{array}{ll} While & Alg(g) \notin v(\cdot) \\ do & \{A\lg(g) = g(D) + \mathbb{N}\mathbb{o}\mathbb{i}\mathbb{s}\mathbb{e}(\mu(\cdot), b(\cdot), q(\cdot))\}, \end{array}$$

where $\mathbb{N}\mathbb{o}\mathbb{i}\mathbb{s}\mathbb{e}(\cdot)$ is the noise function set, which generates noise satisfying the $(h(\cdot), \delta(\cdot)) - DP$ condition, $\mu(\cdot)$ is the expected value of generated noise, $b(\cdot)$ is the scale parameter function used in controlling the range of operating distribution, and $q(\cdot)$ is the utility function which controls the probability expectation of a certain result being generated with the noise-processed data. In practice, the distribution of noise and the parameters of the algorithm should be selected according to application scenarios and information categories.

(3) Design of algorithm parameter: based on the users' requirements of privacy preserving strength and usability and in combination with the value range Ran of privacy information vector $\boldsymbol{I}$, the expected value of query numbers $t(\cdot)$, etc, we can determine the the specific parameters of noise distribution. Specifically, μ is related with the mean demand of outputs, and since $b(\cdot)$ is related with $h(\cdot)$, the sensitivity of dataset Δg, the value space or value set S of noise adding and so on, we can infer that $b(\cdot) = b(h(\cdot), \Delta g, S)$. Moreover, $q(\cdot)$ is related with the social experience value of querying result from S, therefore, $q(\cdot) = q(S, v(\cdot))$.

(4) Combination of algorithm: differential privacy mechanism has the following features:

Post-processing property. If $Alg_1(\cdot)$ satisfies $\varepsilon - DP$, the combined algorithm $Alg_2(Alg_1(\cdot))$ also satisfies $\varepsilon - DP$, where $Alg_2(\cdot)$ is an arbitrary, including randomized algorithm.

Sequential composition. If $Alg_1(\cdot)$ satisfies $\varepsilon_1 - DP$, and for arbitrary s, $Alg_2(s, \cdot)$ satisfies $\varepsilon_2 - DP$. Then the combined algorithm $Alg(D) = Alg_2(Alg_1(D), D)$ satisfies $(\varepsilon_1 + \varepsilon_2) - DP$.

Parallel composition. If $Alg_1(\cdot), Alg_2(\cdot), \cdots, Alg_k(\cdot)$ are k algorithms that satisfy $\varepsilon_1 - DP, \varepsilon_2 - DP, \cdots, \varepsilon_k - DP$ respectively, and $D_1, D_2, \cdots, D_k$ are k datasets which are disjoint. Then, $Alg_1(D_1), Alg_2(D_2), \cdots, Alg_k(D_k)$ satisfy $max(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_k) - DP$.

Based on the above-mentioned three features, different steps can be combined to construct a differential privacy preserving algorithm supporting different datasets and multiple query statistics.

(5) Analysis of algorithm complexity and efficiency: since the main idea of differential privacy preserving algorithm is adding noise to privacy information, the complexity depends on the noise generating, and the effectiveness of privacy preservation also relies on the size of the noise. These are related to the noise generating parameters such as characteristics of dataset, sensitivity calculations of datasets and so on. As a result, the complexity and the effectiveness of privacy preservation can be depicted as follows:

The complexity of algorithm Alg can be denoted as:

$$C(Alg) = c(\Phi, \Delta g, h(\cdot), \delta(\cdot), \mu(\cdot), b(\cdot), q(\cdot)).$$

The privacy preserving quality of algorithm Alg can be denoted as:

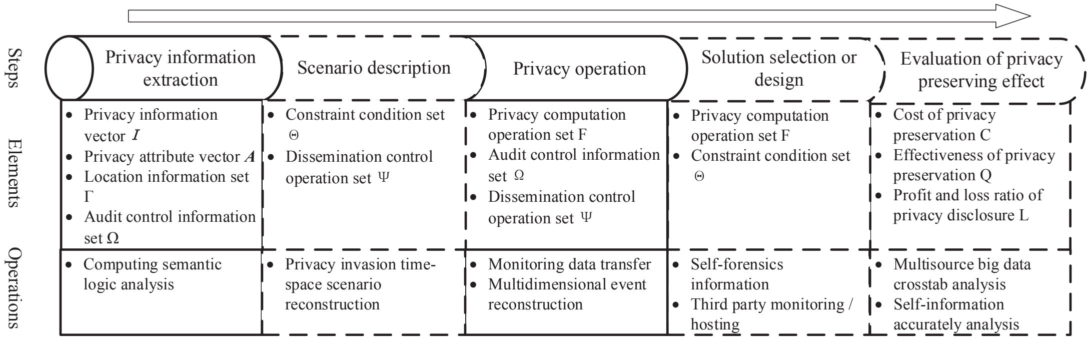
$$Q(Alg) = \Delta\sigma(h(\cdot), \delta(\cdot), \mu(\cdot), b(\cdot), q(\cdot)).$$

