Summary

We believe that Private Information (PI) and privacy must be clearly distinguished. While PI is just a good that can be traded for benefits between any entities, privacy is the valuable and meaningful personal space that is violated when a third party gets hold of your data.

Our first objective is to find an ideal price of PI and privacy. Our main finding here is a generalised model that accounts for both PI and privacy - the PI Equation. It is supported by a Data Valuation Framework, which puts a value on the completeness of the data of an individual and translates qualitative data into a quantity. Together, they help an average person derive the value that he/she should receive in return for a certain amount of data. This enables regulators to set a fair price that data should transact at.

We hypothesize that the difference in prices of PI and privacy is due to market failure, and our model is built based on those assumptions. Eventually, we managed to quantify the factors of imperfect information by grounding our model in reality.

Our model gives valuable insight on the differences between PI and privacy and can flexibly account for different companies and individuals, although it is still the most accurate to model transactions between an average individual and an average company. Sensitivity analysis also revealed the effective bounds of the PI Equation.

The PI Equation supports our next objective, which is to come up with a set of policies. PI is a non-rivalrous and easily transferable good, which is why price regulation is difficult to enforce, even if the ideal price is known. The regulation of price is thus not by direct market intervention.

We suggest three different solutions: strengthening and enforcing existing legislation on data protection, education campaigns to increase the sense of awareness, as well as the promotion and protection of a distributed cryptographic data framework.

Stricter legislation on data protection and higher fines make companies feel the risk of any data breach. Also, the increase of awareness enlightens people on the true price of privacy. Moreover, a robust distributed cryptographic-data framework allows users to benefit from transacting their data without even disclosing it, greatly reducing the risk of selling data.

We then analysed these policies with reference to our model.

Pricing Privacy and Personal Information

ICM Contest Question F

Team # 87374

February 13, 2018

Contents

1 Introduction

1.1 Should There Even Be a Price on Private Information?

Before we start evaluating the cost of private information, let us first consider if there should even be a price on private information and what benefits such a price system could provide.

Current Situation Today, people readily give up their information in exchange for "free" goods and services, for example, by creating an account on a certain platform (and disclosing your private information to do so) to obtain benefits such as electronic books, games, calendar management services or to use a social network platform. Some people even generously fill in their private details on a form just for a chance to win a lucky draw.

The issue with the above situations is that people seem to believe that they have nothing to lose from the exchange as they get "free" goods and services. However, they fail to take into account that their private information has value too. In fact, the value of their private information may be worth more than the few dollars they save from getting the "free" goods and services.

1.2 Risks of Disclosing Data

Most individuals today fail to consider the extent at which their data is being used and the risks that they are taking in disclosing their data.

Unfairness Towards Individuals Once companies obtain private information from individuals, these individuals no longer have any control over how the companies handle their data and how securely their private information is stored. There is also a lack of regulations that prevent companies from trading the obtained data amongst themselves to obtain even more benefits. This is especially favourable for larger companies as they have a monopoly over individuals' data and can use it to rake up high profits at their expense.

Data Breach There are also unexpected risks that arise from sharing your data. Data breaches are increasingly common and have affected large corporations like Uber [9], Yahoo [8] and even companies that hold sensitive data like Anthem Inc. [14]. After a data breach, individuals' information is compromised and likely to be sold on the dark web, which is accessible to

people with malicious intents. Considering the frequency and scale of these recent data breaches and many more that were not mentioned, as well as the personal dangers posed to individuals whose private information has been compromised, it has become necessary for companies to remedy the situation by compensating users in the event of such breaches.

1.3 Why People Still Disclose Their Data

Despite the risks of sharing one's private information, individuals still continue to share their private information freely.

Ignorance The main reason for this is that most individuals are unaware of the real dangers of sharing their private information as well as the commercial value that their private information provides to companies. Ferenstein Wire [5] revealed that common daily scenarios, when phrased differently i.e. explicitly stating the private information extracted in these scenarios, actually resulted in people being less willing to trade in their private information, showing that people did not know the full extent of the private information they were sharing. As a result, individuals do not know how to price their data and often undervalue the actual worth of their data.

Convenience and Necessity Another reason why individuals readily share their data is due to convenience or necessity. There are times when individuals feel that a particular deal is good enough and do not bother to search for an alternative just to protect their private information, or when a service provided is unique and individuals that require the service have no choice but to comply with the terms and conditions issued. For example, social media sites like Whatsapp and Facebook have become popular sites for companies to communicate with their employees, forcing the employees to share their private information with such sites.

The social benefits of sharing data Individuals share data also because they believe in the welfare generated from learning from their data. Many organisations have built up good reputations and credibility through their corporate social responsibility initiatives. For example, technology companies have been experimenting on tools that benefit the society, such as the early detection of health issues [7] and the awareness of suicidal thoughts [2] and terrorist tendencies [10]. Such potential social benefits likely outweigh the occasional privacy concerns of most individuals.

1.4 Commodifying Private Information

There is no standard to base the price of something as intangible as private information against. It is impossible to come up with a general universal cost for an individual's private information, making it unmarketable. Moreover, PI is a non-rivalrous and easily transferable good. That is why price regulation is difficult to enforce, even if the ideal price is known.

Similarities and Differences Between PI, PP and IP Private Property (PP) is easily enforceable as it is easy to build physical barriers to prevent intrusion. An intrusion of PP is highly visible, and intruders are likely liable to condemnation and/or punishments.

Intellectual Property (IP) is less tangible than PP, but still largely enforceable. Despite the rampant pirated media on the Internet, production companies are still able to survive and earn profit through legal means, with the help of piracy controls such as copyright. Moreover, punishments for infringement of IP is stricter on companies to deter the making of monetary profit through illegal means.

