E-Commerce Consumer Privacy Protection Based on Differential Privacy

Guorong Zhong

1International School, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract: With the advent of the era of big data, human society is facing unprecedented opportunities and challenges. In many fields where big data is widely used, e-commerce is undoubtedly a crucial one. While the advanced data analysis technology penetrates into every link of e-commerce transaction, it not only brings about convenience, but also causes many adverse effects. On the one hand, this paper analyzed both the current situation of consumer privacy security on e-commerce platform and the increasingly mature application of differential privacy in various IT fields. It is found to be promising that with its good feature of privacy ambiguity, consumer privacy protection in differential privacy mechanism can meet the requirements of a more robust development of e-commerce in big data era. On the other hand, this paper briefly discussed about consumer protection in legal level, inspired by relevant researches done on GDPR.

1. Introduction

With profound development of network communication technology, mobile devices and intelligent terminals, the type and scale of data owned by society is experiencing explosive growth. According to the 2015 IDC monitoring report, the world has officially entered the ZB era in 2010, and the global data volume will roughly be doubled every two years, which means that the amount of data generated by human in the last two years is equal to the total amount of data generated before.

The broad development prospect of big data has attracted the attention of all sectors of society. Researchers in various fields are trying to understand people's needs from the perspective of data, develop new platforms and explore new opportunities, so as to improve social and economic benefits. Big data has penetrated into every corner of people's life, and optimized people's lifestyle in areas such as medical treatment and transportation. Big data plays an important role as a core driving force, especially in e-commerce. Nevertheless, e-commerce platform has an unshirkable responsibility for the leakage of personal privacy in the process of online shopping. In an integrated online shopping process, traces of consumers will be left on the network in various ways. Online merchants collect these great amount of user information, analyze user characteristics, adjust product release patterns, and eventually enhance the individuality of retrieval results imperceptibly, all in the purpose of obtaining greater economic benefits. According to CNNIC's 2013 research report on the information security situation of Chinese netizens, the proportion of personal information leakage in network shopping has reached 42.9%, which was second only to cases of fraudulent information and fraudulent websites.

This paper attempts to discuss the privacy concerns of online consumers' personal information under the e-commerce platform, and discuss how to strengthen the privacy protection of online
consumers based on differential privacy technology. In addition, by analyzing the situation, causes and effects of the problem, and combining with the legal level, this paper will propose feasible suggestions and assessment countermeasures for personal privacy protection.

2. Background

2.1. The trend of big data in e-commerce
The advent of the era of big data poses the human society unprecedented opportunities as well as challenges. Among the multiple fields that big data has already taken extensive applications, e-commerce would with no doubt be a vital one. On the one hand, penetration of data analyzing technology into every single link of e-commerce transactions brings us about much convenience. On the other hand, however, it triggers a few adverse effects.

2.1.1. Benefits
In 2016, Ning Jiajun pointed out that with nowadays rapid development of e-commerce, data has unconsciously become a flood in the tide of world economy and an important part of the field’s composition. Compared with traditional offline sales, e-commerce possesses a greater capability in producing large amount of data. Mountains of brand, commodity and customer information can form a highly connected aggregation through the internet. Furthermore, the collection, integration and analysis of these information enable online merchants, more quickly and accurately, to locate users’ requirements, figure out new business opportunities, and thereby affect the process of decision-making and improve the efficiency of operations. In 2006, the MIT Technology Review stated that Alibaba Group has completed a transition from the early electric business enterprise to a world technology pioneer, when the double eleven global shopping carnival of 2016 had just ended and created a series of new miracles in the world history of retails. On that single day, turnover reached 120.7 billion yuan, and wireless accounted for 82%. In addition, the carnival peaked 175000 transactions and 120000 payment per second, covering 235 countries and regions[1]. From the Alibaba’s advanced technology innovation, it can be seen that big data’s advantages on high dimensional data quantity and strong data processing ability are perfectly displayed on the e-commerce platform.

