# A VCG Pricing Model of Privacy Protection Service Based on AHP

Mingqi Wang [a, *]

School of Energy Power and Mechanical Engineering, North China Electric Power University, Baoding 071000, China

[a] 792426754@qq.com

## Abstract

**Our paper proposes a pricing model of privacy protection service, and analyzes influence of the basic factors. We establish an VCG (Vickrey-Clarke-Groves) pricing model of privacy protection service based on AHP (Analytic Hierarchy Process). We set up an index system of privacy risk assessment, and we select nine parameters in indicators level under the criterion level. Then, based on the AHP model, we obtain the weights of each indexes. Next, we develop a nine-parameter scoring principle based on characteristics of the individuals and specific domain. According to the value of privacy risk assessment, we figure out the privacy protection budget, and then the privacy protection service level is obtained. Therefore, the final pricing of privacy protection service is determined by VCG pricing principle. We have selected 6 representative information to simulate the model in three fields which are social media, financial transaction and record of health and medical. What's more, we discuss the effects of data protections, the basic elements of the data and data sets on the basic model.**

## Keywords

**AHP; VCG; pricing for the cost of privacy.**

## 1. Introduction

The spread and dependence of electronic communication and social media have become widespread, and people's perceptions of privacy change as well, making it possible as a commodity. There are also significant differences in privacy choices in different areas. If certain groups or subgroups consider personal information to be a personal or community risk, they may be unwilling to relinquish certain types of personal information. The personal choice of cybersecurity, Internet and system security appears to create the risks and rewards of freedom, privacy, convenience, social status, economic benefits and health care. We establish a privacy protection service pricing model. This model will take into account the privacy attributes, and then show its value, and can achieve the level of pricing.

## 2. Assumptions

All pricing policies comply with the value of the product itself, and the laws and regulations are reasonably.

The selection of parameters value in indicator level is completely rational.

The pricing model of privacy protection service is only related to the indicators considered, ignoring other less influential indicators.

The data we collect is sufficient and accurate.

The subjective deviation is small when calculate weights in AHP.

There are no make much difference in the selection of criteria for people in different cultures.

## 3. The Basic Model

In this section, we establish a privacy risk assessment model based on AHP, which reflects the price point of privacy protection on different applications. We first build a privacy risk assessment index system. Then, according to previous literatures, we give the judging matrixes of criterion levels and indicator level and calculate the weights of two levels in the evaluation system. Finally, we illustrate how nine assessment parameters can be combined with personal and field-specific information.

### 3.1 Privacy risk assessment index system

Personal privacy is essentially a manifestation of information. Therefore, our security risk assessment of personal privacy can refer to the information security risk assessment method. Information security risk assessment is a systematic and comprehensive assessment to the systems. Security risks are determined by the likelihood and impact of information security incidents. So we come to the conclusion that the security risk of the information system is finally identified based on the assessment of the possibility and the negative impact of the security incident.[1]

This paper establishes a privacy risk assessment index system as shown in Table 1 after detailed summary, classification and screening, combined with the definition of ITSEC (Information Technology Security Evaluation Criteria) and based on the risk assessment methods of privacy leakage at home and abroad.