However, there is currently no similar protection for Private Information (PI). We have to recognize that once data is disclosed, it is impossible to prevent companies from trading the data among themselves or in any other way beneficial to themselves, especially with the increasing capabilities of technology today.

1.5 Distinguishing between Private Information and Privacy

There is also a need to distinguish between **private information or data** and **privacy**. People are often willing to exchange their private information for benefits such as discounts and convenience but they are also willing to pay a price to protect their privacy. [11]

We believe that the difference lies in the method of valuation - privacy is the valuable and meaningful personal space that is violated when a third party gets hold of your data while private information is a good that can be traded for benefits between any entities, be it individuals or companies.

Generally, people tend to value their privacy, often more than they value their private information. This is evident in the numerous empirical studies conducted to examine the difference between the valuation of personal data and the valuation of privacy [11]. We will further emphasise and differentiate between the two in our models.

2 Assumptions

2.1 Scope of Analysis

The graphs represent an average individual in a single transaction with a single average company. Besides private information, the individual incurs no other costs in the transaction. For instance, the time spent is ignored, or is considered as being absorbed by the Benefit curve (described later).

2.2 Quality of Information

Data Valuation Framework Before attempting to arrive at a price for information, we first attempt to quantify the the quality of information transacted, because different amounts and kinds of information are valued extremely differently to different individuals. We make a huge assumption here by coming up with an idea of a totality of all information, where all the information of an individual can be represented by a number, say 1, and that the value of each piece of data is then a fraction of that totality.

Companies value the quality of data based on the amount of economic benefits they can derive from it, be it extra revenue, market share, or customer loyalty. In other words, they value the Private Information objectively as economic goods. We first identified several categories of data [3]. Then, we listed out some of the information that could benefit such companies (refer to Table 1 below).

Table 1: The Quality of Information of Personal Data

Constituent Factors	Personal Data	Quality of Information
Coordinates (0.0006) Places Visited (0.0011)	Current Location	0.0017
Sites Visited/Clicks (0.0024) Time Spent (0.0024)	Web Browsing History	0.0048
Full Name (0.0028) Company Domain (0.0536)	Email	0.0564
Preferences (0.0225) Spouse Info (0.0440)	Marital Status	0.0665
Medical History (0.0516) Family History (0.0658)	Health Condition	0.1174
Connections (0.4333) Background (0.1202) Interests (0.1998)	Full social network profile	0.7533

We are not considering the information sold by hackers, which is a breach of security and not a breach of privacy. We are also not considering the analytic work that data companies might have done to value-add their data product. Here, we are simply considering the value of such raw PI to trustworthy companies and entities.

An average person values the quality of data transacted based on how much meaning or how intrusive the data is to him, i.e. data is valued not based on how useful it is, but how much privacy it infringes upon. We listed out the different types of data (refer to Table 2 below). [16]

Table 2: The Quality of Information of the Privacy Aspects

Constituent Factors	Privacy Aspect	Quality of Information
Gender (0.0021) Name (0.0060)	Gender and Name	0.0081
Spouse Info (0.0172) Current Availability / Type (0.0134)	Marital Status	0.0306
People Captured (0.0226) Places Captured (0.0226)	Photos and Videos	0.0452
Household Income Range (0.0504) Household Members (0.0122)	Home Address	0.0626
Items Purchased (0.0442) Shops/Sites Visited (0.0339)	Purchase History	0.0781
Current Coordinates (0.0926) Places Visited (0.0244)	Location	0.1170
Credit Card Number (0.0305) Expiry Date and CVN Code (0.1004)	Credit Card	0.1309
Date of Birth/Nationality (0.0812) Account User Name (0.0696)	ID Number	0.1508
Account Access (0.0956) Related Passwords (0.0876)	Passwords	0.1832
Medical History (0.1934)	Medical	0.1934

This is how an average composition of information quality might look like. As we have no way of finding a universal average of the price of different data types, these values have been obtained by reverse-engineering our model. These values will also be used as examples throughout the report.

3 Base Model

The central question we are tackling in this model is: **What should the price of private information and privacy be?** There are three ambiguous terms here - ”cost of private information”, ”cost of privacy” and ”should”, and all of which must be clearly defined in our base model.

What is meant by *”cost of private information”*? Since we take ”private information” to be the objective economic value of the information, it is simply the price that personal data is traded at between companies. Alternatively, if individuals wish to sell their data voluntarily, it is also considered as the sale of PI and should be judged by this cost as well.

What is meant by *”cost of privacy”*? Since we take ”privacy” to be the protection of the individual's personal and meaningful information against unknown entities, the ”cost of privacy” is the cost that individuals feel that third parties should pay to access certain information.

What is meant by *”should”*? It means that the model sets an ideal standard that we should follow, which is a price that is fair to both the data-provider and the data-purchaser. This also begs the question - why are we not already at the ideal?

3.1 The Initial Idea: the Naive Provider of PI

The graph here (refer to Figure 1) plots the values of different types of PI per transaction for an average and naive individual.

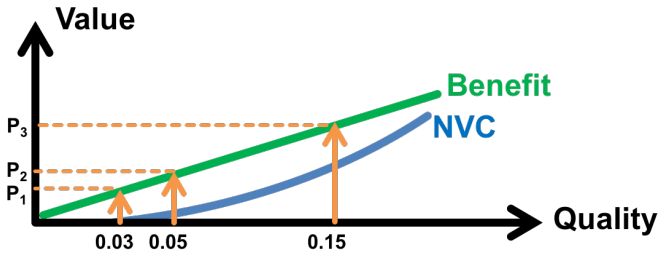

Figure 1: Transaction for an average naive individual

Quality of Information The horizontal Quality-axis refers to the completeness of data. The domain of this axis is 0-1, corresponding to the percentage completeness computed using the Data Valuation Framework.

Value The vertical Value-axis refers to the value obtained by the data-provider or data-purchaser from a certain transaction. The Value-axis is a generalization of Price to include intangible and non-measurable benefits that applications and services provide.

