For office use only	Team Control Number	For office use only
T1 ________________	85722	F1 ________________
T2 ________________		F2 ________________
T3 ________________	Problem Chosen	F3 ________________
T4 ________________	F	F4 ________________

2018

MCM/ICM

Summary Sheet

With the commercialization of private information (PI), how to formulate a reasonable price of PI has become a matter of concern. Considering risks and benefits of selling private information, we develop a price point model for sellers. Then we build a pricing system for individuals, groups and nations according to game theory. Due to the social connection among people, we consider spread of personal belief and network effects in a community.

Firstly, we divide PI into two categories: identify PI and PI in specific domain. Considering threats in specific information domain and vulnerability of individuals, we calculate likelihood of damage occurrence. Based on the above parameters, we give definition of risk. After that, we conduce gain function using amount and importance degree of exposed PI. The **price point model** is developed by gain function with limit of risk. Next, we quantify tradeoff and risk thus endow weights to them. Combined with tradeoff and risk, we define **the cost of PI**. Then, **analysis hierarchy process** (AHP) is used to weigh elements for identity PI and PI in specific domain.

Secondly, based on **game theory**, a **pricing system** is built for individuals, groups and nations. In order to maximum revenue of parties in transactions, we use **double-objectives with particle swarm optimization** (DOPSO) to optimize benefits thus get the optimal pricing strategies. Then, considering the market fluctuations, we use **gravity model** to describe the forces of supply and demand for PI.

Thirdly, depending on whether people are willing to sell PI, we divide people into two groups: sellers and protectors. By means of **susceptible-infection-susceptible** (SIS) in complex networks, **dynamic belief spread** (DBP) model is set up to describe transfer of beliefs. With generation aging, the belief towards PI will also alter. Because more and more PI will be exposed and correlation of PI will increase over time, we calculate of beliefs and benefits in different generations.

Fourthly, due to social connection among people, data sharing of one person may contain privacy of others. To capture network effects of data sharing, we establish **scale-free networks** which degree distribution meeting power law. After defining intimacy of people, we update cost of nodes. Ultimately, we find people with large contacts with others will become monopolists thus influence the whole market price.

Lastly, as **sensitivity analysis**, we update belief of people in protecting PI after a massive data breaching in dark web. After that, we write a memo for decision-maker to put up our recommendations.

Contents

1 Introduction

1.1 Background

Personal information is spreading on the Internet at an unprecedented pace, which not only forms different social network cultures, but also threatens private information security to some extent. With the pervasiveness of social media, privacy information gradually owns its commercial value and different types of privacy information leak may cause enormous risk for individuals. For instance, students are unwilling to share their gender information in the recruitment to avoid gender discrimination. Elders may pay more attention to their illness records and cure situation. How to protect privacy information has become a matter of concern in cyber space.

According to the news, the average economic expenditure in preventing data leak over the last two years is $2.4 million. A survey of 9000 customers conducted by Vanson Bourne [1] show that 70 percent are willing to share their private data to companies in return for discounts. It is common to sell and purchase private information on the web in recent years.

As for privacy protection, there exists an increasing number of solutions to protect private information. It is well-known that difference privacy (DP) by adding Laplace noise into dataset can protect private information efficiently. Additionally, the traditional information encryption also has an irreplaceable effect on privacy protection.

1.2 Problem Restatement

In order to evaluate cost of privacy and the change of personal beliefs, we solve the problems as follows:

- Categorize different individuals into subgroups, and select parameters and measures to model risk of PI, then to develop a price point for PI.
- Considering the tradeoffs and risks, we set up a model for cost of privacy in different specific information domains.
- A pricing system is established to define the value of privacy for individuals, groups and nations.
- Identify the assumptions and constrains and introduce dynamic element to predicate changing of beliefs in different information domains.
- Consider the differences in PI price with respect to ages.
- Taking into account the connection among people, we need to capture network effects of data sharing and consider its influence on our model.
- Consider influence of millions of PI data breach in our model.
- Write a memo to decision-maker.

1.3 Analysis of problem.

In the process of trading private information. Sellers need to weigh up risks and benefits. We need to evaluate the risk that people will undertake and set a reasonable price for PI.

From the perspective of sellers, their private information can be divided into two categories: identity PI and PI in special domains such as social media, health record and financial transactions. Private information should be evaluated by sellers from three aspects: PI type, amount of PI and importance degree of PI. Comprehensively, sellers can be classified into different risk levels with different information domains. With the limit of risk of privacy, price point is developed as the limit of benefits.

In order to set cost of privacy, we need to take tradeoff and risk into consideration. Tradeoff represents the real benefit which can be calculated by gain function. Risk means the potential risk of information itself for sellers at different risk levels in different special domains. So, we can calculate the cost of PI by means of tradeoff and risk.

In the real world, PI is not only evaluated by the estimation of sellers themselves, but also has its utility value for buyers. In order to establish a price system, we need to consider a game process in view of buyers and sellers. At the same time, the influence of market supply and demand on price need to be considered.

2 Assumptions and Notations

2.1 Assumptions

The private information is discrete, independent, Quantifiable. The private information we can value such as age, financial situation, health records is presented in the form of data which is countable and quantifiable.

Private information has both subjective and objective values. Different types of private information have subjective and distinguishing importance degrees for the sellers, but also has utility and business value for buyers.

There exists an increasing number of sources of privacy information. For private information on sale, information owners could take the initiative to sell, and thieves could also conduct illegal sale in cyber space.

There is a limit for people to sell privacy. Because some parts of information would do great damage to the safety of lives and properties if it is exposed to others.

Both parties involved in transactions are completely rational. We consider the PI transaction as a game. Both parties should maximize their own interests under certain conditions.

In social networks, the contact between two nodes is stable. Considering that attitudes or beliefs of PI are spread one by one in networks at a certain rate, the contact need to be stable and not changed over time.

More and more PI will be exposed owing to the development of social network. With the development of social media and meta-date, PI will be exposed more widely on a large scale.

PI of individual may also contain other people's specific PI. In social networks, because of the highly link of individual PI, it is likely that other PI may be included in the sale process unintentionally.

Price is subject to management of market fluctuations and government regulations. Reasonable prices need to comply with the laws of the market economy, but also need to accept the macro-control from the state.

There are no isolated nodes in the social network. In a social network, everyone will build connections with others, and their thoughts and behaviors are more or less influenced by others.

2.2 Notations

Symbols	Notations
$D_i\ (i=1,2,...,m)$	The i^{th} specific type of PI
l_i^k	Subject importance degree of PI for the i^{th} owners in domain k k
tf_i^k	The i^{th} threat factor of PI in domain k
vul_j^k	The j^{th} vulnerability of PI in domain k
$p(vul_j^k \mid tf_i^k)$	Risk level for an individual in domain k
η_{1i}	Discernibility cost of the i^{th} PI
η_{2i}	Importance degree of the i^{th} PI
G_i^k	Gain function of the i^{th} PI in domain k
$R(D_i, D_j)$	Correlation degree of the i^{th} PI and the j^{th} PI
U	Utility value of one's PI for purchasers
i_P	Proportion of population who is willing to sell PI
β_P	Transfer rate from protectors to sellers
μ_P	Transfer rate from sellers to protectors

3 Task 1: Risk and Price Point

3.1 Measures and Parameters

Although personal private information can bring some economic benefits, excessive exposure of private information will cause irreversible damage to the individual's own safety and social life. Therefore, how to reasonably price private information become the main problem.

For an individual, private information she or he can be classified into two categories: identity PI and PI in special domains. Identity PI exists in all kinds of domains, but PI in specific domain is only valuable in special area.

Considering the characteristics of personal identity PI and PI in specific domain such as illness records, financial transaction records， we set up following parameters to describe the privacy transaction process from the perspective of owners or sellers.