3.7 Privacy Computing Language

We propose a Privacy Computing Language (PCL), which can automatically implement formal description, dissemination control, computation and transaction processing in the lifecycle of privacy information. PCL consists of three parts: privacy defining language, privacy operating language and privacy controlling language.

(1) Privacy defining language: privacy defining language aims to describe the data type and the data format of the privacy computation six factors of information M, as well as relevant integrity constraints. Data type mainly includes bit string type, integer type, floating-point type, character type, logical type, table data, metadata, web data, text data, image data, audio data, video data and so on. In addition, privacy defining language is also used to describe computing steps on text, image, audio and video, including privacy information extraction, scenario abstraction, privacy operation selection, privacy preserving scheme selection and design, evaluation of privacy preserving effect and so on.

(2) Privacy operating language: privacy operating language is to describe the behaviors of operating information M, such as modular addition, modular multiplication, modular exponentiation, exclusive or, replacement, disturbance, query, selection, deletion, modification, copy, paste, cut, forward and so on.



Steps	Privacy information extraction	Scenario description	Privacy operation	Solution selection or design	Evaluation of privacy preserving effect
Elements	• Privacy information vector I • Privacy attribute vector A • Location information set Γ • Audit control information set Ω	• Constraint condition set Θ • Dissemination control operation set Ψ	• Privacy computation operation set F • Audit control information set Ω • Dissemination control operation set Ψ	• Privacy computation operation set F • Constraint condition set Θ	• Cost of privacy preservation C • Effectiveness of privacy preservation Q • Profit and loss ratio of privacy disclosure L
Operations	• Computing semantic logic analysis	• Privacy invasion time-space scenario reconstruction	• Monitoring data transfer • Multidimensional event reconstruction	• Self-forensics information • Third party monitoring / hosting	• Multisource big data crosstab analysis • Self-information accurately analysis

Fig. 2. The framework of tracing evidence for privacy invasion behavior

(3) Privacy controlling language: privacy controlling language is used to describe the access control authorization, identification and revocation of information M. The access control permission consists of selection, copy, paste, forward, cut, modification, deletion, query and so on.

3.8 Tracing Evidence for Privacy Invasion

In the framework of privacy computing, privacy invasion and evidence obtaining exist in each step. Tracing evidence for privacy invasion mainly includes four parts: definition of privacy information, determination of privacy violations, obtaining evidence of privacy invasion and tracing the origin of privacy invasion, as shown in figure 2.

Based on privacy computing framework, we abstract the characteristics and processes of privacy invasion, and integrate them with each step of the privacy computing framework. The framework of tracing evidence for privacy invasion behavior is depicted below:

(1) Privacy information extraction: when information M is generated, we deploy scenario logic computing analysis to extract information or label privacy information, so that we can get privacy information vector $\boldsymbol{I}$, location information set Γ and audit control information set Ω as well as privacy attribute vector $\boldsymbol{A}$. This phase is mainly for identifying and defining privacy information.

(2) Scenario description: by abstracting the scenario of information, we can obtain constraint condition set Θ and dissemination control operation set Ψ. This phase provides criteria for judging privacy invasion. If the above conditions are not satisfied, we judge that privacy invasion has happened.

(3) Privacy operation: according to the limitations of scenario, we assign executable operations to each privacy information component. In turn, we form the privacy computing operation set F. Further, we construct the dissemination control operation set Ψ. We record all privacy operations that information subject executes on the information, and then generate or update the audit control information set Ω. Operations beyond the above two sets are judged as privacy invasions.

(4) Solution selection or design: during this process, we analyze the operations from selected/designed schemes to check whether they can satisfy the set of privacy computing operations, and whether their behavior, object and result are outside the constraint condition set so that we can try to avoid privacy invasion and take this as a criterion for judging privacy invasion.

(5) Evaluation of privacy preserving effect: In this phase, we analyze and compute the cost of privacy preservation C, effectiveness of privacy preservation Q and profit and loss ratio of privacy disclosure L. If the above indicators do not meet expected goals, privacy invasion behaviors may have occurred, and hence we need to review the whole lifecycle preservation of privacy information.

