Available online at www.HighTechJournal.org

HighTech and Innovation Journal

ISSN: 2723-9535

Vol. 5, No. 2, June, 2024

A Consumer Data Privacy Protection Model Based on Non-Parametric Statistics for Dynamic Data Publishing in e-Commerce Platforms

Jiao Jia [1*]

[1] *Southwest Jiaotong University Hope College, Chengdu, Sichuan, 610400, China.*

Received 07 December 2023; Revised 24 April 2024; Accepted 08 May 2024; Published 01 June 2024

Abstract

Objectives: Consumer data privacy on e-commerce platforms is increasingly crucial. This study aims to investigate privacy protection mechanisms, particularly focusing on personal and corporate secrets. It seeks to understand individual perspectives on privacy and preferences for data disclosure. The primary objective is to explore methods for safeguarding personal information while maintaining data integrity. *Methods/Analysis:* We employ non-parametric statistical techniques to analyze consumer behavior and preferences on e-commerce platforms. This involves examining patterns of data disclosure and identifying sensitive information shared by users. By studying communication dynamics and recording practices, we assess the efficacy of current privacy protection measures. *Novelty/Improvement:* This study contributes to the understanding of consumer privacy protection by emphasizing the importance of non-parametric statistical methods in e-commerce research. Our findings underscore the need for enhanced privacy measures. We advocate for further research and development of innovative privacy-enhancing technologies to address evolving privacy challenges in online commerce. *Findings:* Our research highlights the significance of personal privacy concerns in e-commerce settings. We identify a spectrum of privacy attitudes among users, ranging from strict confidentiality to selective disclosure. Furthermore, our analysis reveals potential vulnerabilities in current privacy safeguards, particularly regarding the collection and storage of sensitive data on e-commerce platforms.

Keywords: Non-Parametric Statistics; Dynamic Data; Electronic Commerce; Consumer Privacy.

1. Introduction

Since everyone lives in a digital age and can access data with the help of modern-aided technologies, mobile security has never been more important. When taking lessons, students were routed to e-commerce platforms, shoppers were able to access e- and mobile commerce, and office workers were able to access online work-from-home modes of internet access both during and after the COVID-19 pandemic. A complete strategy for protecting customer privacy in online commerce environments is suggested by the customer Data Privacy Protection Model, which is based on non-parametric statistics for dynamic data publishing on e-commerce platforms. To address the changing issues of privacy protection, this strategy combines dynamic data publishing procedures with sophisticated statistical methodologies [1, 2].

The approach permits the investigation of customer behavior and preferences while upholding data confidentiality by utilizing non-parametric statistical techniques. The privacy of sensitive data saved on e-commerce platforms is protected by the use of encryption methods and secure communication routes. Dynamic data publication techniques are included in the concept, which permits information to be securely distributed while maintaining customer

* Corresponding author: jiajiao870909@163.com

http://dx.doi.org/10.28991/HIJ-2024-05-02-013

confidentiality. This approach attempts to provide a strong framework for customer data privacy protection in the dynamic and fast-paced world of e-commerce platforms using a combination of encryption, statistical analysis, and dynamic data management.

For dynamic data publishing on e-commerce platforms, the Consumer Data Privacy Protection Model based on non-parametric statistics is a state-of-the-art solution meant to tackle the complex issues related to consumer privacy in online purchasing environments. Fundamentally, this approach acknowledges how crucial it is to protect personal data while maintaining the smooth operation of e-commerce platforms. The concept protects individual privacy while enabling sophisticated consumer data analysis through the use of non-parametric statistical methods [3, 4].