- Private Information: we assume that the Information is quantifiable, discrete and separable， so we define different information as discrete symbols.
- Discernibility cost: it means the amount of exposed PI, when the ratio of exposed information accounting for total number of information is increasing, its value will increase.
- Information importance degree: different information in specific domain has different importance degree, for instance, illness record is more important in health than in the social media. Important degree is a subject estimation from the view of owners themselves. It is evaluated by owners.
- Treat factor: means that threat sources and some events which influence on sellers themselves.
- Vulnerabilities: flaws in the information exposed by individuals. If information is exposed, it is likely that data owners can be harmed by others.
- Risk level: in specific domains, we comprehensively consider how the amount and importance degree of information affect the entire individual private information risk level, put another word, we give a fixed risk level to a specific

man in specific domain. For instance, if an old man has many kinds of illnesses, and his illnesses are all severe, we can consider this man has high risk level.

- Gain function: people could be willing to share their private information in order to exchange some revenue or discounts. Therefore, according to the characters of PI above, a gain function which describes benefits one person received was established.

3.2 Price Point.

Based on the parameters above, we could develop a price point. Firstly, we assume that the set of private information for an individual as $D=(D_1,D_2,...,D_m)$, D_i $(i=1,2,...,m)$ represents the specific type of private information such as name, age, gender, etc.

It is obvious that the importance degree of private information is different in various information domains, we assume l_i^k as the importance degree for information D_i in domain k .

The more private information a person exposed, the more inconvenience and unsafety he or she has. So, we define the discernibility cost, which means the ratio of our exposed PI accounting for the total number of information:

$$\eta_{1i}=\frac{1}{|D|}\ (i=1,2,...,m)$$

where, $|D|$ mean the total amount of information one person has. Then we define the importance degree of PI exposed to individual;

$$\eta_{2i}=\frac{l_i^k}{\sum_{i=1}^{m} l_i^k}$$

The more importance the information has, the more risk one person should undertake when it exposes.

Next, according to the references, we give the definition of risk: [2]

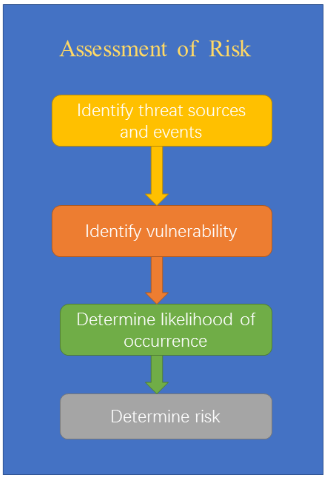

Assessment of Risk
Identify threat sources and events
Identify vulnerability
Determine likelihood of occurrence
Determine risk

Fig. 1 Schematic – The Process Defining Risk

Firstly, we introduce treat factor, which means that threat sources and some events which has influence on sellers themselves. We set the treat factor as $\{tf_i^k\}$ $(i=1,2,...,n)$. Based on the treat factor, we need to identity vulnerabilities existed in PI, then we consider the set of vulnerabilities in specific domain $k:\{vul_j^k\}$ $(j=1,2,...,l)$.

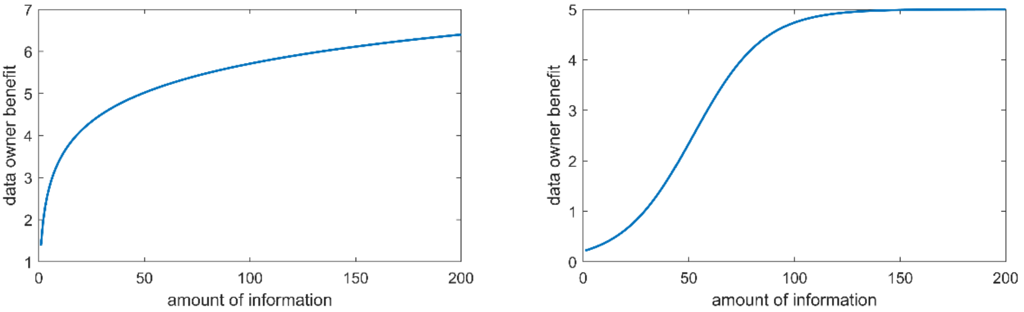
Based on the risk factors, we can calculate the likelihood of vulnerabilities occurrence. It is directly combined by amount of PI and importance degree of PI. So, we define likelihood of vulnerabilities occurrence of information D_i in domain k:

$$\Delta_q p(vul_j^k \mid tf_i^k) = C_k \eta_{1q} \eta_{2q}$$

where, C_k is a constant number.

Integrating all the likelihood of vulnerabilities of a person in domain k, we can calculate the **risk level** of one person:

$$p(vul_j^k \mid tf_i^k) = \sum_{q=1}^{m} \Delta_q p(vul_j^k \mid tf_i^k) = \mathrm{C}_k \sum_{q=1}^{m} \eta_{1q} \eta_{2q}$$



(a) (b)

Fig. 2 Payment Schemes

According to a previous research [3], the author proposed that a probable payment scheme can be expressed as the figure (a). We improve this scheme by introducing logistic function (b). Because when people sell a little amount of their information, they will get a little payment for the uselessness of their data. However, when more data obtained, people have right to get more payment. Besides, it is logical that the payment should be finite. Thus we use logistic function as gain function of PI.

In order to evaluate the benefits or interests when one person is willing to sell private information, we use **logistic function** to describe it.

$$G_i^k = \frac{C_G}{1 + e^{-\frac{(\eta_1 - \mu_1) + (\eta_2 - \mu_2)}{\lambda}}}$$

The function means the more information sold with conresponding importance degree, the greater benefits and interests one gains. Because the benefits have limit and do not increase infinitely, we use logistic function to restrict its growth. μ_1, μ_2 is the threshold which means that only when the amount of PI on transactions and importance of PI reach to a certain level, benefits can be received by sellers. C_G is a constant value which is used to adjust the revenue. λ is used to adjust the influence of amount and importance degree of PI on individual revenue.

In order to develop a price point for sellers, we need to comprehensively consider the risk and benefits. Because risk level must be less than a certain threshold, the amount of PI and importance degree of PI must be less than a threshold correspondingly. The benefits also have limit due to these restrictions. We consider the limit of benefits as **price point**.

Results:

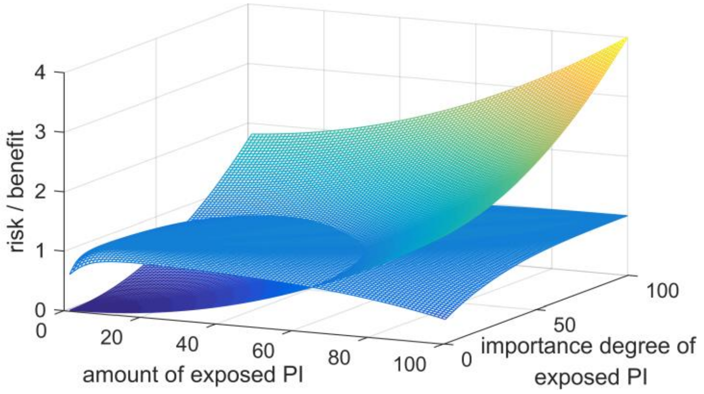


Fig. 3 Intersection of Risk and Benefit: Curve of Price Points

According to the amount of exposed PI and importance degree of it, the risk level and benefit for one individual is calculated. The two surfaces' intersecting line refers to the curve of price points.

4 Task 2: Price Model

4.1 Weight of Elements

Privacy is categorized into 2 types: 1) identity information, and 2) specific domain privacy (on social media, behind financial transaction, and in health/medical records). For type 1 together with the privacy on social media, a social survey [4] gives attitudes of the mass toward these information, which can be considered average, objective privacy importance. Not finding relevant researches, for the financial-transaction and medical-records privacy, we use Analytic Hierarchy Process (AHP) method to compare each two elements at a time, based on subjective experience. The survey result and the AHP method separately contribute to the weights of each factors.