Naive Valuation Curve The curve in orange is the NVC. Every point on the curve refers to the amount of value naive individuals currently expect when they disclose their information to a trusted company. The naive individuals are unaware of how much their data is really worth, and the true risks involved in providing others with their data.

Delta Offers The vertical lines showing the deltas indicate an offer. An example of such an offer could be a $5 discount voucher if you disclose your payment information. If the payment information is worth 0.2 on the quality scale, this corresponds to a delta offer with a height (value) of $5 at 0.2 on the quality scale. The delta offers intersect with the NVC at the minimum value at which the naive individual will exchange his or her information.

Benefit Curve Next, we depict a graduated offer system, referring to services which offer increasingly better services with the more information you provide. This is similar to aggregating many separate offers (delta peaks) over the 0-1 quality scale. This results in a Benefit curve.

Let us take the use of Facebook as an example. Providing your basic credentials (0.02 quality) gives you access to post your information, shown as benefit P_1. Then, you have the option to provide data of your workplace and past schooling experience (extra 0.03 quality) in exchange for the convenience of adding your past colleagues and classmates as friends, shown as benefit P_2. Additionally, you can provide your credit card information (extra 0.10 quality) in exchange for gaming services, giving you a total benefit of P_3.

3.2 The Enlightened Provider of PI

So how do we find an ideal price for **privacy** in our model? We can borrow the Economics concept of Market failure. We understand that if the individual becomes more informed about 1) the true value of his/her data, and 2) the true risks of sharing data, he/she will request higher prices and value for the PI that he/she provides.

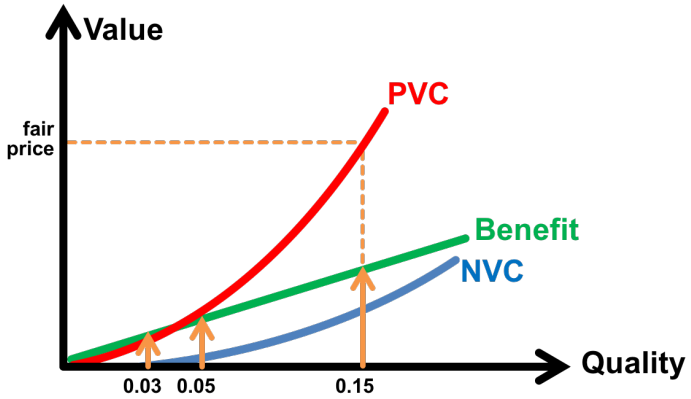

Figure 2: Comparing the Naive and Enlightened Provider of PI

Privacy Valuation Curve Referring to Figure 2, the PVC represents the true valuations of revealing sensitive personal information by an enlightened individual. We can compare this to the NVC, in which a naive provider is willing to provide his meaningful data for a fraction of the true value.

Ideal Price The ideal price would then be the intersection between the PVC and the Benefit Curve. This is the fairest price for both the individual who provides the data and the company that buys the data.

3.3 The Risk-Ignoring Companies

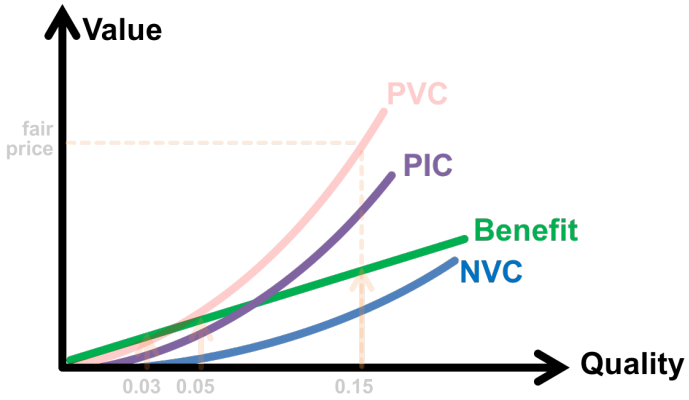

Figure 3: Graph of the Market Valuation of PI

How do we then find an ideal price for **Private Information**? Again, we borrow the concept of Market Failure and observe that while companies have close to perfect information of the value of PI, they ignore the risks it brings to the individuals when they sell the information. Thus, they demand a price between the NVC and PVC i.e. the PIC (refer to Figure 3) when they sell the PI to other companies.

This curve has a grounding in reality and we can find it by searching up the price of transactions of the different kinds of data.

3.4 The PI Equation

$$V = AB(e^{ln(M)\cdot x^{\frac{1}{b+c}}} - 1)$$

This is the equation that models the cost of privacy and PI of individuals, which is a generalization of the PVC, the PIC and the NVC.

The Parameters where:
V = value (in dollars)
x = data completeness (from 0 - 1)
A = risk factor (interpreted as net worth multiplier, which are the losses should the data lead to a security breach)
B = discount factor (a % discount which models the reduction in stinginess of data when it is used for a good cause)
M = maximum price (at complete info of 1) which is currently set to 3500 for privacy and 100 for personal data
b = level of knowledge of value
c = amount of risk considered

3.4.1 The Exponential Behaviour of the PI Equation The equation should be a general model that can compute the value of both PI and privacy. We adopted an exponential form for the equation as we felt that, the more information that people know about us, the more connections and inferences they can make, and the more the sense of violation that we feel. For example, knowing that someone is part of a Facebook group with extreme political viewpoints and that another person is a teacher, is much less intrusive than knowing that the same person is both a teacher and has extreme political viewpoints.

Another way of seeing it is that when people know more pieces of information about us, the amount of information they can infer is not limited to a connection between only those two pieces, but also the connection between

all of the pieces of information that they have seen so far. Thus, mathematically, this is not merely a quadratic (finding connection between 2 pieces of information) or cubic relationship (finding connection between 3 pieces of information), but rather an exponential relationship which considers all the subsets of all the information provided.