The study investigated the moral ramifications and potential for data phishing by consumers who used unencrypted electronic apps downloaded on cell phones through m-commerce sites. Using a combination of methods, the study checks the secure information of M-commerce data, conducts semi-structured discussions with consumers and industry experts, and analyzes pertinent literature [5]. The study was based on conventional hidden stochastic modeling. Notably, we employed a model-free method that eliminates a specific reaction time distribution shape. This has the significant benefit of preventing results that could be untrustworthy when an incorrect reaction time allocation is anticipated [6]. The study proposed a dual-channel feature extraction module-based multivariate time-series anomaly detection approach. Using the spatial characteristics of a short-time Fourier transformation (STFT) and the graphed attentiveness system, respectively, the module targeted the spatial and temporal characteristics of the multivariate information. The efficiency of the model in detecting anomalies is therefore enhanced by fusing the two characteristics [7]. The article addressed customer privacy security between crypto-currency and adaptable clients, as well as the significance of e-commerce online purchasing software's dependability. Based on that, the article used theoretical inquiry and empirical investigation to explore the effects of consumer platform-based shopping portals on credibility as well as the elements and mechanisms that govern consumer shopping behavior. It also creates a survey verification table for block-chain technology and mobile client privacy safeguarding [8]. The research exposed a variety of factors that contribute to security lapses that compromise the integrity of online transactions. The report identified a range of actions that organizations could take to combat the growing risk to the security of web-based companies. Users' attention has been drawn to discussions about privacy and security issues in the field of information sciences and data privacy [9]. The study suggested a threshold secret exchange system with a verified threshold homomorphic encryption approach to create a secure and provable statistical analysis strategy for an e-commerce platform. By using a unique distribution model to provide secret shares, our method reduced the need for secure channels by approximately 40% when compared to a typical criterion privacy sharing scheme [10].

The limitations of earlier research are addressed in this study, which focuses on the following areas: consumer perceptions and behaviors regarding data privacy in e-commerce; the effectiveness of current security measures in online transactions; and the lack of research on workable solutions for improving privacy and security in e-commerce platforms, specifically about data phishing and cryptography techniques.

Contribution

The paper presents a thorough analysis of privacy protection strategies, focusing on corporate and individual secrets. This richness makes it possible to comprehend the several facets of privacy concerns in e-commerce in a nuanced manner. The research improves the depth and dependability of the analysis by introducing novel analytical strategies that are suited to the complexity of e-commerce privacy data through the use of non-parametric statistical methods.

It highlights the need for further research and development of novel privacy-enhancing technologies in e-commerce, defines a range of user privacy attitudes, highlights flaws in the security measures in place, and promotes tailored privacy solutions. The second part of this study provides the method of this research, the third part explores the result analysis, and the fourth part discusses the study's conclusion.

2. Research Methods

In methodology, we used the algorithm of correlation factors of non-parametric statistical and multiple correlations that described the coefficients of correlation in four steps. Figure 1 shows the process of methodology.

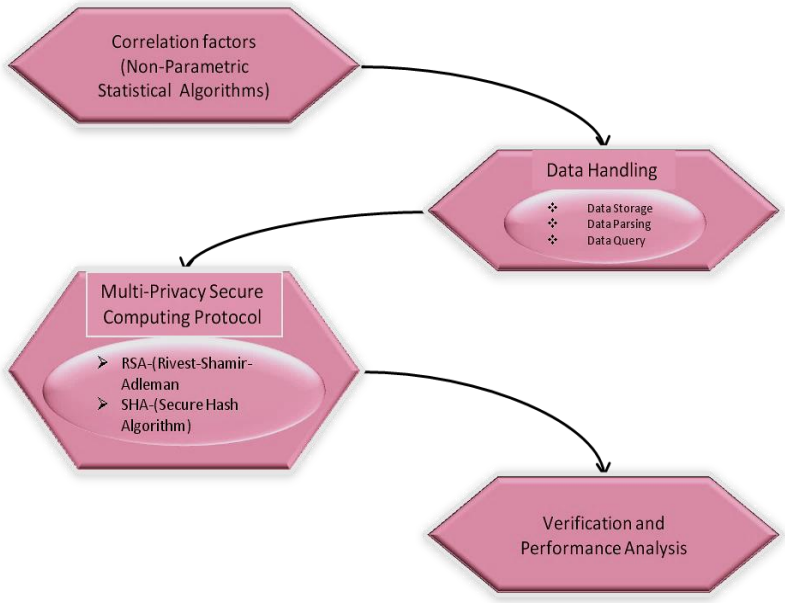

Figure 1. The process of methodology

2.1. Correlation Factors of Non-Parametric Statistical Algorithms

The concept of "correlation" is derived from the investigation of height genes in the human body. Correlation can be defined as "when one variable changes, another variable changes" and the statistics that measure correlation is a correlation. In these data, the duty is bigger and the correlation degree is higher, but the duty is low and the correlation degree is low. In addition, the link is directional, when one variable increases, this change is called direct ratio, reduced when the other variables, this change is called negative correlation.