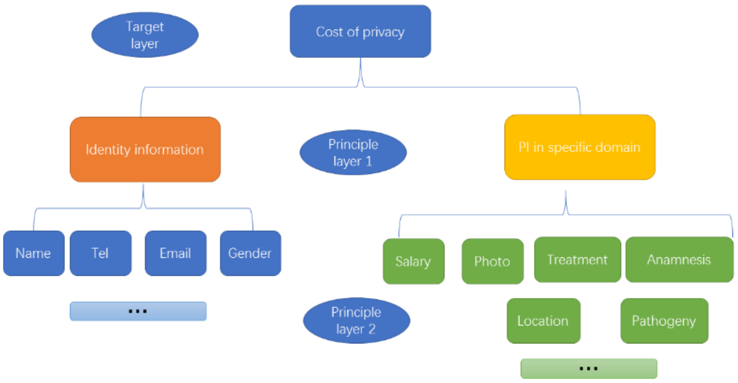


Fig. 4 Analysis Hierarchy Process

The comparing matrixes of AHP is set as

	Time	Attribute	Purpose	Amount	Opposite side	Balance
Time	1	3	1	1 / 3	1 / 2	1 / 3
Attribute	1 / 3	1	1 / 2	1 / 3	1 / 4	1 / 6
Purpose	1	2	1	1 / 2	1	1 / 5
Amount	3	2	2	1	1	1 / 2
Opposite side	2	4	1	1	1	1 / 3
Balance	3	6	5	2	3	1

	Pathogeny	Examination results	Treatment	Medicine	Anamnesis	Family illness history
Pathogeny	1	1	2	1	1 / 4	1 / 3
Examination results	1	1	3	2	1 / 3	1 / 2
Treatment	1 / 2	1 / 3	1	2	1 / 5	1 / 4
Medicine	1	1 / 2	1 / 2	1	1 / 6	1 / 5
Anamnesis	4	3	5	6	1	2
Family illness history	3	2	4	5	1 / 2	1

The two matrix's *CI* values are 0.0341, 0.0377 << 1.24 = *RI*(6). The literature and our AHP method returns the following result.

Table 1 Privacy Elements and Weights

Identity	Name	Tel	Email	Gender	Age	Education
weight	0.440	0.420	0.025	0.077	0.023	0.018
On Social Media	Salary	Location	Photo	Hobby	Occupation	
weight	0.179	0.494	0.193	0.032	0.102	
Behind Financial Transaction	Time	Attribute	Purpose	Amount	Opposite side	Balance
weight	0.102	0.049	0.104	0.197	0.161	0.387
In Medical Records	Pathogeny	Examination	Treatment	Medicine	Anamnesis	Family illness history
weight	0.098	0.131	0.068	0.060	0.386	0.256

4.2 Cost of privacy

According to the weights we obtained, we can calculate gain function. When we conduct private information transaction, we need to weigh the pros and cons of gains and losses. Too much private information exposed will bring both benefits and risk. So, we take tradeoffs and risks into account to evaluate cost of privacy in different domains.

- Risks: Private information itself has potential risk to be exposed and cause damage to owners consequently. In specific domains, people with high risk level have a higher likelihood to be exposed because their PI is more sensitive and more important than others.
- Tradeoffs: in real transactions, sellers need to take gain and loss into account to evaluate its cost. By weighing up gains and loss, sellers could have a better knowledge of their actual cost revenue. Here we treat G_i^k as tradeoff.

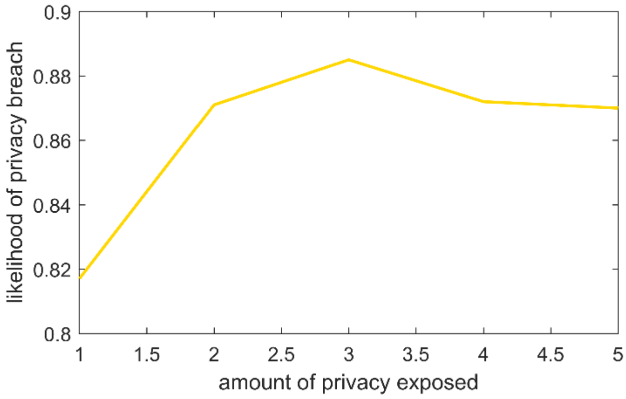


Fig. 5 Resembling Log-function: Risk to Exposed Information

According to a survey [5] in previous time, the author proposed the relationship between likelihood of privacy breach and amount of privacy exposed. We will later define the risk function as a **log function**, based on this shape.

Then we give mathematical formula of risk:

According to the concept of information entropy and relevant references, we can quantify the risk of information exposure.

We assume the i th threat factor in domain k as tf_i^k, the j th vulnerability in domain k as vul_j^k.

Then we get $risk_{ij}$

$$risk_{ij} = \ln(1 - p(vul_j^k \mid rf_i^k))$$

Accordingly, we get **total risk** for all PI of an individual.

$$risk = \sum_{i=1}^{n}\sum_{j=1}^{l} a_{ij} risk_{ij}$$

where a_{ij} means the adjustable coefficient of risk.

We give the definition of tradeoff considering the gain function into account, for information D_i in specific domain k:

$$Tradeoff_k = b_i G_i^k$$

where b_i means the adjustable coefficient of tradeoff.

Then we assume that multiple pieces of information of one person are sold together, so tradeoff will be scaled larger under the effect of relevance degree of information.

$$Tradeoff = e^{\sum_{i \neq j}^{m}\sum_{j}^{m} R(D_i,D_j)} \sum_{i=s}^{t} b_i G_i^k$$

where, $R(D_i,D_j)$ means the relevance degree of information D_i, D_j

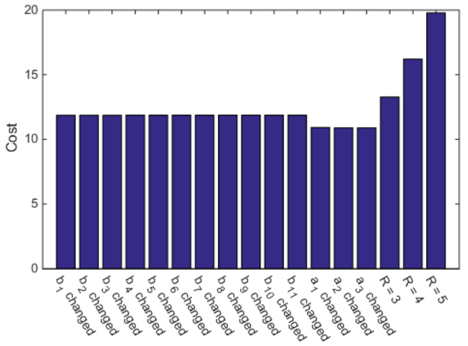
It is obvious that if different kinds of PI are sold together, the cost will be higher. For instance, the value of combination of name and figure is higher than value of name or figure alone. Therefore, we can calculate the correlation degree between two types of PI:

$$R(D_i,D_j) = l_i^k \cdot l_k^k \cdot (|D_i| \cap |D_j|)$$

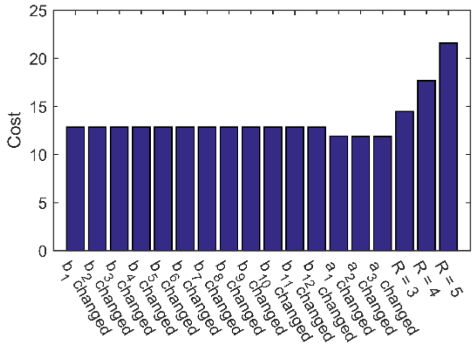
when information D_i is sold, $|D_i| = 1$, otherwise $|D_i| = 0$.

Finally, we get the cost of privacy, which containing tradeoff and risks, and it means that the **cost of privacy** comes from the exposure risk of PI and revenue obtained.