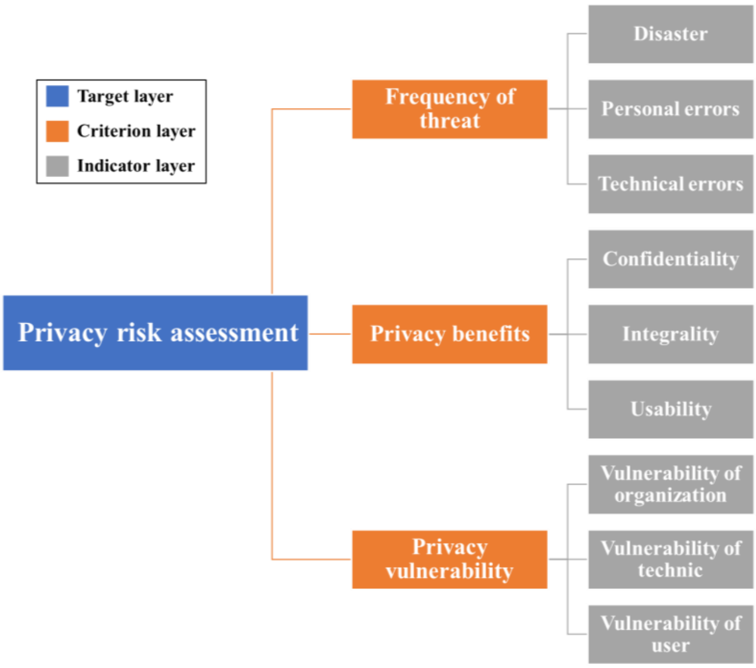


Figure1. Privacy risk assessment index system

### 3.2 Privacy risk assessment model based on AHP

#### 3.2.1 Determine the weights of the criterion level

Step 1: Define the relative important scale

Relative importance scale: for any two elements $C_i$ and $C_j$, use $a_{ij}$ to represents the ratio of $C_i$'s and $C_j$'s Influence on O. The results of comparison can be expressed by the paired comparison matrix:

$$\mathbf{A} = (a_{ij})_{n\times n}\ , \qquad a_{ji} = \frac{1}{a_{ij}} \tag{1}$$

Where $a_{ij}$ is set according to the one-nine method by Stayy.

Step 2: Give the judging matrix

The judging matrix of the criterion level is as following:

$$\begin{bmatrix} 1 & 5/2 & 4/3 \\ 2/5 & 1 & 1/2 \\ 3/4 & 2 & 1 \end{bmatrix}$$

Similarly, we can get the judging matrix of the indicator level which will be discussed below.

Step 3: Calculate the eigenvectors

We figure out the eigenvectors in matlab, and the result is:

$$w = [0.4634 \quad 0.1814 \quad 0.3551]^T$$

Thus, the Privacy risk can be expressed as:

$$T = \omega_1 F + \omega_2 B + \omega_3 V \tag{2}$$

Step 4: Do the consistency check

According to the AHP theory, CI discribes the degree of inconsistency of the A matrix. we can see the expression of consistency indicator CI is:

$$CI = \frac{\lambda - n}{n-1} \tag{2}$$

And the RI table in the AHP is as follows:

Table 1. The numerical table of random consistency index RI

| n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| RI | 0 | 0 | 0.58 | 0.90 | 1.12 | 1.24 | 1.32 | 1.41 | 1.45 | 1.49 | 1.51 |

In our problem, the value of n is 3, so we can obtain RI=0.58.

The expression of consistency check ratio CR is:

$$CR = \frac{CI}{IR} \tag{3}$$

We can obtain the results of the consistency test, which is CR=0.0004<0.1. Thus the eigenvector can be used as the weights of the criterion level.

### 3.2.2 Determine the weights of the indicator level

The establishment of indicator level's judging matrixes adopts the principle of scoring by experts. The whole process is similar to the weights' determination of the criterion level. To simplify the description, we give the results of weights of the criterion level directly which are presents in table 2.

Table 2. Comprehensive weights of Privacy risk assessment model

| Criterion level | Weights | Indicator level | Weights |
|---|---|---|---|
| **Frequency of threat** | 0.4634 | Disaster | 0.4745 |
| | | Personal errors | 0.2497 |
| | | Technical errors | 0.2758 |
| **Privacy benefits** | 0.1814 | Confidentiality | 0.4470 |
| | | Integrality | 0.2375 |
| | | Usability | 0.3155 |
| **Privacy vulnerability** | 0.3551 | Vulnerability of organization | 0.3739 |
| | | Vulnerability of technic | 0.2440 |
| | | Vulnerability of user | 0.3821 |

Combining the above sections, we can get the expression:

$$C = w_F \sum_{i=1}^{3} w_{F_i} F_i + w_B \sum_{i=1}^{3} w_{B_i} B_i + w_v \sum_{i=1}^{3} w_{v_i} V_i \tag{5}$$

Substitute the values of all weights into equation (5), we can obtain a method to assess Privacy risk.