$\mathrm{cov}(X, Y)$ Covariance measures the degree to which two random variables vary together. If the covariance is positive, it indicates that as one variable increases, the other tends to increase as well and vice versa. If the covariance is negative, it means that as one variable increases, the other tends to decrease.

Here $\sigma_X\sigma_Y$ denotes the standard deviation that measures the spread or dispersion of a random variable's values around its mean. It provides a measure of how much the values of a random variable deviate from the mean. μX and μY represent the means (or expected values) of the random variables X and Y respectively. E (Expectation operator), this operator calculates the expected value, which is the mean of a random variable when averaged over all possible outcomes.

The relationship number is defined as the quotient of the covariance and standard deviation between two variables:

$$\rho_{X,Y} = \frac{\mathrm{cov}\,(X,Y)}{\sigma_X\sigma_Y} = \frac{E[(X-\mu X)(Y-\mu Y)]}{\sigma_X\sigma_Y} \quad (1)$$

This is the population correlation coefficient. By estimating the covariance and standard deviation, between two sets of data points *X* and *Y*, where n is the number of data points, X_i and Y_i are individual data points from each set and $\bar{X}$ and $\bar{Y}$ are the means (or averages) of the respective data sets. r Indicates a perfect positive linear relationship, Pearson correlation coefficient can be obtained:

$$r = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \quad (2)$$

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma X}\right)\left(\frac{Y_i-\bar{Y}}{\sigma Y}\right) \quad (3)$$

where the standard score, the sample mean and sample standard deviation of the sample respectively. $\frac{X_i-\bar{X}}{\sigma X}$、$\bar{X}\sigma X X_i$ Since it is similar, $\mu_x = E(X), \sigma_X{}^2 = E[(X - E(X)^2)] = E[X^2] - E^2[X], Y.$

And:

$$E[(X - E(X))(Y - E(Y))] = E[XY] - E[X]E[Y] \quad (4)$$

Therefore, the correlation coefficient can also be expressed as:

$$\rho_{X,Y} = \frac{E(XY)-E(X)E(Y)}{\sqrt{E(X^2)-(E(X))^2}\sqrt{E(Y^2)-(E(Y))^2}} \quad (5)$$

For the sample Pearson correlation coefficient:

$$r_{xy} = \frac{\sum x_i y_i - \overline{nxy}}{(n-1)s_x s_x} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}} \quad (6)$$

Partial correlation coefficient: It's a random variable. $X, Y, \xi_1, \xi_2, \ldots, \xi_p p + 2$

When the partial correlation coefficient refers to the correlation coefficient and after removing the influence of the third variable: $p = 1\xi_1 X.YXY$.

$$r_{XY1} = \frac{(r_{XY} - r_{Y1}r_{X1})}{\sqrt{1-r_{X1}^2}\sqrt{1-r_{X1}^2}} \quad (7)$$

This represents Pearson correlation coefficient. $X, Y, r_{X1}, r_{r_1}XY, X\xi_1, Y\xi_1$. When the type of promotion, the partial correlation coefficient is the number of the correlation between and after removing the influence. The formula is as follows:

$$p = 2\xi_1,\ \xi_2 X, YXY \quad (8)$$

$$r_{XY12} = \frac{(r_{XY1} - r_{Y21}r_{X21})}{\sqrt{1-r_{X21}{}^2}\sqrt{1-r_{Y21}^2}} \quad (9)$$

$$r_{XY133} = \frac{\left(r_{XY12} - r_{Y312}r_{X312}\right)}{\sqrt{1-r_{X312}^2}\sqrt{1-r_{Y312}^2}} \quad (10)$$

Semi-partial correlation coefficient:

If only to have an impact, it's called a semi-partial correlation coefficient. The formula is as follows: $\xi_1 x$

$$r_{Y(X11)} = \frac{(I_{XY} - r_{Y1}r_{X1})}{\sqrt{1-r_{X1}^2}} \quad (11)$$

Multiple correlation coefficients:

Describe the multiple correlation coefficients with the correlation between variables, mainly divided into four steps:

***Initialization*:** analysts must be in Certificate Authority (CA) identifying the certification bodies. The CA allows the analyst to generate public keys and authentication keys at the beginning [11, 12].