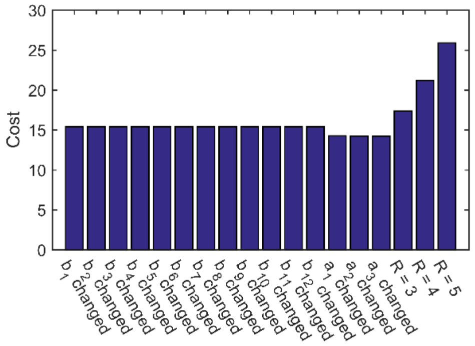
$$Cost = Tradeoff + Risk = e^{\sum_{i \neq j}^{m}\sum_{j}^{m} R(D_i,D_j)} \sum_{i=s}^{t} b_i G_i^k + \sum_{i=1}^{n}\sum_{j=1}^{l} a_{ij} risk_{ij}$$



(a) social media



(b) financial transaction



(c) health/medical records

Fig. 6 Impacts by Changing Parameters

According to the model for cost of privacy, we calculate the cost in social media, financial transactions and health/medical records. We change weights of tradeoff (represented by b_i), weight of risk (represented by a_i), and weight of relevance (represented by R) in turn, and under a same numeric level. From the figures above, changing weights of tradeoff, risk and relevance are all able to modify the cost, where, altering the relevance degree of PI will trigger the cost significantly. The relevance of information play an important role in cost.

5 Task 3: Price System

When private information is commercialized and converted into a commodity, the seller needs to enter the market for the transaction. Taking into account the market fluctuations, relationship between supply and demand, we set up a pricing system for sellers, groups and countries. With the guidance of game theory and pricing theory, we discuss two game processes: individuals-groups transaction and groups-nations transaction.

We assume that the both partners of decision- making in the game all know strategies of each other completely. From an objective point of view, buyers need to consider the utility value of information with respect to a certain price.

We uniformly define the cost of privacy as C, sale price as R, utility value of PI as U ,we assume that PI can be sold from individuals to groups or from groups to nations.

Personal information was divided into two categories: identity PI and PI in specific domain according to task 2. We assume the set of identity PI is $\{x_1, x_2, ..., x_n\}$, corresponding price is $s_1, s_2, ..., s_n$ .Similarly, we assume the set of PI in specific domain is $\{y_1, y_2, ..., y_m\}$, corresponding price is $t_1, t_2, ..., t_m$.

Game process 1: individuals-groups.

We assume the cost of privacy as

$$C_1 = \varepsilon Cost$$

ε is an adjustable coefficient of cost of privacy, The value of ε will increase with the increase of importance degree of PI. We assume the initial value of ε as 1. Then we get the price of PI:

$$R_1 = \sum_{i=1}^{n} s_i + \sum_{j=1}^{m} t_j$$

For individuals, they are willing to get the max revenue with certain limit of risks they are to undertake. Sellers do not want to expose too much information to avoid causing too much damage, according to analysis above, we obtain the following **optimization objective**:

$$\begin{aligned} &\max\ R_1 - C_1 \\ &\text{s.t.}\ \ Risk<L \\ &\qquad \frac{1}{|n-m|} < K_1 \\ &\qquad s_i > 0, t_j > 0 \end{aligned}$$

where, K_1 is a constant value. n, m is the amount of PI.

Then we define the information utility value for groups:

$$U_1 = \left(\prod_{i=1}^{n} l_{x_i}^{\frac{o_i}{100}}\right)\left(\prod_{j=1}^{m} l_{y_j}^{\frac{r_j}{100}}\right)$$

o_i, r_j separately means the preference degree of groups to identity PI and specific domain PI.

For groups, they want to obtain the max revenue from the utility value of PI with restricted cost. Buyers not only want to obtain more PI, but also hope to get the information which has great relevance with each other, so we develop this **objective** as following:

$$\begin{aligned} &\max\ U_1\text{-}R_1 \\ &\text{s.t.}\ \ R_1<Pr \\ &\qquad \sum_{\substack{i=1\\ i\neq j}}^{m+n}\sum_{j=1}^{m+n} R(D_i, D_j) > K_2 \\ &\qquad m+n>K_3 \\ &\qquad s_i > 0, t_j > 0 \end{aligned}$$

where, Pr means the prime cost for groups, K_2, K_3 are constant value.

Based on the optimization objectives we develop, we can obtain the **Nash equilibrium** of the game and get the optimal strategy for individuals and groups. As a result, we can obtain the price of PI.

Game process 2: groups-nations

Firstly, groups buy PI from p individuals, then we get the cost of PI for groups:

$$C_2 = pR_1$$

The price for nations:

$$R_2 = p(\sum_{i=1}^{n} s_i^{'} + \sum_{j=1}^{m} t_j^{'})$$

Groups want to get the max revenue under the limit of cost. At the same time, they also want to sell PI as more as possible.

$$\begin{aligned} &\max\ R_2 - C_2 \\ &s.t.\ \ C_2 < De \\ &\qquad p\cdot(m+n)>K_4 \\ &\qquad s_i^{'} > 0, t_j^{'} > 0 \end{aligned}$$

where, De, K_4 are constant value.

For nations, the utility value of PI:

$$U_2 = \left(\prod_{i=1}^{n} l_{x_i}^{\frac{o_i}{100}}\right)^{\alpha}\left(\prod_{j=1}^{m} l_{y_j}^{\frac{r_j}{100}}\right)^{\beta}$$

where α, β represents national effects. Nations want to get max benefit with the limit of cost. Nations not only want to obtain more PI, but also hope to get the information which has great relevance with each other.

$$\begin{aligned} &\max\ U_2 - R_2 \\ &s.t.\ \ R_2 < De \\ &\qquad \sum_{\substack{i=1 \\ i\neq j}}^{m+n} \sum_{j=1}^{m+n} R(D_i, D_j) > K_5 \\ &\qquad p\cdot(m+n) > K_4 \\ &\qquad s_i^{'} > 0, t_j^{'} > 0 \end{aligned}$$

where, K_5 is a constant value.

The above Game Model proposed a problem of two optimization goals. The model includes two kinds of variables unsolved. The first is how many components of a certain privacy (a vector) join in the dealing. The number of identity elements is *m*, the number of specific domain privacy elements *n*. The second is the price respectively corresponding to each element, which sums to R_1.

The two goals of this model is contradictory, in which an increase of the goal 1 is accompanied with a decrease of the goal 2, so there is no single optimum solution to achieve each goal. In such case called multiple objectives, the non-inferior (undominated) solution is defined to refer to a set of solutions under which no better states can be found. For instance, the process of Pareto optimality is to optimize at least one goal, on the condition that the other goals are not harmed.

Carlos et al. [6] improved the **Particle Swarm Optimization** (PSO), originally used in single-objective optimization, changing the mechanism particles searching the vector space. By using a particle set called secondary repository, the dominated particles guide their own flight in later iteration.

To get it adjusted to our model, a vector named *c*[*i*] is introduced, in order to fit the **Integer Programming problem**. Its value, 1 or 0, represents whether or not an element *i* is sold. *c*[*i*] is updated based both on its history and randomness. It tends to be valued as the number that it is more often valued. The main algorithm in our work is like

Table 2. Multi-Objectives Particle Swarm Optimization

```
1. initialize the population POP, position x, velocity v
        FOR i = 1 : Max = 400
            initialize POP[i]
            initialize x[i] % which is an element's price.
            initialize c0[i] % 1 refers message i being sold; 0 refers not being sold.
            initialize v[i]
2. evaluate each particles
        FOR i = 1 : Max
            score(POP[i]) = -1•[c[i]• x[i]^T -C1 , U1 -c[i]• x[i]^T]
            IF IsSatisfyingConstraints(POP[i]) == 0
                score(POP[i]) = [-Inf , -Inf]
3. store the positions of undominated particles in repository REP
        h = 1;
        FOR i = 1 : Max
            IF isdominated(POP[i]) == 0
                REP[h] = POP[i]
                h ++
4. initiate memory of each particle
        FOR i = 1 : Max
            p_best [i] = POP[i]
5. iteration loop, updating particles
        WHILE k < k_max
        DO
                h = round(length(REP)•rand)
                FOR i = 1 : Max
        $v[i] = w \cdot v[i] + r_1 \cdot (p\_best[i] - POP[i]) + r_2 \cdot (REP[h] - POP[i])$
        $x[i] = x[i] + v[i]$
        $c_k[i] = \text{round}(\frac{\sum_{j<k} c_j[i]}{\text{length}(c_0[i])} \cdot \text{rand})$
        $\text{score}(POP[i]) = -1 \cdot [c_k[i] \cdot x[i]^T - C_1, U_1 - c_k[i] \cdot x[i]^T]$
        IF IsSatisfyingConstraints($POP[i]$) == 0
            $\text{score}(POP[i]) = [-\text{Inf}, -\text{Inf}]$
        IF $\text{score}(POP[i]) < \text{score}(p\_best[i])$
            $p\_best[i] = POP[i]$
    $h = 1$;
    FOR $i = 1 : \text{Max}$
        IF isdominated($POP[i]$) == 0
        $REP[h] = POP[i]$
        $h$ ++
    $k$ ++
ENDWHILE
```

The 400-iteration algorithm returns back a **Pareto Curve** for the game model, as shown in Fig. 7. It is evident that all 50 particles generally moving close to a certain line, which means the stable solution of the variables is a pair of $[m,n]$ with fixed values. With deeper analysis, it appears this optimized state is the one that all privacy components are included in the dealing. For other cases such as privacy in health/medical records, or a deal between a group and a country, the curve's tendency is totally the same, albeit different positions.