3.4.2 The Knowledge Factors Captured in the PI Equation The first knowledge factor that we had identified previously was the knowledge of the value of the PI which we denote by a factor b. This was the difference between the value demanded by naive data providers (NVC) and the price that companies demanded (PIC) in exchange for data. In other words, $b = 0$ or close to 0 for naive data providers. Our subsequent analysis showed the expected average value for b to be around 0.6811 for a company.

The other knowledge factor that we identified was the knowledge of the risk involved in selling your data, which we denote by a factor c. The key difference that we identified between privacy and PI is the awareness of the risks involved. For companies, $c = 0$.

The total knowledge factor is $b + c$. The higher the knowledge factor, the steeper the gradient of the curve.

3.4.3 The Boundary Conditions of the PI equation The PI equation also has 2 important boundaries. Firstly, the value demanded at 0 data should be 0. Furthermore, the value demanded at quality of data = 1 should be the complete profile of a person. Thus, we include a parameter M which refers to the maximum price demanded for privacy and PI. Mathematically, it is the limit of V as x approaches 1. We found that numbers online suggest the price of complete privacy of a person is worth about $3500. Since companies pay up to $57 per user profile when they buy a company, our team estimated the cost of a full profile to be objectively worth about $100. [1]

3.4.4 Consideration of non-average people and non-profit organizations After computing the value of the data for an average individual, we added two parameters, A (risk factor) and B (discount factor), to consider for a non-typical transaction. For an average individual in an average transaction, $A = 1$ and $B = 1$.

The risk factor (A) captures the idea that the richer or higher in status one is, the more one stands to lose from disclosing personal information. A can range from 0 (an extremely poor person) to any large number. It can

be interpreted as the ratio of the net worth of a person to the net worth of an average individual. This value can also be altered to factor in a person's level of social standing and influence (for example, a country's president might not be very rich but he is very influential).

The discount factor (B) considers data transactions that might be used for the common good, such as for security purposes (anti-terrorism). The discount factor can range from 0-1, where 0 means the purpose of the data is so good that it should not be charged, and 1 means that it is for a commercial purpose.

3.5 Grounding the Values

3.5.1 Obtaining the Data The objective value of PI can be represented by the amount of money that companies are willing to pay for each transaction. Here, we mainly used these three sites to obtain data points. [15] [16] [1] These sites somewhat corroborate in terms of the value of information and are recent sources of information obtained from companies themselves. Two of them are obtained from the transaction prices of data companies while one is obtained from the selling prices of data-based companies.

While finding out the value of privacy is much harder, our team based these values on a survey which asked individuals how much they thought their data was worth. This was because individuals' subjective opinions of the cost of their own PI is often more representative of the price of privacy. Furthermore, the way the company asked the question in this particular survey, "put a cost on a chunk of information that a trusted third party would have to pay to access", suggested the high commercial nature of the transaction, which we think is a good reminder of the value of the PI to companies. [17]

3.5.2 Comments on Data Collection Methods In general, it was hard to find reliable data on how much privacy and PI are worth. Not only were there few studies that provided insights into this issue, the data varied wildly from site to site. Also, the surveys were done on different continents in different years. To standardize, our team always used the most recent data available for a statistic (mostly 2017) and also used data from the US as much as possible. We did not use any data obtained from the prices in the dark web as we feel that those are highly biased values for data that serve different purposes than commercial ones.

Here, we discuss some assumptions that we made, some advantages of our data methods, and some disadvantages.

Using the data from individual surveys might be biased as studies have shown that the values assigned to personal data are highly context-dependent. This means that the results of the survey are extremely sensitive to the phrasing of the questions. There is also a lack of market verification as we cannot check for the validity of these hypothetical prices. [3]

While using the data from companies is good because they are market-verified, they might also not be too accurate because they include the benefits from network effects. They also include the prices of data collection and refinement. [11]

3.5.3 Fitting the numbers Using the values obtained online (which are organized into a table in the Appendix), we fitted the numbers to the PI equation to obtain the general PIC and PVC (refer to Figure 4).

By rearranging the weights of the Data Valuation Framework appropriately and reasonably, we came up with two equations for an average PVC and PIC, which are:

$$\text{PVC} : V = (e^{ln(3500)\cdot x^{\frac{1}{2.6886}}} - 1)$$

$$\text{PIC} : V = (e^{ln(100)\cdot x^{\frac{1}{0.6811}}} - 1)$$

Since $b = 0.6811$, this suggests that an average companies' level of knowledge of the price of information is at about 0.6811. For an average consumer conscious of his privacy and of the value of his information, $b + c = 2.6886$.

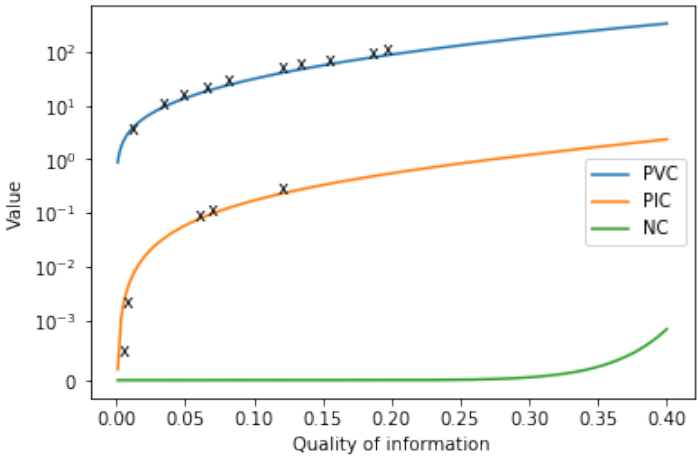

Figure 4: Our model is based on real price points

4 Analysis

4.1 Strengths

Our main strength of the model is its ability to differentiate between the value of privacy and PI. The current market prices of PI are simply the companies' transacted values. Thus, we can use the model to determine not only the ideal price for privacy, and also how far away the current prices are from the ideal price.