***Data collection and storage*:** When massive data is generated, the remaining data in each processing process is collected by the collection center, distributed randomly through a secure channel and transmitted to multiple data servers for storage.

***Cipher text operation process*:** The user's data is received by each data server. Implement the confidential data between multiple servers and network steps to save the cipher.

***Information retrieval*:** The information server to confirm the user's recognition and instructions after receiving message parsing instructions will conform to the requirements of the data back to the analytic program. Analysts in their private keys are used to encrypt and classify the corresponding analysis conclusions.

The mode of the database is a database that contains multiple data servers. In this system, all data is true, that is to say, all the data that follow the convention and the execution time, will be as far as possible from the execution of this agreement for more information. In the process of use, there are more than three data, whereas, in a real environment, the number of data servers will be more than three. In this mode, when multiple data servers exist at the same time, all the servers in this mode can guarantee their security, so that one of the servers in this mode cannot be attacked.

This model mainly studies the inside and outside of the two different ways of being offensive and defensive. Here, outside of the attack, it was aimed at the pattern of the attacker hoping to obtain the user's data from the system without knowing the private key. Internal damage is, for example, a malicious action by a system participant, such as a server or an analyst, or if they want to learn and recover a confidential message from a program. The model includes three aspects: between the user and the server, the server, and the analysis of the communication with the server [13, 14].

User profiles can be decimal numbers. Under this condition, the data collection center puts the data into an integer and the distributed data, data unit, data and other data are transmitted to the corresponding three data servers. The collection center is classified into three and its transmission through the security channel respectively to the data server, the server for the following operations: $S_1,\ S_2,\ S_3$.

S_1- Select a random number, to encrypt, get the cipher-text and storage; $r_1 \in Z_N^* \lambda\ C_\lambda = g^\lambda r_1^N (\mathrm{mod} N^2)$

S_2- Select a random number, to encrypt, get the cipher-text and storage; $r_2 \in Z_N^* \mu\ C_\mu = g^\mu r_2^N (\mathrm{mod} N^2)$

S_3- Select a random number, to encrypt, get the cipher-text and storage. $r_3 \in Z_N^* v C_v = g^v r_3^N (\mathrm{mod} N^2)$

Because it has the following properties:

$$\begin{aligned} C_\lambda * C_\mu * C_v = \\ &= E(\lambda, pk)\, E(\mu, pk)\, E(v, pk) \\ &= \left(g^\lambda r_1^N\right)\left(g^\mu r_2^N\right)\left(g^v r_3^N\right)(\mathrm{mod}N^2) \\ &= g^{\lambda+\mu+v}(r_1 r_2 r_3)^N(\mathrm{mod}N^2) \\ &= E(\lambda+\mu+v, pk) \\ &= E(\eta, pk) \end{aligned} \tag{12}$$

So, the encrypted information is stored in three servers, only the cipher-text multiplication and in the case of knowing the private key can get complete information. λ、μ、νη. So, the illicit close sex of the user's data can be guaranteed, the premise is the three data servers that do not separate the data together.

2.2. Based on Nonparametric Statistics Consumer Information of Electric Business Platform Security

Safety Guarantee 1: Data storage security, information storage security, and cipher key storage security certificate: early in the program, through secret channels, open authentication, and encryption. In this article, both methods are based on the minimum version of Secure Hash Algorithm (SHA)-3, r=40, and SHA-3 is the minimum, which provides efficient security for most programs. In the process of data collection, using SHA-3 produced an arbitrary number. Due to the use of a smaller cryptographic system, data confidentiality, authenticity, and integrity can be maintained between the receipt center and the data servers. The user and the server communication, according to SHA-3, produces information isolation if it is from the outside or if the server was invaded or colluded with other servers. In the absence of a key, the user is unable to determine the block, and data segmentation is necessary to avoid an internal attack on the data server. So, the data stored on the server is guaranteed.