### 3.2.3 Analysis of the result

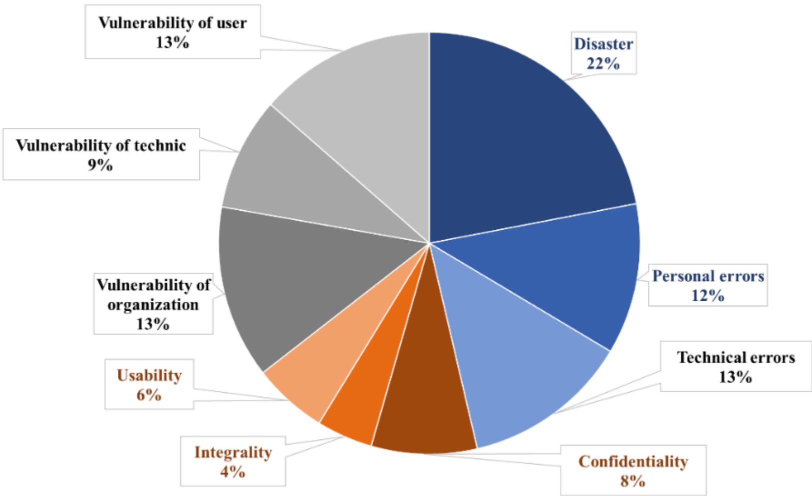


Figure2. Weight vector of indicatior level. Indexes in three criteria levels are located in different areas whose colors are blue, gray, and orang. The blue area represents Frequency of threat, while the orange area represents Privacy benefits and the gray area represents Privacy vulnerability.

Through the AHP model, we get the weights of nine parameters in the indicator level. Among them, "disaster" has the largest weight and " integrality " takes the least weight. The above figure is in accordance with the law of our statistical data. Although the weights of each index is different, the value of any parameter in the indicator level may affect the final pricing result. Therefore, we should rigorously and rationally choose the value of each parameter in the indicator level.

### 3.3 Principle of assigning parameters in indicator level

In order to simulate the privacy risk accurately, we should take characteristics of the individuals and characteristics of the specific domain of information into consideration. In this paper, we choose nine parameters as indicators considering factors such as age, identity, industry and domains. Table 1 in appendix clearly shows the relationship between them.

Based on the likelihood of occurrence, we divide the 9 parameters into the following five ranks. The higher the rank is, the greater the probability of occurrence.

Table 3. Definition of Parameter ranks.

| Rank | Lowest | Lower | Moderate | Higher | Highest |
|---|---|---|---|---|---|
| Score | 0~20 | 20~40 | 40~60 | 60~80 | 80~100 |

Table 4. Definition of probability level

| Rank | Definition |
|---|---|
| Highest | This indicator is highly probable and in almost all cases almost inevitable |
| Higher | This indicator is more likely to occur and in most cases is likely to occur |
| Moderate | The likelihood of this indicator occurring is medium and may in some cases occur |
| Lower | This indicator is less likely to occur and is generally not likely to occur |
| Lowest | This indicator is unlikely to occur in very rare circumstances |

### 3.4 Simulation of the model

**Get the privacy risk assessment scores in three areas**

Based on the nine indicators and their evaluation criteria given in the basic model, we search the statistics of the National Institute of Statistics (references) and data statistics websites (references) on the communications sector, finance and health. We select 6 sets of information for evaluation from the 3 fields of social media, financial transaction and record of health, and medical, and then get the result of the scoring in the following table.

Table 5. The scoring results of 9 parameters

| Domains | Groups | $F_1$ | $F_2$ | $F_3$ | $B_1$ | $B_2$ | $B_3$ | $V_1$ | $V_2$ | $V_3$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Social media | National policy on the site | 27 | 70 | 89 | 34 | 56 | 78 | 86 | 81 | 52 |
| | WeChat chat history | 78 | 57 | 48 | 29 | 67 | 53 | 41 | 56 | 67 |
| Financial transactions | Bank card password | 89 | 73 | 21 | 13 | 79 | 46 | 12 | 24 | 75 |
| | Business accounts | 87 | 86 | 59 | 58 | 86 | 28 | 87 | 34 | 45 |
| Record of health, and medical | Results of physical examination | 74 | 27 | 12 | 12 | 68 | 57 | 76 | 54 | 52 |
| | Flu records | 95 | 94 | 98 | 67 | 24 | 68 | 65 | 67 | 34 |

Combined with the AHP model, the evaluation indexes of the privacy risk assessment of the representative 6 sets of information from the 3 fields are as follows:

Table 6. the evaluation indexes of the privacy risk assessment

| Domains | Groups | T |
|---|---|---|
| social media | National policy on the site | 66.78528 |
| | WeChat chat history | 39.19844 |
| Financial transactions | Bank card password | 56.35972 |
| | Business accounts | 59.66692 |
| Record of health, and medical | Results of physical examination | 57.2646 |
| | Flu records | 62.9961 |

To simulate the above 6 sets of information from the 3 fields in our model, we can calculate the index of privacy risk assessment. The result suits for reality well, and T is related to the value of information. We can conclude that the value of information on the "national policy on the site," is higher, while the individual information of "Wechat chat history" is of lower value. In fact, the national policy on the website are of more information and influence, and personal WeChat chat information is less important and influential. Therefore, its privacy risk assessment value is smaller.