Another strength of the model is its illustration of the relationship between the knowledge of the value of PI (b) and perceived risk (c) on the price of PI. This can give us insight on the effectiveness of education and awareness policies.

Our model also finds a way (the Data Valuation Framework) to account for the differing values of different types of information for different individuals. Using the framework, we can also use the perceived values of privacy to find the value of a certain piece of information from the perspective of an individual.

4.2 Weaknesses

One weakness of the model is that the model only works best for an average individual and company. This is because many of our approximations were based on the average person/company, which is inevitable as we do not have values or data for special cases, and we cannot really account for it.

The Data Valuation Framework is highly subjective. In fact, even individuals trying to use the framework for themselves might have a problem quantifying the value of their information. As such, the use of this framework introduces much bias in the model.

The accuracy of our model is also highly dependent on the accuracy of survey results and the veracity of the data we collected. They are also subject to the assumptions we have discussed above.

Our model also failed to take into account network effects. Data obtained by companies could be worth a lot more just by virtue of being collated in a large number.

4.3 Sensitivity Analysis

Sensitivity to the value of an average person's full profile M represents the maximum value. The graph plotted (refer to Figure 5) shows us a family

of 3 values: 1000, 3500, 5000. As the values of M increase, the slopes get steeper and steeper similar to a normal exponential family.

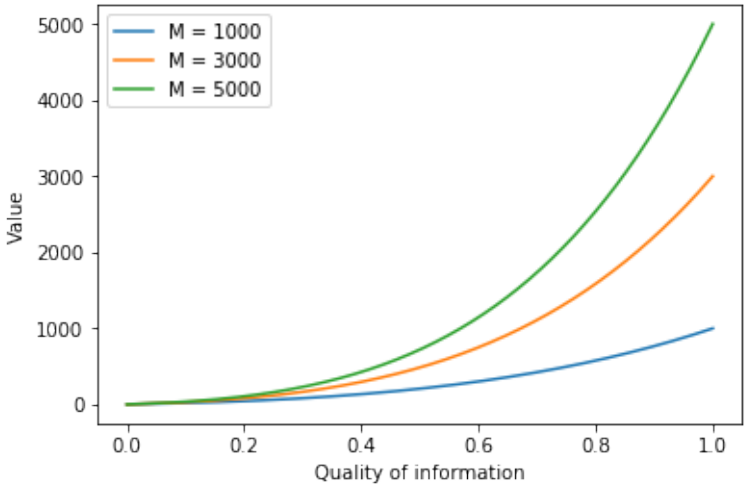

Figure 5: Sensitivity Analysis Graph w.r.t M

From the very naive to the very paranoid In general, $b + c$ (information and knowledge of risk) represents the slopes of the curve. The graph plotted (refer to Figure 6) shows us a family of 5 values: 0.001, 2.689, 5.0, 8.16, 14.0.

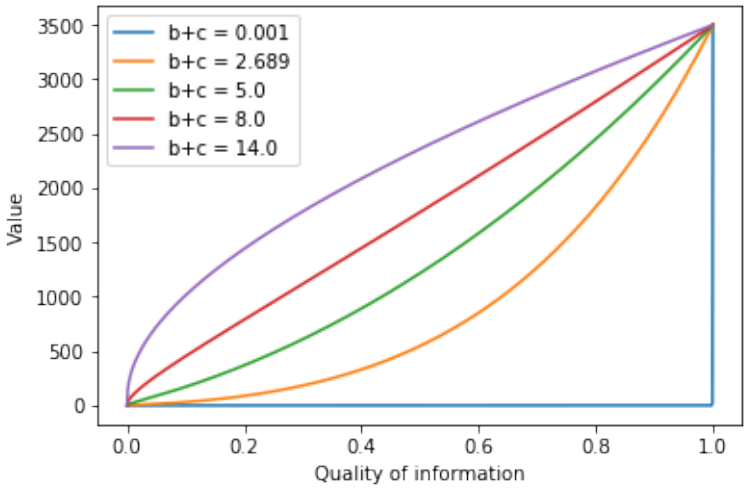

Figure 6: Sensitivity Analysis Graph w.r.t. $b + c$

For a very naive person, whose everything else is average but $(b + c)$ is

close to 0, we observe that the value of privacy is extremely low all the way, until quality of info = 1, where it shoots up to the value of M. The graph looks like an L, which is not realistic as it will probably just be a low value throughout in real life. As the values increase, the slopes generally get more and more concave. At the value of around 8.0, the graph becomes approximately linear. At values of higher than 8.0, the graph becomes logarithmic, which means the model is no longer applicable.

Hence, our model is only really realistic when b+c is below 8.16, and also shows unintended behaviour at quality of data =1 when $b + c$ is too low.

We also found that this value of 8.16 that we found is dependent on the value of $ln(M)$. We found that the value of a "linear"-looking line, or the breaking point of exponential behaviour, happened when $b + c = M$ approximately.

4.4 Network Effects of Data

Other factors external of the individual can also affect the value of the individuals data. Individuals are part of a network, and when more and more individuals join the same network, network effects come into play, increasing the value and the desirability of each individuals private information.

While our model currently cannot quantify and take into account network effects, we think that our model has the potential to do so with more data. By observing the changes in b and M that occurs over different network sizes, we can create a separate model that finds the parameters b and M as a function of network size, hence quantifying the network effect.

4.4.1 Impact of Network Effects on Our Model and Policy Network effects have a significant impact on our model and policy as they heavily influence the value of the individuals private information. By properly utilizing network effects, the value of an individuals private information will rise significantly.