$\frac{C}{D}$ Likely represents the result of a cryptographic operation, perhaps encryption or decryption.φ represents Euler's totient function.$\sum_{i=1}^{n} x_i y_i$ This indicates a summation operation, where $x_i$and $y_i$are variables that are summed up from $i = 1\ to\ n$. pk^r represents raising a public key pk to the power of r.

$g^{r(sk_1+sk_2+sk_3+sk_4)}$ This part appears to involve exponentiation of a base g raised to the power of the sum of multiple secret keys $sk_1 + sk_2 + sk_3 + sk_4$all raised to the power of r. $(\mathrm{mod}N^2)$This indicates that the entire expression is taken modulo N^2, where N is likely a large composite number.

The password we can obtain is Analysts calculated according to the private key: sK denotes the secret key, A is likely some input value, $and\ D$ is the result of the operation, computed by rising A to the power of sk_4 modulo N^2.

$$sk_4 D = A^{sk_4}(\mathrm{mod}N^2) \tag{13}$$

So, there are:

$$\begin{aligned} \frac{C}{D} &= \frac{\phi^{\sum_{i=1}^{n} x_i y_i} pk^r}{g^{r(sk_1+sk_2+sk_3+sk_4)}}(\mathrm{mod}N^2) \\ &= \frac{\phi^{\sum_{i=1}^{n} x, y_i} g^{skr}}{g^{r(sk_1+sk_2+sk_3+sk_4)}}(\mathrm{mod}N^2) \\ &= \phi^{\sum_{i=1}^{n} x_i y_i}(\mathrm{mod}N^2) = (1+N)^{p\sum_{i=1}^{n} x y_i}(\mathrm{mod}N^2) = 1 + (p\sum_{i=1}^{n} x_i y_i)N(\mathrm{mod}N^2) \end{aligned}$$

We can get,

$$\begin{aligned} T_{xy} &= L\left(C/D(\mathrm{mod}N^2)\right)/p \\ &= \frac{1+(p\sum_{i=1}^{n} x_i y_i)N)-1}{Np} \\ &= \sum_{i=1}^{n} x_i y_i \end{aligned} \tag{14}$$

Similarly, through the above analysis, we can also, verify the effectiveness of the method. $\sum_{i-1}^{n} x_i a_i, \sum_{i-1}^{n} y_i a_i$So, during the period of data storage, three separate integer confidential information is transmitted via a secret channel to three data servers and by the nonparametric statistics password of algorithms for data security operation in security fields, are stored in the server S values. The value of S typically depends on the specific cryptographic protocol that is used. If they don't have to deal with the problem of nonparametric statistics, the encryption program will become very safe and an outside attack also cannot access the file, according to the provisions of the agreement, as long as all the

passwords are multiplied by time and then decrypted, can obtain confidential information. The user data is carried out by the user's public key cryptographic processing, unless the three data servers do not spread its merger, otherwise, the user data we preserve the right of privacy, without a user's private key, is unable to obtain by the information users.

In the process of user registration, the user and the data server are done on a secret channel access and transmission of the private key. Because of three data servers and the user's computing device has large operation and transmission performance, therefore, Advanced Encryption Standard (AES) is a good safe passage. The key can be made by a public key encryption system, such as a Diffie-Hellman key exchange or Rivest-Shamir-Adleman algorithm (RSA). Encrypt using AES and Digital Signature Standard (DSS) channel, to ensure the user with the confidentiality, authenticity and integrity of the data server.

Safety Guarantee 2: The security of data parsing - even a breach of one or more servers accomplice, also won't leak personal data [15, 16].

The experimental results show that all of the data processing is carried out by three sets of data servers for encrypted information exchange, realizing the information exchange of three data servers. The user's data is always carried out by the user's public key password, even if the internal attack on the other two data servers, without the user's private key, is also unable to obtain the user's information. In addition, the data server can also use a public way, namely the three data communication between the server and the three sets of data to the servo communication, namely a password-based approach to secondary password exchange, to ensure the correctness of the three-servo end communication and integrity. The server is not credible, there is no crack, and even if someone within the attack or used the most important password, 8 could not parse out all the information. So, this method can guarantee the safety of data.