2.1.2. Drawbacks
As is mentioned before, big data has provided huge amount of users’ behavior data on the one hand, making it an indispensable impetus for the development of e-commerce. On the other hand, its limitations and drawbacks are hindering the steady progress of e-commerce as well. It includes difficulties in information filtering process due to low data density, and privacy issues in which information is abused or stolen. Among them, the information privacy has no doubt to be the biggest barrier for e-commerce platform to gain further development in the era of big data. When the profound development of internet technology is taking place, the openness of data on e-commerce platform and the frequency of data interaction, together with the risk in privacy leakage are on the increase. Nowadays threats to users’ private information on e-commerce platform come from mainly two aspects. One is private browsing habits exposed through traces on web page, the other is key information such as online payment password being stolen. First, as we can see from a complete process of online shopping, a series of steps from user registration, web browsing, order formation, online payment to the final logistic transportation will be taken. What canny e-merchants can do is far more than accessing these seemingly indistinctive cumbersome of a single user, they will also try putting fragment information together and obtain a user image as perfect as possible, and if this happens, users’ privacy is in great danger. Second, electronic finance has won favor of the public due to its convenience and efficiency, but in the meantime, security flaws still exist in even advanced preventive measures like face recognition and fingerprint payment. Thus, how to improve the poor security situation and perfect the e-commerce consumer privacy protection mechanism is a thorny problem to the long-term development of e-commerce.
2.2. Definition of consumer privacy and reasons for consumer privacy leakage
The concept of privacy is pretty extensive since it is involved in numerous fields of social science. Nevertheless, scholars from all walks of life have not agreed on a precise and unified concept of privacy so far. In a manner, privacy is described as a multidimensional, flexible and dynamic concept, and it changes with the experience of life. It is an overlap concept which is both secret, confidential and anonymous, safe and ethical, meanwhile depending on special situations (such as time, location, occupation, culture and reasons), therefore it being extremely hard to be defined in general[2]. From the perspective of e-commerce consumers in the era of big data, the traditional privacy definition is changing from information regarded too sensitive for a person or a group to let others know (such as diseases and bank deposits) to information that is able to identify personal characteristics (such as id number and daily life tracks).

Correspond to the definition of consumer privacy above, this paper’s point of view on consumer privacy leakage is incidents that any information concerning consumers’ identity and therefore has any economic value was inadvertently abused, maliciously stolen or illegally traded. The cause of consumer privacy leakage is complex and extensive, which can be roughly classified into lack of personal information protection consciousness, economic value of privacy information being impressive and imperfection of relevant legal system. Driven by factors given above, online merchants will make use of bias privacy policy to maximum interests of their own, which results in consumer privacy leakage.

3. Current situation of privacy protection in e-commerce
E-commerce is a booming field in the past resent few years as public dependence towards the internet is gradually enhanced. Online consumer privacy protection has also become one of the hottest topics at present. The lack of privacy protection technology and relevant legislation on e-commerce platforms undoubtedly brings huge resistance to online consumer privacy protection.

3.1. Observation of e-commerce consumer behavior
Specific online consumer behavior is sure to be closely related to any privacy leakage that may be caused by a series of links of shopping on the Internet. In the process of online shopping, there are two kinds of consumer behavior that lead to the disclosure of consumers' privacy. Firstly, entering sensitive information (such as website or bank account password) in network environment is the main cause of direct consumer privacy leakage. Many consumers are accustomed to relying on the password memorizing function provided by the website, but are often lack of clear concept and high alert to the detailed record principles that cookies on the devices will follow behind screen. Secondly, binding registration regulations and sharing behavior in online shopping also have a high risk of indirectly revealing consumers' privacy. When making online purchase, it is inevitable for consumers to register on the online platform provided by merchants and fill in relevant personal information (name, id number, etc.). In this way, our personal information is tied together indiscriminately. Rapid development of internet technology also makes increasingly keen on sharing shopping information. However, what they do not know is that such sharing behavior will virtually strengthen the association between two accounts and further expose the common consumption habits of both parties.

3.2. Recent trend of privacy protection
The high degree of freedom and interactivity of big data result in great changes in the traditional way people make purchase. Nevertheless, information processing technology is also improving in the mean time, leaving many privacy vulnerabilities for criminals. Information leakage in online shopping is not individual case any longer, but a focus issue affecting the development of the entire e-commerce industry. For example, e-commerce merchants can reorganize the original data set according to data relevance by multi-directional processing seemingly useless information, so as to obtain the consumer privacy that ought to be hidden in the first place. It is quite obvious that traditional privacy protection technologies are not enough to withstand the further developed internet mechanism, and are unable to
cut off the way consumer privacy leakage takes place on e-commerce platform. A brand new theoretical mechanism of privacy protection, which is more suitable for the development demands of the era of big data and of better quality, needs to be developed and applied urgently.