4.5 Applicability In The Real World

4.5.1 Comcast case study On September 18, 2015, Comcast paid $100 compensation to each of their 75,000 customers. They were also fined $33 million. Comcast had published their personal information even though those customers had specifically paid a fee for their information to be kept private - a privacy fee. If we calculate the information leaked using our

Data Valuation Framework for privacy, including Gender and Name, ID Number and Home address, we get the value of data for an average person to be 0.2215. If we use our PVC to find the quality of data that Comcast perceived each customer had lost such that they deserve $100 compensation, we get the number 0.216. The closeness of these values show the real-life applicability of this model. [6]

We feel that this model might also be useful in calculating the compensation for the other massive data breaches, such as Yahoo!'s 3 billion account breach.

4.6 Future Analysis

As the saying goes, "data is the new oil." As more and more insights are being developed with data, data is also getting more and more precious and expensive. This effect will be even more obvious, as more and more people are aware of the dangers of sharing their information, as well as the value that companies get out of it.

On the flip side, as more and more data become available, it could also be that the same set of data depreciates over time.

5 Policies

As PI is non-rivalrous and an easily transferable good, price regulation is difficult to enforce, even when the ideal price is known. With this in mind, we propose three distinct solutions.

5.1 Legislation

In the short term, we should pass pro-privacy laws, and enforce them. We should strengthen and execute current privacy laws, like increasing fines, or requiring companies to have robust measures in place to protect individuals' data to minimise chances of data breaches. This legislation will provide the necessary legal foundation to protect individual privacy.

We should also regulate the trading of private information. The intention to sell data should be made explicit to individuals when they contribute their data. We should also enforce that companies allow individuals to download their data any time they want, to allow them to find out exactly what data has been collected from them.

5.1.1 Illustration On Our Model With these laws in place, we believe that companies will at least be cognisant of user privacy in their decision making process. By increasing the fines for a data breach, companies are forced to consider more risks before collecting data, which increases the c for them. The amount of fines for a data breach can be calculated from the model as a function of the c that the regulator hopes to achieve in companies.

We hope that extra risks borne by companies can lead to more stringent measures in processing user data, and avoid the trading of non-explicitly consented personal information to protect their reputation. Moreover, if companies are found to be violating the laws, a punishment can also be extrapolated from the model, which is the total price of all the privacy breached.

By letting individuals know that their data has been commercialised, and exactly what data has been taken from them, individuals are more informed about the value of their information and the risks that a particular service is putting them through. This increases $b + c$, the awareness of value and risk in individuals.

5.1.2 Ineffectiveness of Current Laws Despite current laws punishing a mishandling of individuals' data, we still hear data breaches happening from time to time, as mentioned in the introduction. The root cause is that these companies store data on their servers, which are vulnerable to intrusion. While companies that are more focused on security and privacy like banks would invest more money, firms less focused on security (like Uber) do not have the incentive to make the their data framework fully secure.

We still cannot totally prevent companies from trading the private information without the knowledge of users. However, legislation, while limited in effect, acts as the legal foundation for the individual's right to privacy.

5.2 Awareness Campaign

We recommend the decision maker to start information campaigns to publicise the importance of privacy. Though press releases and published information, the decision maker can inform the dangers of disclosing your information. We should also encourage publicity of data breaches to illustrate how real the problem is.

5.2.1 Illustration On Our Model We assume that everyone will have a clearer understanding of the importance of privacy, leading to an increase in $b + c$. We illustrate this by a movement of the NVC towards the PVC (refer to Figure 7). In the example of Facebook, people may no longer want to use Facebooks messaging function anymore, because those messages might be leaked. If the new PVC no longer intersects the delta function, the user stops using the service.

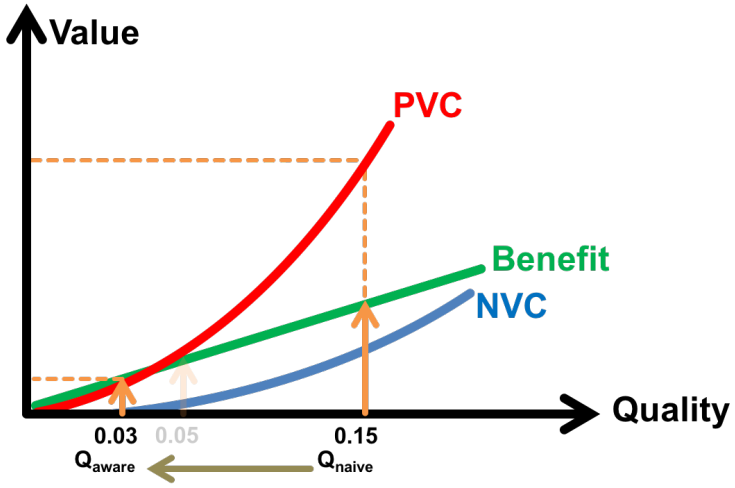

Figure 7: Effect of increasing awareness on privacy

However, realistically, we will not expect the people to be enlightened immediately. Despite Snowden's highly publicised revelations of NSA's privacy overreach, most of the Americans are still apathetic to the affair. People with high privacy awareness have largely not changed many of their other habits either. One reason could be that alternative technologies such as the Tor web browser are inconsistent and slow. Therefore we need to nurture alternative yet convenient technologies that fully respect the privacy of the users.

5.3 Restructuring the Internet

We propose supporting and protecting the development of a decentralised and cryptographically-secure Internet that allows users to participate equally.

5.3.1 Decentralised Cryptographic Data Framework A decentralised cryptographic data framework allows users to sell the benefits from the usage

of their data without even disclosing their data. It is possible with a few recently developed methods. There is a group of volunteers currently developing OpenMined [13], a distributed data framework with deep learning, federated learning, homomorphic encryption as well as blockchain smart contracts. It assumes a fully rigorous framework and a trusted and verified third party known as the Oracle. A more in-depth description of this framework can be found in the Appendix.