Safety Guarantee 3: The security of the data query is that only a legal order can get the result of the operation.

Proof: At the time of the visit, permissions authorized by the name will be signed by the user's manual key and then used by the server authentication key to ensure the validity of the license and its correctness using the above-mentioned safe passage for the analysts and server communication. To ask for information, the resulting statistics are encrypted, as shown in Figure 2, and only after the authentication and certification of authorized users can they use this password to crack it to prevent malicious invasion from the outside.

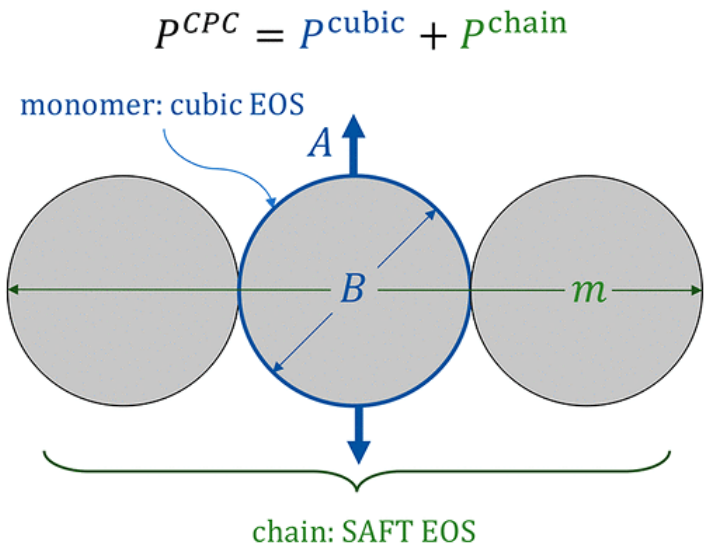

Figure 2. Based on the parameters of the dynamic data system

Even if people with authority did not receive the news, even with all the strength, the people can get the password but can only get a small piece of data. It is impossible to determine what kind of things they are in the absence of clear data. However, there is no place where statistical integration can ensure the patient's information. Even if an authorized user sends the password back through a server, he can't crack it without the cooperation of three servers, so the hacker will only collect some data but not disclose the user's data.

Ultimately, only the data that can be used can be obtained when the user cannot statistically understand the data of the individual user. In this paper, the digital signature criterion is selected to ensure the correctness and integrity of communication between users and servers. This method guarantees the security of the data [17].

The above explanation satisfies the security requirements for the system:

Data storage security: This system will use all data transmitted to multiple data servers for storage, and in the security operation of encrypted data, it can also ensure its security.

Data parsing security: No matter which server or multiple servers collude; the user's data will not be disclosed during the association calculation.

Security of data query: only certified analysts can enter the system, and the certified instructions can get the numerical value of the calculation, while unauthorized analysts and illegal instructions cannot get the algorithm. Even if the server and the analyst collude, the user's private data will not be disclosed.

3. Result Analysis

3.1. Dynamic Data Plays a Certain Role in the Protection of Consumer Information on e-Commerce Platforms

The primary protocol is a strong answer to the problem of heterogeneous consumer data scattered across various platforms. Utilizing non-parametric statistical encryption approaches, the protocol guarantees safe data transfer and analysis while maintaining user privacy. Secure communication lines and data server isolation strengthen the system's defenses against both internal and external threats. In Crypto++5.6.0 Benchmarks, it takes one microsecond to perform a 1,024-bit simulation. He writes programs in Microsoft Visual C++2005SP1 and uses AMD Holon 8354 on Linux. According to this conclusion, in a 1024-bit Pallier decryption system, a simulation with the rest of the Chinese theory takes a few microseconds. Taking the maximum computationally complex correlation as an example, when three data servers are networked by 100 Gb network sites, the computation and establishment time of communication are 1 minute and 1 second. The algorithm in this paper is to support parallel computing. If a data server is running 10 computers at the same time, the overall running time will be reduced to 16 seconds in the double correlation calculation.