4. The Analysis of Privacy Protection in E-commerce

4.1. Privacy protection in Differential privacy

Differential privacy is a new definition of privacy mechanism, proposed by Dwork in 2006, which is different from traditional privacy ones[3]. Under the definition of differential privacy, the result of the data set computation is insensitive to the change of a specific record. Therefore, the risk of privacy disclosure of a record due to its addition to the data set is controlled in a very small and acceptable range. Attackers cannot obtain accurate individual information by observing the calculation results[4]. In recent years, a great many achievements have been made in the research of differential privacy in numerous fields. Based on existing researches, this paper summarizes and assumes the theoretical development of differential privacy and its application in the field of consumer information protection in e-commerce.

4.1.1. Definition

Differential privacy is a vital law on defining personal privacy, with extremely rigorous mathematical definition while the common output is usually the result of some kind of data mining, whose content is determined by the type of query. On the e-commerce platform, the common output can be expressed as the identity query result in the form of "whether a customer buys a certain item", the count query result in the form of "how many customers buy a certain item in total", or the histogram query result in the form of "how many customers buy a certain item every four hours".

The definition of differential privacy and its applicability to consumer privacy protection on e-commerce platforms cannot be understood without considering the situation that the common output will face in reality. When the consumer data set D is collected by the e-commerce platform, traditional privacy information processing method can theoretically protect personal privacy to a certain extent by deleting identifier attributes (such as name, ID card and ID number). However, this is far from enough in practice. Other attributes without identifiabilities themselves also exist in the dataset collected, such as birthday, gender, place of residence, browsing and purchase record. It is of great chance that an attacker combines the information with identifier attributes obtained from elsewhere with information that has no identity symbols in D to obtain the intersected data set T, which represents the victim set of privacy leakage.

In view of the above situation, the new privacy mechanism, differential privacy, proposed in 2006 will solve the real problem that traditional privacy cannot satisfy. Differential privacy describes the contribution of individual privacy to the common output. By adding noise to the original data, privacy attackers cannot infer the data content other than query results based on background knowledge.

The theoretical definition of differential privacy is given below, with reference[5].

**Definition 2.1 (ε-Differential Privacy)** An algorithm A satisfies ε-differential privacy (ε-DP) if and only if for any datasets D and D' that differ on one element, we have

\[
\forall t \in \text{Range}(A) : \Pr[A(D) = t] \leq e^\epsilon \Pr[A(D') = t]
\] 

(2.1)

The condition (2.1) is equivalent to

\[
\forall t \in \text{Range}(A) : \Pr[A(D) = t] \leq e^\epsilon \Pr[A(D') = t]
\] 

(2.2)

Where D and D ' are adjacent data sets with only one set of data difference; A(D) and A(D ') are output results based on K. \Pr[A(D)] belongs to Sj is the risk of privacy leakage. The privacy parameter \epsilon is the privacy budget. The smaller the value of \epsilon, the higher the privacy protection
In short, differential privacy ensures that individual data has no impact on the common output of the data set, making it impossible for attackers to obtain individual privacy through public output and background knowledge, thus achieving the purpose of privacy protection.

4.1.2. Technology

A mechanism that satisfies the differential privacy definition can be seen as an interface between the private data set and the common output, ensuring that all common output about a particular query is similar. In other words, the common output is almost independent of any individual, and the degree of dependence is controlled by the privacy budget set by the data owner. That is to say, the existence or absence of an individual cannot significantly change the public output in this mechanism. The main method to implement the differential privacy mechanism is to add noise, and the two commonly used noise mechanisms are Laplace mechanism and exponential mechanism respectively, and Laplace mechanism satisfies differential privacy protection by applying appropriate noise to the data of numerical query results, which we will make further discussion.