### 3.5 The VCG mechanism cost pricing model based on AHP

#### 3.5.1 Differentiated privacy protection

Differential privacy protection is a privacy protection technique based on data distortion that uses noise-adding techniques to distort sensitive data while keeping certain data or data attributes constant.

It ensures that the processed data can retain certain statistical properties in order to data mining and other operations.

Definition of Differentiated Privacy Protection: Given two adjacent data sets D and D ', a privacy protection algorithm A, and Range (A) is A range of values. If algorithm A outputs O (O∈Range (A)) arbitrarily on datasets D and D ' and the following formula is satisfied:

$$\Pr[A(D)=0] \le e^{\varepsilon} * \Pr[A(D')=0] \tag{6}$$

Then algorithm A satisfies ε-differential privacy. Among them, ε is called the privacy protection budget. the smaller ε is, the higher the degree of privacy protection is. The larger the ε is, the lower the degree of privacy protection is. The definition of differential privacy provides a theoretical basis for the classification of privacy protection services.[2]

The better the privacy protection service level is, the higher the privacy protection degree is. In the definition of differential privacy protection, privacy protection budget ε reflects the degree of differential privacy protection. Then each level corresponds to a value range of ε. Combined with the above result we can get the corresponding ε to the five ranks of risk evaluation , and the range is as follows:

Table 7. Privacy risk assessment scores and the corresponding 5 ranks of ε

| Rank | Lowest | Lower | Moderate | Higher | Highest |
|---|---|---|---|---|---|
| Score | 0~20 | 20~40 | 40~60 | 60~80 | 80~100 |
| ε | 0~0.2 | 0.2~0.4 | 0.4~0.6 | 0.6~0.89 | 0.8~1.0 |

### 3.5.2 VCG pricing model

VCG (Vickrey-Clarke-Groves) mechanism is aimed at that the price of digital products or services can not be evaluated. It is based on Vickrey's single-item auction, and promoted to more general price Auction mechanism by Clarke and Groves. The bidder firstly submits the quotation for each lot, and then the auction system distributes the lot to each bidder in a socially optimal manner. Each bidder pays the price equal to that part of the increase the sum of the value obtained by the other bidders when the bidder does not appear.[3]

The VCG mechanism satisfies the principle of incentive compatibility and individual rationality at the same time. This mechanism realizes the consistency of individual utility and the overall social wealth. Not only can it be able to motivate the bidder to bid on the real value of the lot, it can also achieve the optimal distribution of the society.

**Step 1: determine the best match**

The pricing mechanism is the auction of k privacy protection service levels. The bid of k bidders were $b_1$, $b_2$, ..., $b_k$, to buy a certain level of service.

Define the social welfare function:

$$\omega(\bullet) = \sum_{j=1}^{k} V_{ij} \tag{7}$$

Where $V_{ij}$ denotes the privacy valuation of the privacy service level of the bidder bidding on $b_j$ corresponding to the privacy protection intensity $V_i$. The expression $V_{ij} - P_i$ can reflect the rewards of bidders buying Vi.

Through the optimal matching, the auction system has obtained the matching method of maximizing the value of ω (•) of the social welfare function. It assign a privacy protection level to each type of bidder, and then set the price for each level according to the principle of compensation loss in the VCG mechanism.

There are biograph (x, y), if you can find a program with the largest number of a group of matches, we record it as the maximum match. If it satisfies

| x | = | y | = number of matches

We call the program one perfect match.