(6) Evidence tracing: when a privacy invasion occurs, it is necessary to analyze the tracing source from the four phases above in order to trace the main entity of privacy invasion. Based on six tuples of privacy information and the third-party monitoring or trusteeship, we need to identify and define privacy information, and judge privacy invasion behaviors. Then, through the correlation of each step of privacy computing framework, we can obtain the evidence of abnormal behaviors and find out the source of the invasion so that we can realize evidence tracing.

4 Further Research Trends

Dynamic Privacy Metric: the data controlled by large-scale Internet companies flows across information systems, national network boundaries, and ecosystem. Since the existence of a variety of data types as well as diverse application scenarios, future research on privacy metric may focus on three aspects: the measurement method of privacy information suitable for multimedia scenarios, the dynamic adjustment mechanism of privacy metric, and the automatic mapping of privacy metric to constraint conditions and policies. Solving the core problem of dynamic privacy metric under huge data sets can support the scenario adaptive privacy control, especially in case of big data flowing unpredictable through random paths.

The Fundamental Theory of Privacy Preserving Algorithm: focusing on atomic operations for privacy preservation of different information types and privacy preserving requirements, we need to study the fundamental theory of privacy preserving primitive. In terms of encryption based reversible privacy preserving primitive, the main focus is to develop highly efficient ciphertext computation theories such as fully homomorphic encryption, partial homomorphic encryption, ciphertext search and ciphertext statistics. In terms of perturbation irreversible privacy preserving primitive, the main focus is to improve differential privacy models and to develop new theoretical methods in information theory.

Evaluation of Privacy Preserving Effect and Performance: to conduct evaluation of privacy preserving effect and performance, we need to further investigate how to establish a scientific and reasonable quantification system under which we can propose quantitative methods to evaluate indicators for privacy preserving primitive and primitive combinations. The indicators include privacy leakage, data utility, primitive complexity, etc. By this, we can provide guidelines for the design, comparison and improvements of privacy preserving schemes.

Privacy Computing Language: the grammatical system of privacy computing language, including statement definition, programming interface, fusion of privacy preserving primitives, etc., should be studied to provides a convenient and platform-independent programming tool for the implementation of complex privacy preserving schemes, so as to support the deployment of the privacy preserving mechanism in the complex interconnected information system.

Decision Criteria and Forensics of Privacy Violations: based on the description of privacy information by privacy computing framework, we can combine scenario perception, privacy information operation determination, privacy information constraint condition matching and so on, to carry out the study of multi-factors joint decision criteria for privacy violations, and therefore determine the quantitative threshold of decision. In order to solve the key problem in spatial-temporal scenario reconstruction of privacy violations, we should design practical forensic schemes based on the forensic information embed in privacy information description, third party monitoring, as well as cross-over multi-element big data analysis.

5 Conclusion

With the quick development of technologies such as Internet, mobile Internet and Internet of Things, data are aggregated together through Cloud services, generating big data. It has the typical characteristics such as massive, diverse, etc. Big data provides the public with personalized service, which has changed the way we work and live profoundly. However, information service is facing serious privacy leakage problems during the lifecycle of information flow which includes collection, storage, processing, publishing, destruction, and etc. Existing privacy preserving solutions mainly focus on one scenario by providing certain preservation on some particular aspects. They have not yet formed into a theoretical system. Therefore, the proposed concept of privacy computing and its framework aim to set up a whole lifecycle preservation on privacy information, including privacy computing framework, formal definition of privacy computing, four principles in privacy computing, algorithm design criteria, evaluation of privacy preservation, privacy computing language etc. Meanwhile, the privacy computing framework can support the privacy information exchange, extended authorization of privacy information circulation and the forensics tracking of privacy invasion in cross-platform scenario; the aim of the privacy computing language is to satisfy description unambiguity, platform irrelevance and computational consistency, which are able to support the layered cross information system implement of privacy preservation. Based on our proposed privacy computing framework, we implemented the differential privacy preserving mechanism in Baidu DuerOS. At the end of this paper, we prospect the research development trends of privacy computing. We expect that privacy computing could guide the practical research on privacy preserving technologies, and also guide the exploitation of privacy preserving subsystem in large-scale information system via promoting achievements of privacy computing. We also expect that the privacy computing could provide theoretical supports for enacting the criterion of privacy preservation and evaluating the ability of privacy preservation.