The complexity of the three algorithms in the scheme is shown in Table 1.

Table 1. Complexity Analysis

	Method 1	Method 2	Method 3
Meter complexity	6 nm	12 nm	6 nm(m − 1)
Communication complexity	0 (nm)	0 (nm)	0 (nm^2)

In the process of cipher-text calculation, if each consumer has an attribute and three data servers cooperate, algorithm 1 is used to calculate equal cipher-text values (where is the number of combinations). $mC_x, C_y, C_a C_n^1 = mC_n^1$ Equal ciphertext values are calculated by algorithm 2; $C_{x^2}, C_{y^2}, C_{a^2} C_n^1 = m$ Equal ciphertext values are calculated by algorithm 3. $C_{xy}, C_{xa}, C_m C_n^2 = m(m-1)/2$. Finally, the statistical value of the correlation coefficient is calculated [18, 19].

Therefore, the computational complexity of the algorithm is respectively: algorithm 1 requires sub-modular operation in total and the computational complexity is 6 nm6 nm. In algorithm 2, a total of sub-modular operations should be performed, and the computational complexity is 12 nm12 nm. Algorithm 3 requires a total of sub-modular operations and the computational complexity is $6\,nm(m-1)6\,nm(m-1)$. The communication complexity is respectively: in algorithm 1, the data server needs to carry out round communication, and the communication complex complexity is $(3n-1)m0(nm)$; in algorithm 2, the data server needs to carry out round communication, and the communication complexity is O(nm). $(6n-1)m$; in algorithm 3, the data server needs to carry out roll communication and the communication complexity is $(6n-1)m(m-1)/20(nm^2)$. Figure 3 and Table 2 show the Execution time for different MBs, among the cryptography algorithms AES takes the lowest time to encrypt the data.

Table 2 shows execution times (in seconds) for the Advanced Encryption Standard (AES), Data Encryption Standard (DES), Elliptical Curve-Diffie-Hellman (ECDH), and BF cryptographic algorithms across different file sizes (MBs). As file size increases, execution times generally rise. AES tends to perform better than DES, ECDH, and BF across all file sizes.

Table 2. Execution time of cryptographic algorithms

File Size (MBs)	Execution Time (Sec)			
	AES	DES	ECDH	BF
100	15	17	17	18
200	24	26	27	25
400	56	58	59	57
600	81	84	85	84
800	105	117	116	115
1000	161	170	171	169