The process of constructing the optimal match is similar to the bipartite graph's optimal matching process. The process of optimal matching is as follows:

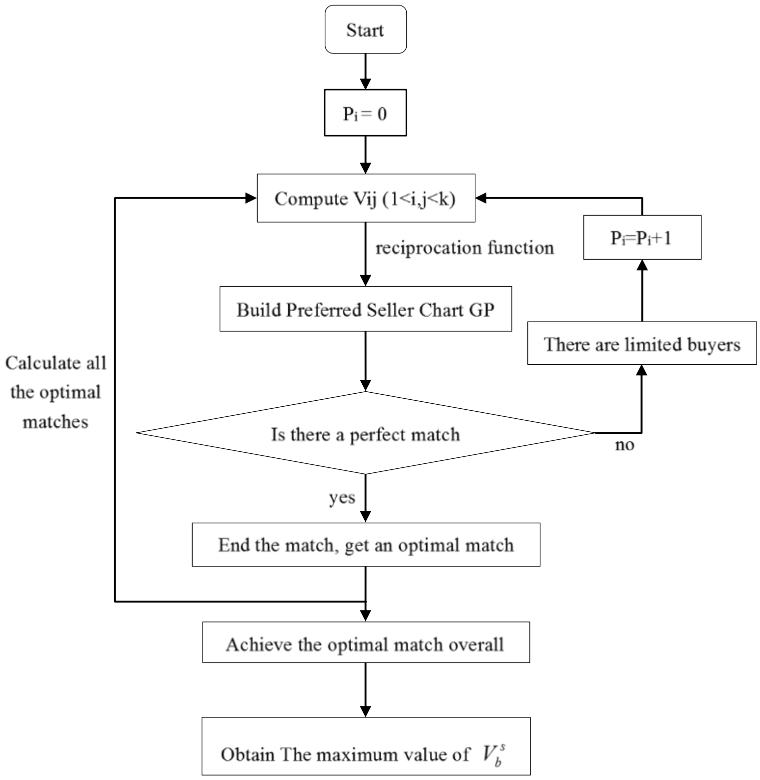


Figure 3. The process of optimal matching

Step 2: Calculate the VCG price for rank i privacy protection service

Let S be a set of privacy protection service levels and B be a set of auction prices, while $V_B^S$ is the sum of all the privacy evaluations of all bidders after the best match. When making optimal match, the service of $L_i$ rank is matched with bidders bidding for $b_j$, then:

$V_{P-j}^{S-i}$ : The sum of the privacy estimates of the remaining matches removing the set of Li-bj matches.

$V_{B-j}^{S}$ : The sum of privacy valuation of the remaining bidders rerun the optimal match, while bidders do not participate in the auction biding on $b_j$.

Therefore, the VCG price of privacy protection services of the level i is:

$$m_i=V_{B-j}^{S}-V_{P-j}^{S-i} \tag{8}$$

So the VCG price of privacy protection services at all levels can be derived.

According to the data in the above two tables, we combine with the VCG algorithm. Then we obtain the relative value of the representative 6 sets of information from the 3 fields using the formula (9) in the matlab. The result is as follows:

Table 8. The results of pricing.

| Information | WeChat chat history | Bank card password | Physical examination results | Business accounts | Flu case record | National policy on the site |
|---|---|---|---|---|---|---|
| Relative value | 0 | 6.63 | 7.19 | 8.90 | 10.70 | 13.22 |

## 3.6 Summary of the pricing model

Overall, our pricing model simulates the process of converting private information into a vendable product. It ultimately enables the pricing of privacy. The main process of the model is as follows:

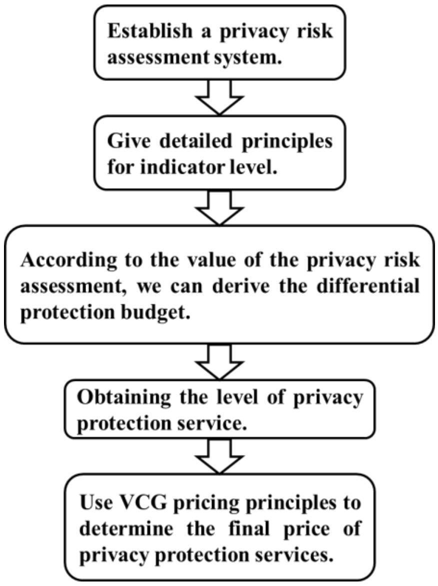


Figure 4. The process of the basic model

According to the model, we can conclude that the determination of the parameters in the indicator layer directly determine the final pricing of privacy protection service. And there exists the following relationships: the higher the index layer score is, the higher the privacy risk assessment value is, and the higher the price of privacy protection service. Therefore, in the following study, we mainly study the factors that affect the change of the indicator level, and then draw the influence on the final pricing.