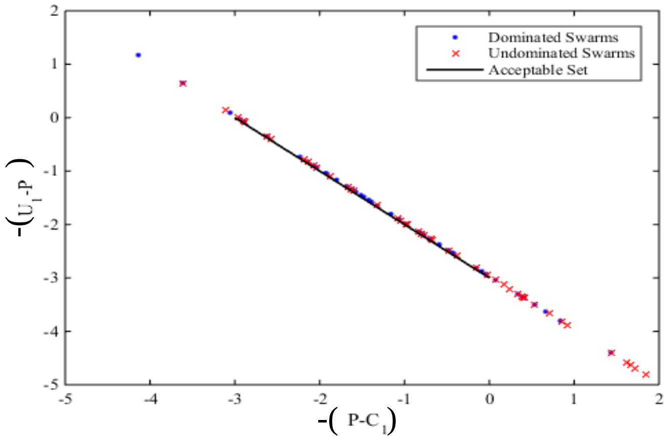


Fig. 7 Pareto Curve for Pricing Financial Transaction Privacy, between individual and group

What is different between these separated states (points in the figure) is, their corresponding selling price are diverse. Any of them is a non-inferior solution. To choose an exact price, we consider a compromise between seller and buyer, as well as laying particular stress on a certain side. We determine the selling price P to be 18.17 for the dealing between individual and group, and 25.02 for the dealing between group and country, with respect to the transaction privacy. The value is determined by making the net profit increase 5% comparing to the center point, for the favored side.

Considering the relationship between supply, demand and price, we try to calculate the force of supply and demand using **gravity model and Newton's second law**.

From the perspective of Newton's second law, we analysis the influence of market supply and demand. We assume $m_{country}$ as the demand of nation to PI, m_{group} as the amount of PI which groups can offer to nations, $a_{country1}$ as the change rate of identity PI price over time in process 2, a_{group1} as the change rate of identity PI price over time in process 1. $a_{country2}$ as change rate of specific domain PI price over time in process 2, a_{group2} as change rate of specific domain PI price over time in process 1.

Only considering country and group, for the identity PI, the relationship between force and change rate is:

$$\frac{m_{country} \cdot m_{group}}{e^{-o_i} \cdot s_i^{'}} = m_{country} \cdot a_{country1} = m_{group} \cdot a_{group1}$$

For the specific domain PI, the relationship can be expressed as:

$$\frac{m_{country} \cdot m_{group}}{e^{-r_j} \cdot t_j^{'}} = m_{country} \cdot a_{country2} = m_{group} \cdot a_{group2}$$

Using the change rate, we can update the price in process 1 and process 2 as follows:

$$s_i^{'}(t) = s_i^{'}(t-1) + a_{country1}$$

$$t_j^{'}(t) = t_j^{'}(t-1) + a_{country2}$$

$$s_i(t) = s_i(t-1) + a_{group1}$$

$$t_j(t) = t_j(t-1) + a_{group2}$$

When individuals have control to sell their own data, we assume that when individuals increase the price of PI, groups are more willing to sell them to nations than directly use them.

In the game, we can give different weights for groups to deal with PI:

$$\lambda \max\ (U_1\text{-}R_1)\ \ (0<\lambda<1)$$

$$\mu \max\ (R_2\text{-}C_2)\ \ (\mu>1)$$

The force between supply and demand can explain: when the price is increasing, the attracting force and changing rate will decrease, and vice versa.

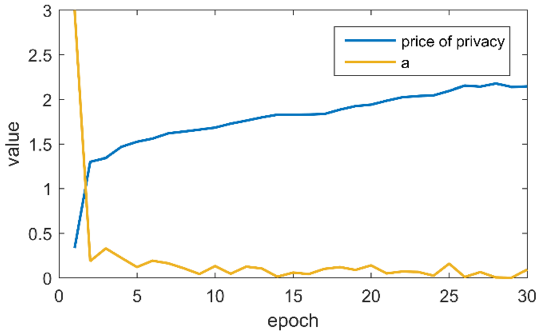


Fig. 8 Cost of Privacy – Epoch under Attraction Force

The blue line shows if attraction force is positive, the price of privacy will increase over time, and the tangerine line shows a decrease tendency of acceleration (change rate of price).

6 Task 4: Dynamic Belief-spreading Model

Constrictions:

- In task 3, we formulate price according to market economic system. However, some certain records are subject to macro-control from nations. Due to lack of data, we do not quantify the influence of adjustment from nation.
- the dynamic change of price is not taken into consideration.
- People from different cultures have different attitudes about privacy, the Price fluctuations due to cultural differences is not taken into consideration.

In real life, the price is not only influenced by the market supply and demand, but also will be subject to some macro-control from the country. In order to maintain social stability and improve market prices over instability, the country will adopt certain policies and regulations to adjust prices. For the sake of simulating and observing privacy selling from the scope of all over the world, we set up a dynamic spreading system to describe the change of cognition or attitudes about PI by means of **susceptible-infection-susceptible** (SIS) model in complex networks [7].

Epidemic models are used to describe the situation of illness spreading in real networks. On the basis of contacting between each other, not only the pathogens, but also cultures, cognitive concepts and beliefs will spread one by one. For the people in real networks, whether they are willing to sell their privacy depends on their beliefs, cognitive concepts and attitudes. According to SIS model, we divide the people into two categories：willing to sell and unwilling to sell. We assume that the attitude of being willing to sell can spread in the networks. The dynamic element we introduced is population who are willing to sell their privacy.

We assume the total global population as N, population who is willing to sell their privacy as I_P, population who is unwilling to sell their privacy as $N-I_P$.So, the proportion of the population who is willing to sell is:

$$i_P=\frac{I_P}{N}$$

The proportion of the population who is unwilling to sell is:

$$1-i_P=1-\frac{I_P}{N}$$

Then we can assume the attitudes of being willing to sell could be transmitted from unwilling people to willing people in a unit time as β_P .The willing people will transfer to unwilling people at a rate: μ_P. In the complex network, we can assume an individual has $\langle k\rangle$ contacts, which means the average degree of node in network.