5.3.2 Capabilities of a Crypto-Data Framework Users can now allow their private data, such as their chat or search history, to be used for training a model without divulging the data at all. The user can also be compensated fairly every time the data is being used. Similarly, companies can utilise peoples data without ever knowing or accessing the data.

5.3.3 Effect of the Crypto-Data Framework on our Model As people are guaranteed that their data will not be shared, more people are willing to share the benefits of their data at a low price. In other words, the value of c decreases sharply, even close to 0. This will decrease the PVC curve to low levels close to the PIC (refer to Figure 8) and allow companies to utilise data for cheaper prices. We think that this is the win-win situation that allows users to benefit from their data without risk, and for companies to use data without spending a fortune.

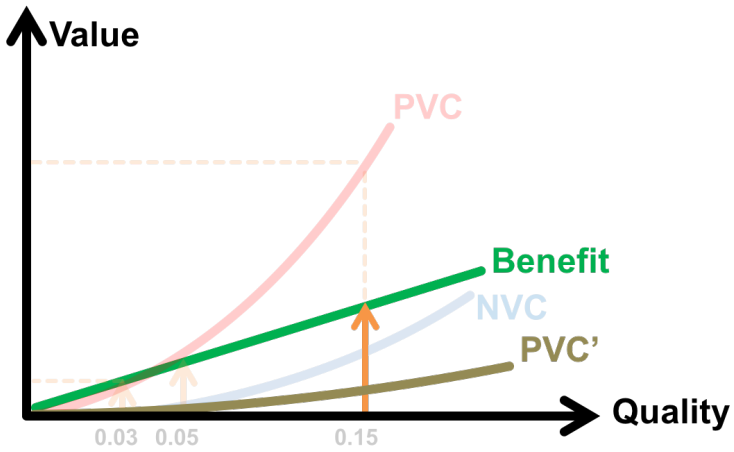

Figure 8: Effect of the Crypto-Data Framework

We foresee that people "paranoid" about their own data will be the first

adopters of such decentralised services. As more people realise the benefit of not having their data stolen, they will also move on the decentralised data framework.

5.3.4 Limitations of developing the Crypto-Data Framework While the decentralised artificial intelligence framework is in development, the state-of-the-art cryptographic algorithms were still the same ones we used since the start of the millennium. [12] The important question to consider: why there is such a lack of privacy-focused services in the first place? This could be mainly attributed to the lack of awareness of data privacy from individuals, which we hope will improve over time through our education campaigns.

Moreover, it is difficult to start a decentralised application given the pervasiveness of the major players in the tech market. Diaspora [4] is proposed to be a decentralised social media framework, but its size pales in comparison to Facebook, and the benefits of privacy are insufficient to convince the users to move away from the comfort and wide network on major social networks. There is no competing alternative version of the platform that is decentralised and focuses on privacy.

5.3.5 Support Required Therefore we call for the support the growth of decentralised frameworks so that the user has full control over their data and can benefit from any use of their data. It is best that we start supporting and protecting the decentralisation of the Internet for the good of the people and companies.

6 Conclusion

In summary, we have quantified personal information in terms of normalised quality of information using the Data Valuation Framework. We then used the quantities to provide a pricing structure for PI and privacy. Using the pricing model, we then recommended the policies of stricter legislation of data, greater awareness, and the wider adoption of a decentralised cryptographic data framework.

7 Policy Memo

Dear Decision Maker

In the age of big data, where "data is the new oil", we feel the need to re-emphasise the status of privacy as an inherent human right. Private information (PI) is being sold by the millions at data companies while the individuals themselves, whose data is being sold, are given barely any benefits. Yet individuals are the ones at risk when data breaches occur. PI such as addresses, phone numbers, medical information, marital status and so on can seem trivial to the some, but it can mean a lot to some others.

We hypothesise a few reasons for this problem: people do not know the real value and risk of sharing their PI, and companies also do not care about the risks they bring to individuals when their data gets sold. In other words, market failure due to imperfect information and inequitable risk.

It is also not simple to develop a policy for re-pricing of PI and privacy for two reasons. Firstly, PI and privacy are clearly different things - while PI is just a good with an objective value that can be traded for benefits, privacy is the valuable and meaningful personal space that is violated when a third party gets hold of your data. Differentiating between the prices of the two is already a difficult task. Secondly, PI is also a non-rivalrous and easily transferable good, which is why price regulation is difficult to enforce even if the ideal price is known. The usual methods of direct market intervention and taxes are thus unfeasible methods of price regulation.

In order to support strategies, we first present our method to find a fair price for PI and privacy.

Our method is centred around two main models: the Data Valuation Framework, and the PI Equation. The Data Valuation Framework is a method to transform some types of data into a value between 0-1, with 1 being the complete profile and information of a person. Thus, the Data Valuation Framework provides a numerical value on the quality of certain types of information. After calculating a quality-value, it is then utilised in the PI Equation.

The PI Equation is specifically built contextually on the hypotheses of our problem - that market failure is caused by imperfect information and inequitable risk. By assuming the exponential nature of the information curve and setting certain boundary conditions, we managed to ground our model in reality using data obtained from a few recent studies. We realised that our model gives valuable insight on the differences between PI and privacy. As our model focuses on a transaction between an average indi-

vidual and average company, it is also suitable for use at any scale - be it groups, companies and entire nations - as long as the knowledge and risk parameters of the entity have been empirically determined.

Comparing the data from our models to the Comcast data breach in 2015, we managed to obtain values of privacy very similar to the valuation by the authorities. This shows that our model has true practical value.

Now that we can have a quantity for the ideal price of PI and privacy, we offer our solutions to this problem.