References

[1] H. Scherzer, R. Canetti, P. A. Karger, H. Krawczyk, T. Rabin, and D. C. Toll, “Authenticating mandatory access controls and preserving privacy for a high-assurance smart card,” in *Proc. of European Symposium on Research in Computer Security*. Springer, 2003, pp. 181–200.

[2] H. Lindqvist, “Mandatory access control,” *Master's Thesis in Computing Science, Umea University, Department of Computing Science*, vol. 87, pp. 1–104, 2006.

[3] J. M. McCune, T. Jaeger, S. Berger, R. Caceres, and R. Sailer, “Shamon: A system for distributed mandatory access control,” in *Proc. of Annual Conference on Computer Security Applications*. IEEE, 2006, pp. 23–32.

[4] D. Slamanig, “Dynamic accumulator based discretionary access control for outsourced storage with unlinkable access,” in *Proc. of International Conference on Financial Cryptography and Data Security*. Springer, 2012, pp. 215–222.

[5] R. Sandhu and Q. Munawer, “How to do discretionary access control using roles,” in *Proc. of the 3rd ACM Workshop on Role-based Access Control*. ACM, 1998, pp. 47–54.

[6] N. Li, “Discretionary access control,” in *Collection of Encyclopedia of Cryptography and Security*. Springer, 2011, pp. 353–356.

[7] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, “Role-based access control models,” *IEEE Journal of Computer*, vol. 29, no. 2, pp. 38–47, 1996.

[8] A. F. Dafa-Alla, E. H. Kim, K. H. Ryu, and Y. J. Heo, “Prbac: An extended role based access control for privacy preserving data mining,” in *Proc. of 4th Annual ACIS International Conference on Computer and Information Science*. IEEE, 2005, pp. 68–73.

[9] F. Li, Y. Wang, L. Yin, R. Xie, and J. Xiong, “Novel cyberspace-oriented access control model,” *Journal of Communications*, vol. 37, no. 5, pp. 9–20, 2016.

[10] V. Goyal, O. Pandey, A. Sahai, and B. Waters, “Attribute-based encryption for fine-grained access control of encrypted data,” in *Proc. of the 13th ACM Conference on Computer and Communications Security*. ACM, 2006, pp. 89–98.

[11] J. Bethencourt, A. Sahai, and B. Waters, “Ciphertext-policy attribute-based encryption,” in *Proc. of IEEE Symposium on Security and Privacy*. IEEE, 2007, pp. 321–334.

[12] J. Shao, R. Lu, and X. Lin, "Fine: A fine-grained privacy-preserving location-based service framework for mobile devices," in *Proc. of International Conference on Computer Communications*. IEEE, 2014, pp. 244–252.

[13] L. Sweeney, "k-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002.

[14] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, "Incognito: Efficient full-domain k-anonymity," in *Proc. of the 2005 ACM SIGMOD International Conference on Management of Data*. ACM, 2005, pp. 49–60.

[15] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," in *Proc. of IEEE 22nd International Conference on Data Engineering*. IEEE, 2006, pp. 24–24.

[16] F. Liu, K. A. Hua, and Y. Cai, "Query l-diversity in location-based services," in *Proc. of International Conference on Mobile Data Management: Systems, Services and Middleware*. IEEE, 2009, pp. 436–442.

[17] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," in *Proc. of IEEE 23rd International Conference on Data Engineering*. IEEE, 2007, pp. 106–115.

[18] D. Rebollo-Monedero, J. Forne, and J. Domingo-Ferrer, "From t-closeness-like privacy to postrandomization via information theory," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 11, pp. 1623–1636, 2010.

[19] C. Dwork, "Differential privacy: A survey of results," in *Proc. of International Conference on Theory and Applications of Models of Computation*. Springer, 2008, pp. 1–19.

[20] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *Proc. of IEEE Symposium on Foundations of Computer Science*. IEEE, 2007, pp. 94–103.

[21] R. Dewri, "Local differential perturbations: Location privacy under approximate knowledge attackers," *IEEE Transactions on Mobile Computing*, vol. 12, no. 12, pp. 2360–2372, 2013.

[22] A. Blum, K. Ligett, and A. Roth, "A learning theory approach to noninteractive database privacy," *Journal of the ACM (JACM)*, vol. 60, no. 2, pp. 1–25, 2013.

[23] R. L. Rivest, L. Adleman, and M. L. Dertouzos, "On data banks and privacy homomorphisms," *Foundations of secure computation*, vol. 4, no. 11, pp. 169–180, 1978.

[24] H. Zhu, F. Liu, and H. Li, "Efficient and privacy-preserving polygons spatial query framework for location-based services," *IEEE Internet of Things Journal*, vol. 4, no. 2, pp. 536–545, 2017.