So, we can get dynamic equations:

$$\frac{dI_P}{dt}=\beta\langle k\rangle\frac{I_P(N-I_P)}{N}-\mu_P I_P$$

Divide both sides by N at the same time：

$$\frac{di_P}{dt}=\beta_P\langle k\rangle i_P(1-i_P)-\mu_P i_P$$

Then we solve equations by integrating both sides and get answers as follows:

$$i_P(t)=(1-\frac{\mu_P}{\beta_P\langle k\rangle})\frac{Ce^{(\beta_P\langle k\rangle-\mu_P)t}}{1+Ce^{(\beta_P\langle k\rangle-\mu_P)t}}$$

C is a constant value, and when $t=0$, $i=i_0$:

$$C=\frac{i_0}{1-i_0-\frac{\mu_P}{\beta_P\langle k\rangle}}$$

According to our model, $i_P(t)$ reflects the privacy protect concept. The larger of its value is, the weaker of people’s protect concept for privacy. People who has weak protect concept are always willing to sell their PI. With more and more people change their concept of privacy, the value of tradeoff will change as follow:

$$tradeoff_P=tradeoff\cdot i_P$$

In turn, the value of μ_P,β_P will be influenced by tradeoff and risk.

$$\beta_P=e^{-tradeoff_P/risk}$$

$$\mu_P = e^{-risk/tradeoff_P}$$

We initiate value $i_P(0) = i_0$, μ_P=μ_0, $\beta_P = \beta_0$, according to equations above , we can calculate the value of $i_P(1)$ and update the value of μ_P and β_P. Likewise, we can update the value of each variable during the loop in algorithm and get $i_P(t)$ at arbitrary time.

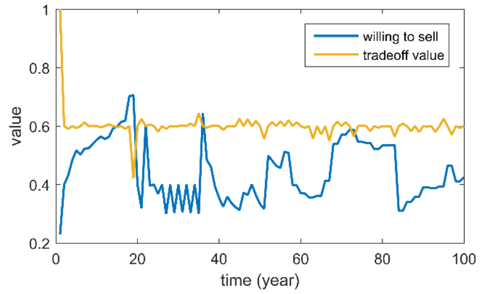


Fig. 9 Changing Attitude with Time

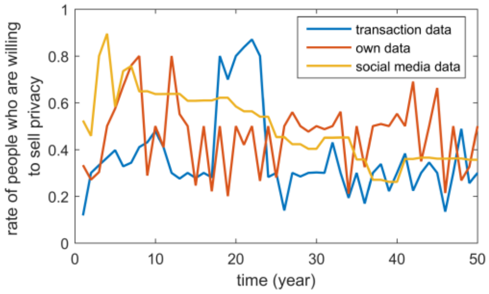


Fig. 10 Changing Attitude in 3 Domains

In order to compare the proportion of people who are willing to sell, we need to make a certain scaling. The blue lines of Fig. 9 shows the proportion of population willing to sell privacy with respect to years. The tangerine line shows the change of tradeoff over time. From the chart above, it is concluded that the value of tradeoff is relatively stable and the attitudes towards PI have a large fluctuation.

Fig. 10 shows a changing attitudes towards specific domains in PI under different ages. The proportion of people willing to sell privacy is stable within a certain range.

7 Task 5: Generational Differences in PI

People of different generation may have very different attitudes toward privacy. More and more adolescents may hold an open attitude towards privacy and they are always willing to share their privacy in social media or exchange it for interests. On the contrary, traditional seniors prefer to protect their privacy.

In order to find generational differences in attitudes or beliefs toward private information, we take ten years as a generation and push forward for a hundred years, put other word, we need to observe change in 10 future generations. With generation ages, there are three kinds of varieties:

With the development of social media, more and more PI will be exposed in all kinds of domains. Privacy sales and privacy leaks will become more and more common, which may cause the disparity of l_i^k gradually decreasing over the time

Correlation degree among PI exposed will increase, so the value of $R(D_i, D_j)$ will increase.

According to the changes above, we can update proportion of population who is willing to sell: $i_P(t)$ for each generation, then we can update tradeoff for each generation.

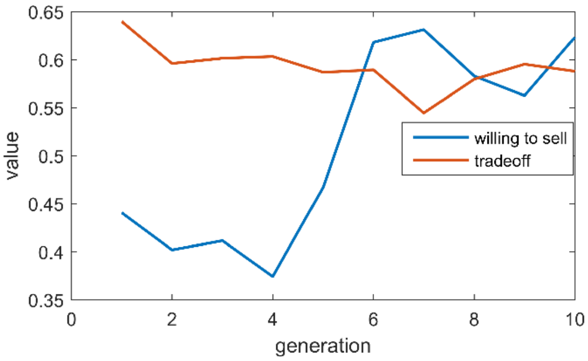


Fig. 11 Changing Attitude and Tradeoff with Generation

Red line shows the change of tradeoff value over generation, and blue line shows the proportion who are willing to sell PI. Through simulation, we can find the gap among the importance of various information is narrowing in concept with generation aging. People are increasingly willing to sell PI with generation aging. In addition, we can again ensure that the tradeoff is stable over time.

Table 3 Comparison among PP, IP & PI

	Personal Property	**Intellectual Property**	**Private Information**
similarity	1 They all have potential value for future and entity value now. 2 They all contain parts of value with timeliness: Currency will devalue over time (PP). The validity of intellectual property is limit (IP). Medical records are time-limited(PI)		
difference	Without Timeliness such as pension	Without Timeliness Transfer of ownership	With Timeliness No sell ownership

8 Task 6: Measuring Network Effect

In this section we consider people’s social network in community. Through continuous interaction in the real world, people's attitudes, concepts, and PI prices may change dramatically. It is assumed that individuals living in the same community have strong homogeneity, and their behaviors interlink with each other.

Due to the evolution and development of social networks, a large amount of private information is publicly shared. Information from one may unintentionally have a subtle correlation with PI from others. For instance, a group photo sharing on Facebook also includes the appearance information of others. Therefore, those who have larger number of friends, put another way, have more connection with others, will have a better knowledge of others’ PI. They are called “**hubs**”. A hub is more likely to become a monopoly, thus affect the market price.

In real social networks, hubs are not common. On the contrary, nodes with small degree can be easily observed, which means most people have a relatively limited connection with the community. According to the survey [8], in social network the distribution of a node’s degree meets the power law, which we call as scale-free network. On the basis of characteristic of social network, we can build a network with network effects of sharing data.

We assume that there exist N people in community, considered as nodes in networks. $cost_i$ means the cost of privacy for people i, because PI which one person share may contain extra PI from others. Put another way, what one person share consists of his/her own PI, together with others.

Then, we define a distance d_{ij} from node i to node j, which represents intimacy between two people. Note that the closer two people are, the more information from one will be covered in the other's sharing. In the network we build, if the degree of node i is k, the amount of other's PI being covered is Φ_i^k,which can be calculated as:

$$\Phi_i^k = \Psi_i \cdot D_i + cost_i$$

where, $\Psi_i = [cost_1, cost_2, ..., cost_{i-1}, cost_{i+1}, ..., cost_N]$, means the initial cost of privacy which merely contain his or her own PI. $D_i = [\frac{1}{d_{1,i}}, \frac{1}{d_{2,i}}, ..., \frac{1}{d_{i-1,i}}, \frac{1}{d_{i+1,i}}, ..., \frac{1}{d_{N,i}}]$, which means reciprocal distance from neighbors to node i [9].

Because social networks have different aggregation factors, a large community can be divided into some sub-communities. Considering that people are very intimate in a sub-community, PI that one share has great influence in sub-community and even in entire community.

We can use three-dimensional vectors (x_i, y_i, Φ_i^k) to represent the position and the amount of PI which a node has. We assume that distance between two nodes represents the level of intimacy between two people. Therefore, if distance between two nodes

$$\sqrt{\left|x_i - x_j\right|^2 + \left|y_i - y_j\right|^2} < \varepsilon$$

These two are considered to lie in a same sub-community.