Suppose function $f$ is a query function providing $\epsilon$-DP, $f(n) = \text{count}(I) + \text{noise}$, where noise is a noise subject to some random distribution. Let's say $X$ is equal to noise, then we have

$$\forall t, \frac{\Pr[f(D) = t]}{\Pr[f(D') = t]} = \frac{\Pr[\text{count}(D) + X = t]}{\Pr[\text{count}(D') + X' = t]} \leq e^\epsilon$$

Take $d = \text{count}(D) - \text{count}(D')$, then we can obtain

$$\forall x, \frac{\Pr[X = x]}{\Pr[X' = x + d]} \leq e^\epsilon$$

In order to ensure that the above formula is constant, we need to ensure that $d$ is less than or equal to delta $f(\Delta f)$, delta $f$ (sensitivity) is defined as follows

$$\Delta f = \max_{D \neq D'} |\text{count}(D) - \text{count}(D')|$$

And from here we can see that the distribution of the noise is actually the Laplace distribution

$$\text{Lap}\left(\frac{\Delta f}{\epsilon}\right)$$

we take it as $\beta = \frac{\Delta f}{\epsilon}$, then we have

$$\Pr[\text{Lap}(\beta) = x] = \frac{1}{2\beta} e^{-|x|/\beta}$$

$$\frac{\Pr[\text{Lap}(\beta) = x]}{\Pr[\text{Lap}(\beta) = x + d]} \leq e^\beta \leq e^\beta = e^\epsilon$$

Then can be obtained, which satisfies the mathematical definition of differential privacy.

In short, Laplace mechanism is to add noise to the output of the original function. When noise of $f$ is 0, that is, when count value is the output, $Pr$ probability is at the highest level. When $f(D)$ create noise with the probability of $Pr$, $f(D')$ create noise in the probability of $Pr'$, they output the same value equals the value of count’, and the $Pr$ and $Pr'$ probability ratio are controlled within the $\exp(\epsilon)$[6].
4.1.3. Application
Due to good characteristics, differential privacy has developed into a recognized privacy standards, and is applied in gradually from academia to industry. Google[7] took the lead in deploying differential privacy mechanism in Chrome browser in 2014, so as to discover the phenomenon of webpage hijacking without knowing the settings of users’ home pages. Apple[8] announced on WWDC in 2016 that the differential privacy mechanism was deployed in iOS10 to count the frequency of emoji used and update the dictionary without knowing users’ keyboard input. Microsoft[9] also began to collect application usage data by using differential privacy mechanism in the creator update of Windows in the autumn of 2017.

In conclusion, if the thriving e-commerce platform wants to achieve steady and long-term development, it must possess professional knowledge of privacy protection to make necessary modifications to the traditional data mining mechanism, so that the new mechanism can meet the requirements of differential privacy protection. In this way, the security of user data sets which maintain the operation of e-commerce platforms will be ensured, and the privacy information of consumers will be protected. This complies with the basic literacy requirements of data owners in the era of big data.

4.2. Privacy protection in legal level
The progress of computing ability, storage capacity and advanced network technology enables companies a wider range of collection, processing and connection of data. What’s more, companies now tend to put data into various use, such as personalized service and marketing, which on the legal level, leaves even bigger obstacles in personal information protection. Although the personal information protection act has long been included in the legislative agenda of our China, due to its complexity, however, the process has been very slow. By far, no specialized laws concerning personal information protection or privacy rights have been issued, some of the relevant regulations are only involved in other laws in haste. Because the concepts of these laws and regulations are too general and not systematic, along with the ambiguity in the definition of personal information and privacy in China, victims of privacy leakage incidents on e-commerce platforms in the country often suffer losses when they resort to law, while the illegal merchants enjoy profits as owners of vast amount of information.

On the contrary, as early in the 1960s, data privacy laws have first appeared in European and American countries. In May 2016, the European parliament voted to pass the new General Data Protection Regulation (GDPR), which aims to more vigorously curb the abuse of personal information, strengthen the protection of personal privacy, and then force to replace the original data protection directive in May 2018[10]. After the implementation of GDPR, more than 500 million people in 28 member states of the European Union can be directly protected by it. It is worth mentioning that although the UK has started the process of leaving the European Union, GDPR has also been approved and officially launched since May 25 in the country. It is without doubt that GDPR will be another historic shift in the process of personal privacy protection in Europe.