[25] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proc. of International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 1999, pp. 223–238.

[26] R. Lu, X. Liang, X. Li, X. Lin, and X. Shen, "Eppa: An efficient and privacy-preserving aggregation scheme for secure smart grid communications," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 9, pp. 1621–1631, 2012.

[27] C. Gentry, *A fully homomorphic encryption scheme*. Stanford University, 2009.

[28] E. Bayer-Fluckiger, "Ideal lattices," *A panorama of number theory or the view from Baker's garden*, pp. 168–184, 2002.

[29] Z. Brakerski, C. Gentry, and V. Vaikuntanathan, "(leveled) fully homomorphic encryption without bootstrapping," in *Proc. of the 3rd Innovations in Theoretical Computer Science Conference*. ACM, 2012, pp. 309–325.

[30] A. Lopez-Alt, E. Tromer, and V. Vaikuntanathan, "On-the-fly multiparty computation on the cloud via multikey fully homomorphic encryption," in *Proc. of the 44th Annual ACM Symposium on Theory of Computing*. ACM, 2012, pp. 1219–1234.

[31] C. Gentry, A. Sahai, and B. Waters, "Homomorphic encryption from learning with errors: Conceptually-simpler, asymptotically-faster, attribute-based," in *Collection of Advances in Cryptology*. Springer, 2013, pp. 75–92.

[32] H. Zhu, F. Wang, R. Lu, F. Liu, G. Fu, and H. Li, "Efficient and privacy-preserving proximity detection schemes for social applications," *IEEE Internet of Things Journal*, vol. 0, no. 0, pp. 1–1, 2017.

[33] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *Proc. of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2011, pp. 325–334.

[34] S. Oya, C. Troncoso, and F. Pérez-González, "Back to the drawing board: Revisiting the design of optimal location privacy-preserving mechanisms," in *Proc. of ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2017, pp. 1959–1972.

[35] C. Y. T. Ma and D. K. Y. Yau, "On information-theoretic measures for quantifying privacy protection of time-series data," in *Proc. of ACM Symposium on Information, Computer and Communications Security*. ACM, 2015, pp. 427–438.

[36] P. Cuff and L. Yu, "Differential privacy as a mutual information constraint," in *Proc. of ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2016, pp. 43–54.

[37] Z. Jorgensen, T. Yu, and G. Cormode, "Conservative or liberal? personalized differential privacy," in *Proc. of International Conference on Data Engineering*. IEEE, 2015, pp. 1023–1034.

[38] S. Asoodeh, F. Alajaji, and T. Linder, "Notes on information-theoretic privacy," in *Proc. of Communication, Control, and Computing*. IEEE, 2015, pp. 1272–1278.

[39] A. Gervais, R. Shokri, A. Singla, S. Capkun, and V. Lenders, "Quantifying web-search privacy," in *Proc. of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2014, pp. 966–977.

[40] Y. Cao, M. Yoshikawa, Y. Xiao, and L. Xiong, "Quantifying differential privacy in continuous data release under temporal correlations," *IEEE Transactions on Knowledge and Data Engineering*, vol. 0, no. 0, pp. 1–1, 2018.

[41] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, "Quantifying location privacy," in *Proc. of IEEE Symposium on Security and privacy*. IEEE, 2011, pp. 247–262.

[42] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, "Protecting location privacy: optimal strategy against localization attacks," in *Proc. of the 2012 ACM Conference on Computer and Communications Security*. ACM, 2012, pp. 617–627.

[43] C. Kiekintveld, J. Marecki, and M. Tambe, "Approximation methods for infinite bayesian stackelberg games: Modeling distributional payoff uncertainty," in *Proc. of International Conference on Autonomous Agents and Multiagent Systems-Volume*. ACM, 2011, pp. 1005–1012.

[44] A. Srinivasan, J. Wu, and W. Zhu, "Safe: Secure and big data-adaptive framework for efficient cross-domain communication," in *Proc. of the First International Workshop on Privacy and Security of Big Data*. ACM, 2014, pp. 19–28.

[45] X. Wu, T. Wu, M. Khan, Q. Ni, and W. Dou, "Game theory based correlated privacy preserving analysis in big data," *IEEE Transactions on Big Data*, vol. 0, no. 0, pp. 1–1, 2017.

[46] Z. Zhang, S. He, J. Chen, and J. Zhang, "Reap: An efficient incentive mechanism for reconciling aggregation accuracy and individual privacy in crowdsensing," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 12, pp. 2995 – 3007, 2018.