1) Legislation. In the short term, we propose passing pro-privacy laws such as increasing the fines for a data breach, and enforcing regulations to be more transparent about commercialisation of PI.

We propose that fines can be calculated using the cost of privacy calculated in our model. The effectiveness of enforcing transparent commercial interests using the data can also be quantified and measured in our model.

2) Awareness Campaign. We propose starting an information campaigns to publicise the importance of privacy. Again, the effectiveness of the campaign can be quantified and measured in our model, which can help by being an indicator of progress.

3) Restructuring the Internet. We propose supporting and protecting the development of a decentralised and cryptographically-secure Internet that allows companies to utilize users' data in their analysis without accessing or viewing the users' PI.

We propose that the use of a decentralised cryptographic data framework can help both the individuals and the companies by reducing the risks of sharing data to nearly zero. This is phenomenon is also captured in our model. In such a way, individuals can trade their own PI without the risks, and can thus objectively treat it as a good.

We are in the midst of what is termed as "the fourth revolution" - Artificial Intelligence - that development is fuelled heavily by data. As time passes, data will only become more and more sought after and valuable. We beseech you to consider our proposal to solve the problem and implement them as soon as possible.

Regards
Team 87374

8 Appendix

8.1 PVC and PIC data points

For the graph in section 3.5.3

Privacy Aspect	Quality of Information	Value
Gender and name	0.008102356	2.9
Marital status	0.030562818	8.3
photos and videos	0.045238707	12.2
Home address	0.062646252	17.4
Purchase history	0.078099395	22.6
location	0.116972815	38.4
credit card	0.130906643	45.1
SSN	0.150798963	55.7
passwords	0.183232622	75.8
medical	0.193439847	82.9

Personal Information	Quality of Information	Value
Full web profile	0.753268558	20.0
Current location	0.001710917	0.0004
Marital Status	0.066454192	0.09
Web Browsing history	0.004763487	0.0018
Email	0.056358333	0.07
Health condition	0.11744972	0.22

8.2 How does a company train a model on OpenMined

(Elaboration for section 5.3.1) The data scientist provides the initialised model to the Oracle. The Oracle will negotiate with nodes that are fully controlled by individuals to request to train on their data. The Oracle passes a model to the individuals node for training on the data. The model passed is homomorphically encrypted to protect the learnt information. Then the model is then returned to the oracle, and the user is compensated through the contract if the information they provided is evaluated as correct and useful. Then the Oracle seeks other users to train the model on, until the training is complete.

References

[1] Ciro Borriello. How much is your online data worth, 2017. URL https://thegreatdissonance.wordpress.com/2017/06/03/how-much-is-data-worth/. [Online; accessed 12-February-2018].

[2] Josh Constine. Facebook rolls out AI to detect suicidal posts before theyre reported, 2017. URL https://techcrunch.com/2017/11/27/facebook-ai-suicide-prevention/. [Online; accessed 12-February-2018].

[3] Jeff Desjardins. How much is your personal data worth?, 2016. URL http://www.visualcapitalist.com/much-personal-data-worth/. [Online; accessed 12-February-2018].

[4] Diaspora. About, 2017. URL https://diasporafoundation.org/. [Online; accessed 12-February-2018].

[5] Gregory Ferenstein. Consumers are willing to give up some privacy for the right products, 2016. URL https://readwrite.com/2016/01/18/pew-privacy-study/. [Online; accessed 12-February-2018].

[6] Pauline Glikman, Nicolas Glady. What's the value of your data?, 2015. URL https://techcrunch.com/2015/10/13/whats-the-value-of-your-data/. [Online; accessed 12-February-2018].

[7] Eric Horvitz. Are there benefits to giving up our privacy?, 2014. URL http://www.bbc.com/future/story/20140220-can-giving-up-privacy-help-us. [Online; accessed 12-February-2018].

[8] The National. Yahoo suffers worlds biggest hack affecting 1 billion users, 1900. URL https://www.thenational.ae/business/yahoo-suffers-world-s-biggest-hack-affecting-1-billion-users-1.185328. [Online; accessed 12-February-2018].

[9] Eric Newcomer. Uber paid hackers to delete stolen data on 57 million people, 2017. URL https://www.bloomberg.com/news/articles/2017-11-21/uber-concealed-cyberattack-that-exposed-57-million-people-s-data. [Online; accessed 12-February-2018].

[10] BBC News. Facebook's AI wipes terrorism-related posts, 2017. URL http://www.bbc.com/news/technology-42158045. [Online; accessed 12-February-2018].

[11] OCED. Exploring the economics of personal data a survey of methodologies for measuring monetary value. 2013. URL http://www.oecd-ilibrary.org/science-and-technology/exploring-the-economics-of-personal-data_5k486qtxldmq-en. [Online; accessed 12-February-2018].

[12] National Institute of Standards and Technology. Guideline for using cryptographic standards in the federal government. 2013. URL http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-175B.pdf. [Online; accessed 12-February-2018].

[13] OpenMined. About, 2017. URL https://openmined.org. [Online; accessed 12-February-2018].

[14] Reuters. Anthem to pay record $115M to settle lawsuits over data breach, 2017. URL https://www.nbcnews.com/news/us-news/anthem-pay-record-115m-settle-lawsuits-over-data-breach-n776246. [Online; accessed 12-February-2018].

[15] Financial Times. How much is your personal data worth?, 2013. URL https://ig.ft.com/how-much-is-your-personal-data-worth/. [Online; accessed 12-February-2018].

[16] TotallyMoney. What is your personal data worth. URL http://www.totallymoney.com/personal-data/infographic/. [Online; accessed 12-February-2018].

[17] TrendMicro. How much is your personal data worth? survey says, 2015. URL https://www.trendmicro.com/vinfo/us/security/news/internet-of-things/how-much-is-your-personal-data-worth-survey-says. [Online; accessed 12-February-2018].