Next, supposing there exists n people and c sub-communities in the entire community, the position of a sub-community is (β_m, γ_m) which can be calculated as:

$$\beta_m = \frac{\sum_{i=u}^{u+n-1} x_i}{n}$$

$$\gamma_m = \frac{\sum_{i=u}^{u+n-1} y_i}{n}$$

So the distance between sub-community m and sub-community l is:

$$D_{ml} = \sqrt{(\beta_m - \beta_l)^2 + (\gamma_m - \gamma_l)^2}$$

If sub-community l shares PI to the entire community, the network effects are proportionally allocated to the remaining sub-community based on distance between two sub-communities and amount of PI of sub-community. Thus, the **network effects** can be expressed as:

$$\Delta_m = \frac{\sum_{i=u}^{u+v} \Phi_i^k}{\sum_{i=1}^{c} \Phi_i^k} \cdot \frac{1}{D_{ml}}$$

Then we calculate the network effects on individual in a sub-community

$$\Delta_{m\_i} = \frac{\Phi_i^k}{\sum_{i=u}^{u+v} \Phi_i^k} \cdot \Delta_m$$

For the entire community, the initial total cost of privacy for N people is

$$cost_c = \sum_{i=1}^{N} cost_i$$

After considering the social connections among people, we update integrated cost of privacy for people i with degree k .

$$\cos t_i^k = \frac{I_i^k}{\sum_{i=1}^{N} I_i^k} \cdot \sum_{i=1}^{N} I_i$$

The **integrated cost** of privacy contains PI which comes from itself and other connected nodes.

It is obvious that people with large degree has large integrated cost of privacy whom we consider as hubs. Purchasers will be more willing to purchase PI from hubs who help to obtain more PI with lower price. Therefore, we search for the node with the largest degree. Considering the complex relationship in a large community, we firstly divide them into some sub-communities. In a sub-community, we get the hub as follows:

In the scale-free network, the degree's distribution meets the power law:

$$p_k = Ce^{-\lambda k}$$

We assume that the smallest degree that a node can get is 1, namely $k_{\min} = 1$, which means there is no isolated node in network:

$$\int_{k_{\min}}^{\infty} p(k)\mathrm{d}k = 1$$

Then we find the node with max degree, suppose $k_{\max}$ is the largest degree in this sub-community, which can be calculated by

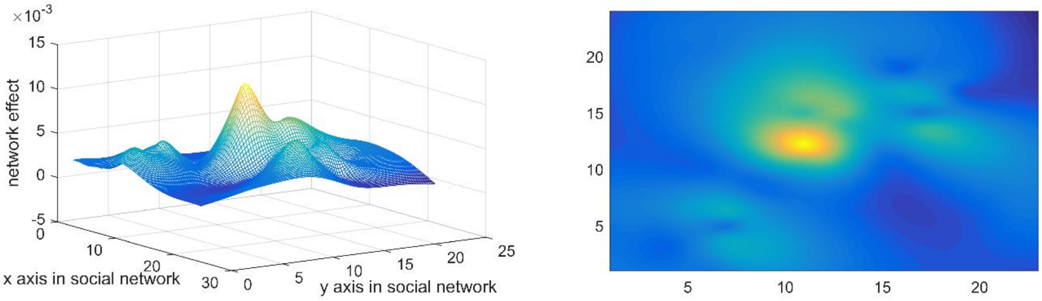
$$\int_{k_{\max}}^{\infty} p(k)\mathrm{d}k = \frac{1}{N}$$

So

$$k_{\max} = k_{\min} + \frac{\ln N}{\lambda}$$

Obtaining a node with $k_{\max}$ in sub-communities, we successfully get a hub in the entire community.

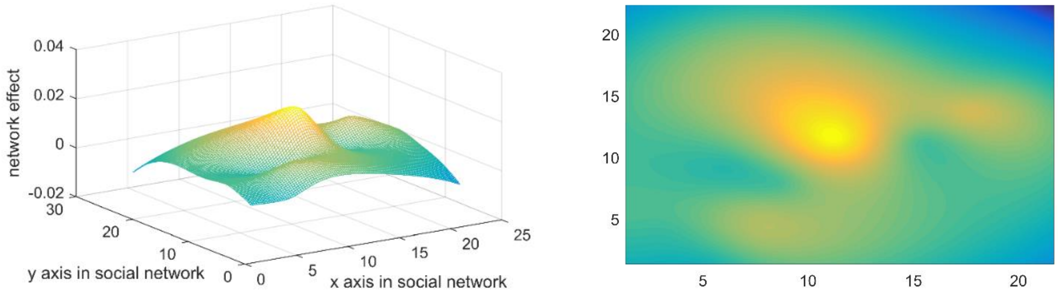
In our pricing system, when the monopolists appear, they will affect the price of the entire market. In the game of transactions, for individuals, due to the change of cost after considering network effect, the price for groups will correspondingly change. For entire community, purchasers could buy PI from hubs who has a large amount of information. As for larger ranges, big customers are also willing to buy PI from communities with large contacts with other communities.



(a) (b)

Fig. 12 Network Effect **on Individual** of Sharing Information

Fig. 12(a) shows a kind of network effect in social network: when one person shares PI, its neighbors will be affected to some extent. The highest peak represents the person who shares PI with the maximum extent of PI exposure, and other peaks the extent of PI exposure for neighbors. In Fig. 12(b), the hotter the color, the stronger the network effect.



(a) (b)

Fig. 13 Network Effect **on Community** of Sharing Information

In Fig. 13(a), the highest peak refers to the position of the community sharing PI, and other peaks the neighbor influenced by network effects. In (b), the hotter the color, the more network effect it receives.

9 Task 7: Sensitive Analysis: the Dark Web

In recent years, information security has become the focus of the world's attention. Although many countries have legislation on privacy breaches, there are still many ways of trafficking and infringing the privacy of others. Dark web [10] is a kind of illegal web with onion browsers [11]. When millions of personal PI are sold on dark web illegally, it will affect our price system and attitudes of sellers.

Firstly, we assume that a dark web steals and sells parts of PI from n community, the proportion of information stolen in the i^{th} community is $rate_i$. Feared by PI breaching, people in a community will reinforce their own privacy protect concepts and be sensitive to sell privacy. From the perspective of attitudes of sellers, at time *t*, the transfer rate from sellers to protecters is μ_P, the transfer rate from protecters to sellers is β_P. At time *t*+1, the transfer rate from sellers to protecters goes to $(1+rate_i)\mu_P$, the transfer rate from protecters to sellers goes to $(1-rate_i)\beta_P$. The proportion of sellers, i_P can be calculated.

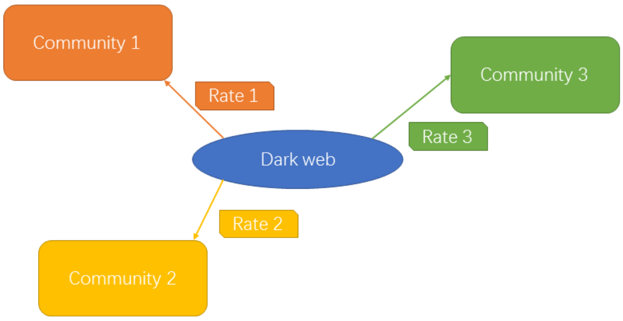


Fig. 14 Schemetic – Network Structure

In our pricing system, if millions pieces of stolen privacy are sold on dark web, $tradeoff_P$ and $cost_P$ will change as i_P changes. According to the consideration above, we can get the loss of people, compared to the reveune before PI is stolen and sold on dark web.