According to the provisions of GDPR, enterprises should obtain the consent of users in the collection, storage and use of personal information, and users should have absolute control over their personal data. Among the 12 updated user rights in GDPR, the right to be forgotten, the right to restrict processing and the right to transfer data are most impressive, and the implementation of the former two is of great practical significance to the protection of consumer privacy on e-commerce platforms. At the policy level of China, we should strengthen the reflection of normative mechanisms in these aspects, carry out severe penalty towards illegal acts and cases that seriously infringe upon personal privacy, and effectively establish the protection system of personal privacy security.

4.3. The Optimization of Privacy Protection in E-commerce
China is a country with a large population base. According to statistics, the number of internet users in China accounts for one-fifth of the total number of internet users in the world. With the development of internet technology and electronic data processing technology, the huge amount of personal
information brought by numerous users under several major e-commerce platforms in China is facing great challenges.

The organic combination of multiple effective approaches is urgently needed to solve this problem. Firstly, individual awareness of privacy protection should not be underestimated. Consumers who participate in e-commerce activities should be more sensitive to protect personal information and keep a vigilant eye on potential privacy leakage flaws in any process involving registration and payment. Secondly, laws and regulations concerning network privacy security also need to be further concerned by the public. It is necessary for Chinese government to be deeply aware of the seriousness of the problem and strengthen its determination to introduce laws and regulations into solving it. The last and most important measure is to introduce the differential privacy protection mechanism, which is more perfect than the traditional privacy protection technology, into the actual operation process of information processing on e-commerce platforms, so as to fundamentally reduce or even eliminate the occurrence of online consumer privacy leakage.

5. Conclusion
The rapid development big data has brought huge opportunities as well as challenges to the field of e-commerce. Nowadays, with the development of mobile internet technology, information carried by the internet is more and more easy to be obtained, and the consequent personal information leakage brings about increasingly acute contradictions. Privacy leakage has gradually become one of the hottest topics in today’s society. From the perspective of the current situation of consumer privacy security on e-commerce platform, this paper has done the following work to improve the consumer privacy security of this platform based on differential privacy technology.

(1) On the basis of reading and studying relevant literature at home and abroad, the concepts of online consumer information privacy and online consumer privacy leakage are sorted out and summarized. Meanwhile, combined with the information interactivity of online shopping behavior on e-commerce platform in comparison with traditional off-line shopping behavior, this paper proposes a new understanding and definition of online consumer privacy leakage.

(2) In the research on the theory of differential privacy technology and its existing industrial applications, this paper proposes a new idea of introducing this privacy protection technology into the field of consumer privacy protection on e-commerce platforms. Differential privacy technology has a good feature of privacy ambiguity, while e-commerce is one of the most focal platforms in data mining and processing nowadays. The combination of these two will certainly be a key step and an inevitable trend of the development of the era of big data.

Reference:
[1] Sohu News. http://www.sohu.com/a/119101428_197955.
[2] Liu Yahui, Zhang Tieying, Jin xiaolong, Cheng Xueqi. Personal privacy protection in the era of big data. Computer research and development. 2015, 52 (1): 229-247.
[3] C. Dwork, F. Mcsherry, K. Nissim, A. Smith. Calibrating noise to sensitivity in private data analysis, in Proceedings of Theory of Cryptography, Springer, 2006: 265-284.
[4] Xiong Ping, Zhu Tianqing, Wang Xiaofeng. Differential privacy protection and its application. Journal of computer science, January 2014, volume 37, issue 1.
[5] Ninghui Li, Min Lyu, Dong Su, Weining Yang. Differential Privacy: From Theory To Practice. 2016.
[6] Open Source in China. https://my.oschina.net/keyven/blog/730740.
[7] U. Erlingsson, V. Pihur, and A. Korolova, “Rappor: Randomized aggregatable privacy preserving ordinal response,” in Proceedings of the 2014 ACM SIGSAC conference on computer and communications security, ACM, 2014, pp. 1054–1067.
[8] Apple Differential Privacy Team, “Learning with privacy at scale,” Apple Machine Learning Journal, vol. 1, 8 Dec. 2017.
[9] B. Ding, J. Kulkarni, and S. Yekhanin, “Collecting telemetry data privately,” in Advances in Neural
Information Processing Systems, 2017, pp. 3574–3583.

[10] Christina Tikkinen-Piri, Anna Rohunen, Jouni Markkula. EU General Data Protection Regulation Changes and implications for personal data collecting companies. computer law & security review 34 (2018) 134–153.