### 3.7 Some influence factors of the model

#### 3.7.1 The consideration of data protection

It is obvious from the above model that in the case of certain data privacy protection, we don't consider the impact of policy or subjective factors on the pricing of privacy. Combining the value of privacy and data protection conditions to determine the final pricing of goods. For example, political policies, market fluctuations, and artificial subjective opinions can't govern the final pricing of privacy.

In the ensuing work, we will focus our attention on political policies, market fluctuations or subjective factors. When privacy gets some data protection, we can change the policy influence or subjective factors parameter values, thus directly affecting the privacy pricing.

#### 3.7.2 The impact of elemental data on the model

According to our pricing model, to price a private information requires obtaining the value of the parameters in indicator level during the privacy risk assessment. And then pricing it. Therefore, when pricing some elements of privacy data, we first need to consider the value of the parameters in indicator level.

To analyze the impact of five different basic elements, such as name, birthday, gender, social security and citizenship number, on the pricing model, we analyze the differences in the selection of their parameters in the following table.

Table 9. Differences in parameter selection of five data elements

| Data elements | Bias of partial parameter selection |
|---|---|
| Name | Integrity, usability, personal errors and vulnerability of user are low |
| Birthday | Integrity, usability and vulnerability of user are low |
| Gender | Integrity and usability are low |
| Social Security | Confidentiality, usability is higher while integrity and vulnerability of technology is lower |
| Citizenship number | Confidentiality, usability is higher; while integrity is lower |

From the above analysis, we can see that Social Security and Citizenship number are more useful and their "Usability" indicators have larger values, so the risk posed by the spill is greater, resulting in a higher privacy risk assessment value. Combined with the calculation of the pricing model, we can conclude that the pricing of social security and citizenship numbers will be higher than other factors.

### 3.7.3 The impact of data sets on the model

We can see from our privacy risk assessment model that the number of data elements in a data set directly affects the values of " integrity " and "usability " in the indicator level. And when the data elements are superimposed, the risk caused by the leakage is much larger than the sum of the risks caused by the single data leakage. For example, a single name will have a lower pricing due to its lower integrity, usability, personal error, and vulnerability of user, our model will give a lower price. While the personal error and vulnerability of user parameters will have a lower value for photo-attached names. However, its integrity and usability parameters will be greatly improved, and the value of disaster will also increase, resulting in the price of the name attached to the person is higher than that of a single name.

## Appendix

The relationship between 9 parameters and characteristics of the individuals and the specific domain of information.

| parameters | Relationship |
|---|---|
| Confidentiality | Researchers have stricter requirements on the confidentiality of information on work, while superstar care less on this concern. Compared with the young, the old demands more on the confidentiality of information. |
| Integrality | The statistics of the national data statistics websites are more complete while those in the local statistics websites are not integrated enough. |
| Usability | The Centers for Disease Control and Prevention can track the transmission of the disease by analyzing the information of the group's information on disease or health, so as to control the disease and have certain Usability to social. The information of a single person on suffering from the common disease is useless in comparison. |
| Disaster | Extremists are more likely to bring disaster, while infants have little probability. |
| Personal errors | There are individuals or businesses who may steal, resell, use improperly personal information, or collect information without the owner’s consent. These phenomena result in personal information embracing a great risk of disclosure. |
| Technical errors | Professionals and non-professionals have much different possibility of technical errors. |
| Vulnerability of organization | The situations can cause certain organizational vulnerabilities such as organizations and enterprises that have access to personal information are not managed well or laws and regulations on this aspects are inadequate. |
| Vulnerability of technic | Some smart devices may lead to certain technical vulnerabilities by letting out personal identity, consumption, communication, finance, social relations and other information. |
| Vulnerability of user | Internet users reveal personal information due to simple passwords or negligence of information behavior. |

## References

[1] Fu Yu, Wu Xiaoping, Ye Qing, Peng hee. Research on security risk assessment of information systems based on fuzzy sets and entropy weight theory [J]. electronic journal, 2010,38 (07): 1489-1494.
[2] ZHANG Xiao-Jian, MENG Xiao-Feng.Differential Privacy Protection for Data Distribution and Analysis [J] .Journal of Computers,2014,37(4):927-949.
[3] Shi Wuchao, Wu Zhenqiang, Liu Hai.A Pricing Mechanism for Differential Privacy Service Based on VCG Mechanism [J]. Computer Technology and Development, 2017,27 (06): 119-123 + 129.
[4] https://wenku.baidu.com/view/a77241c080c758f5f61fb7360b4c2e3f57272580.html