Assuming that total cost of privacy in a community as: $\sum_{i=f_1}^{f_2} cost_i^k$, then we get the **loss** of people i compared with reveune before PI stolen and sold on dark web.

$$Loss_i = \sum_{i=f_1}^{f_2} cost_i^k \cdot rate_j \cdot \frac{\Phi_i^k}{\sum_{i=1}^{N} \Phi_i^k}$$

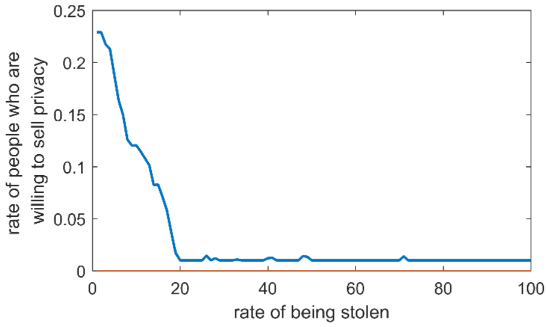


Fig. 15 Willing-Stealing Relation

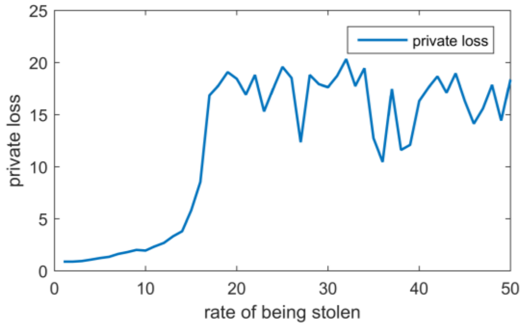


Fig. 16 Loss-Stealing Relation

The proportion of people who are willing to sell privacy decreases when it is easily stolen in dark webs. We conclude that when the rate of being stolen is more than 20%, the willing-to-sell rate goes down forward close to 0. What is stimulated by the stealing activity is the privacy loss.

10 Conclusion

- According to risk and gain function, we develop a price point which is restricted by risk.
- Constantly updating cost of price via altering weight of tradeoff, risk and correlation degree, we conclude correlation degree has the greatest impact on cost.
- Based on analysis hierarchy process, we endow weight of elements in different domains. We find the elements with highest value separately in identity, social media, health/medical records are telephone number, location, anamnesis.
- Positive forces of supply and demand will increase the price of privacy and consequently affect the value of acceleration.
- Under our assumptions that the privacy protection beliefs become weaker, with generation aging, the proportion who is willing to sell privacy will increase dramatically. However, the value of tradeoff maintains stable over time.
- Since intimacy among people, the network effect is obvious during the process of data sharing.
- The more data breach in dark web, the stronger privacy protection belief is.

11 Strengths and Weaknesses

Strengths:

- We help sellers to balance their risks and benefits of selling PI and make a reasonable cost of privacy for them in different specific domains.
- Based on game theory, we build a pricing system for individuals, groups and nations. Considering the demand and supply of market, we help the parties who buy and sell financial-transaction information to get their optimum strategy.
- Considering the population distribution in social network, we use scale-free network to evaluate the impact of social connection on our model. Also, spreading of attitudes to PI are simulated with spread model.

Weaknesses:

- Due to time constraints, we lack a large amount of relevant data in the privacy field to support our model. Not only that, some personal privacy information is hard to find or evaluate.

- Not having to quantitatively consider the impact of authority’s macro-control on market prices, we just conduct such analysis.
- Owing to lack of data, subjective method such as analytic hierarchy process is used to solve the problem. Subjective methods have personal biases, which in turn leads to disjunction with reality.

References

[1] APABourne, V. (2016). Open source is fuelling innovation and cost savings in uk businesses. *Software World*.

[2] Secretary, R. M. B. A. (2011). Guide for conducting risk assessments.

[3] Nget, R., Yang, C., & Yoshikawa, M. (2017). How to balance privacy and money through pricing mechanism in personal data market.

[4] Tian Li, An Jing. (2015). Survey on Social Media Users' Privacy Concerns. *News and Writing* (1), 37-40. (in Chinese)

[5] Khokhar, R. H., Chen, R., Fung, B. C., & Lui, S. M. (2014). Quantifying the costs and benefits of privacy-preserving health data publishing. *Journal of Biomedical Informatics*, *50*(8), 107.

[6] Coello, C. A. C., Pulido, G. T., & Lechuga, M. S. (2004). Handling multiple objectives with particle swarm optimization. *IEEE Transactions on Evolutionary Computation*, *8*(3), 256-279.

[7] Dai, M. M., & Chu, Y. F. (2011). The model of spread path with strongest risk diffusion capability in industrial chain based on complex networks theory. *IEEE International Conference on Grey Systems and Intelligent Services* (Vol.488, pp.753-756). IEEE.

[8] Mclaughlin, S., Laurenson, D. I., & Tan, Y. Y. E. (2006). *Mobile ad-hoc network*. US, US 20060176829 A1.

[9] Newman, M. E. J. (2003). The structure and function of complex networks. *Siam Review*, 45(2), 167-256.

[10] Jardine, E. (2017). Privacy, censorship, data breaches and internet freedom: the drivers of support and opposition to dark web technologies. *New Media & Society*, 146144481773313.

[11] &Amp, M. . Onion browser | tor-entwickler bemängeln "tor browser"-a. | mac & i news-foren. *Heise Zeitschriften Verlag*.

Task 8: POLICY ADVISORY

Honored decision-maker:

Private information commercialization has become an irresistible trend. In the meta-data era, the use of private privacy is very wide. In order to make a reasonable system and establish stable market environment, we put up some recommendations to decision-maker from the following aspects.

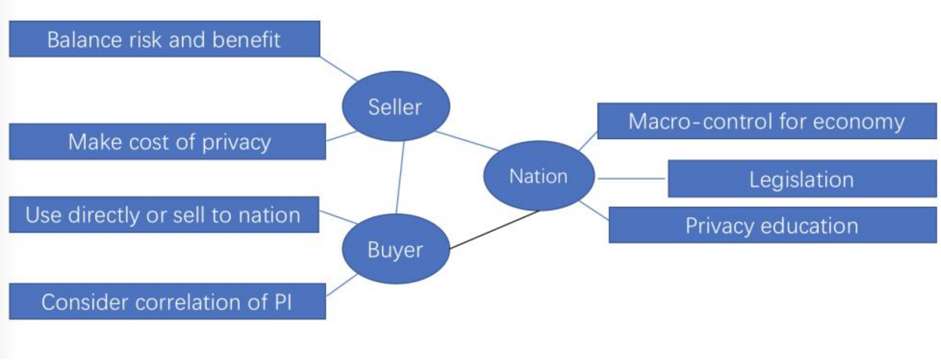
From the perspective of nation, managers should legislate on privacy protection and set up appropriate information security department. Then, the state should conduct macro-control on the market price of privacy and adjust the price within a reasonable range. Countries should make strict restrictions on the content of privacy which can be sold. For instance, private information from some government officers is not supposed to sell. As the value of information has been commercialized, information security should attract enough attention. In order to enhance people's awareness of privacy protection, nations need to publicize privacy security and information security especially to adolescents.

From the view of sellers, first of all, individuals should have the concept of privacy protection, to avoid the sale of their own privacy for the benefit with no limit. Sellers should balance risk and benefit according to parameters we considered. When sellers want to sell their private information, they need to identify their own risk level and evaluate the vulnerabilities and threats. Sellers should constantly adjust their price according to the price point and evaluate cost of privacy based on risk and tradeoff. For private information in specific domain such as social media, financial transaction and health records, sellers should carefully assess the value of different types of information. According to our model, we find selling information with high correlation can dramatically increase costs. So, sellers need to consider correlation of different types of private information in order to avoid suffering damage too much.

For purchasers, purchasers need to evaluate utility value before transaction. Different kinds of private information have distinguished utility values in specific domains. Purchasers need to identify cost-effective private information to get revenue as more as possible. According to our game model, purchasers could directly use PI or sell to others. Purchasers need to choose their characters reasonably on the basis of utility value and price. When buying information, purchasers should comply with laws and regulations.

In addition, in social networks, people are closely connected and people are divided into different communities. Due to the sharing of private information, more and more PI will be exposed and correlation of IP will increase. People should consciously protect their privacy and prevent others from revealing their privacy. According to our model, attitudes or beliefs will spread from one by one. With generation aging, more and more people are willing to sell PI. Therefore, people need to stick to their own concepts.

As for some special phenomenon such as monopolies, nations need to take some measures to control market monopoly. Monopolies will control most of the market information which leads to the change of price. For PI stolen and sold on dark web, nations should manage and restrict.



In short, in order to formulate reasonable price, all of works need to try their best.

Sincerely yours,
Team# 85722.