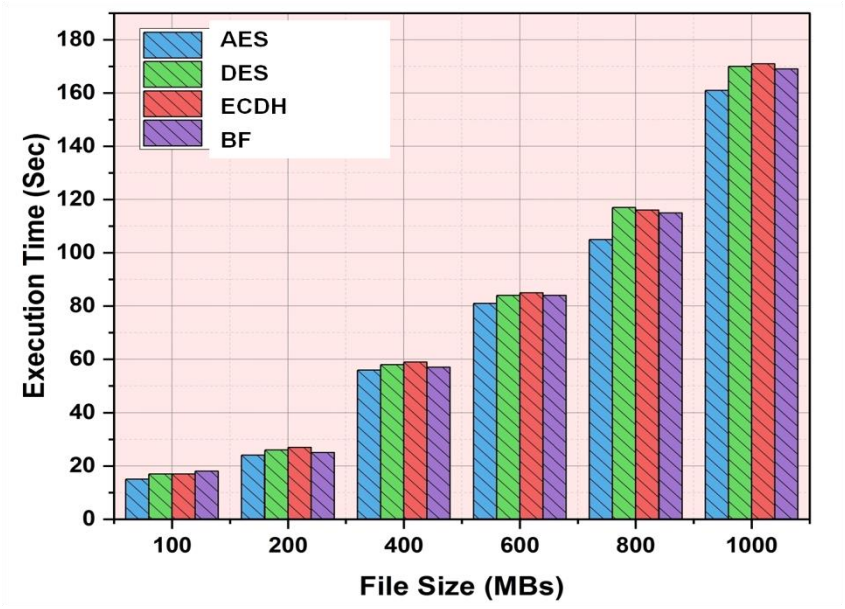

Figure 3. Execution Time of Cryptographic Algorithms

3.2. Multi-Information Security Operation Protocol Based on Non-Parametric Statistics

Problem description: Suppose that a consumer buys or browses goods in a single consumer unit and leaves private information in the data servers of multiple platforms. Due to the differences in time, place, unit, etc., the information left in each data server is different and incomplete, so each server has its own data for the same attribute, the attribute value of the consumer to remember, the information is stored in a server, and the need to each server calculation data contribute to participate in the multilateral security, each server and analysts can get a full analysis of consumer data but can't get any information of other participants storage, where; $\mathrm{nm}(s_n);\ \mathrm{mXix_i}(i = 1,2,\dots,n);\ \mathrm{mx_{i1}}, \mathrm{x_{i2}}, \dots \dots \mathrm{x_{iv}}$

To solve the above problems, this paper proposes a multi-party secure computing protocol (Protocol 1), in which each participant confidentially analyzes its data to obtain complete consumer analysis data, m. For the convenience of description, let $m = 3$. Enter a single attribute of a consumer, where; $\mathrm{nX} = (x_1, x_2, x_3, \dots, x_n);\ x_i = x_{i1} + x_{i2} + \cdots + x_{iv} i = 1, 2, 3, \dots,\ n;\ m = 3.$

Because of user data and server communications, each part of the user's data after a secret channel encryption is sent to the corresponding data server because each server is completely isolated and cannot understand other server information. If a server is not compromised, the system won't receive all the information if a foreign adversary is present. This method can protect the system from internal attacks. Therefore, the data stored on the server is very secure [20, 21].

In this program, the encryption is carried out by a non-parametric statistical encryption algorithm, it is stored in the medical application, and the security calculation is performed in the cipher text. If the problem of non-parametric statistics is not dealt with, the cryptographic system is protected, and the hacked person cannot access the files, thus ensuring the security of the key.

Aiming at the user data existing in different user units, the non-parametric statistical homomorphic key is used to encrypt the user, and the improved non-parametric statistical algorithm is used to establish multiple security measures between the user and the data server to ensure the security of the communication between the user and the data server and that the data will not be leaked at the same time. In our system, if there is no damage to any data, the safety of the user's personal information can be ensured. In addition, access control procedures ensure that only permissive analysis will yield qualified analysis results. On this basis, through the cooperation of three servers, the user's personal information is treated confidentially. Finally, the methods are verified, including reliability analysis and performance analysis. This method can resist both external attacks and internal attacks [22].

Encryption Time: Encryption time refers to the amount of time it takes for a cryptographic algorithm to transform plaintext into cipher-text. This process involves applying a specific encryption algorithm and possibly additional steps such as key generation and initialization vector (IV) generation.

Decryption Time: Decryption time refers to the amount of time it takes for a cryptographic algorithm to transform cipher-text back into plaintext. This process involves applying the decryption algorithm and, if applicable, using the decryption key.

For both encryption and decryption, the choice of algorithm can significantly impact the time taken to perform these operations. Some algorithms are designed for speed, while others prioritize security.

Hardware acceleration techniques, such as using dedicated cryptographic hardware or instruction set extensions, can reduce encryption and decryption times.

In scenarios where real-time processing is required, minimizing encryption and decryption times is crucial. This involves selecting algorithms and key sizes that balance security with performance requirements.

Table 3. Comparison of Cryptography algorithm's Encryption and decryption

Algorithm	Encryption Time (seconds)	Decryption Time (seconds)
AES	3.5	3.5
SHA	7.0	-
RSA	6.0	6.0

In summary, AES tends to have the fastest encryption and decryption times among the three algorithms. Table 3 provides rough estimates of the encryption and decryption times based on the common performance characteristics of these algorithms. AES generally outperforms SHA in terms of speed and security, while RSA is used for key exchange rather than direct encryption and decryption.

For RSA, the encryption and decryption times increase with the size of the data and key size. AES encryption and decryption times remain relatively constant regardless of the data size, assuming the same key size and implementation efficiency. SHA is a hashing algorithm and does not perform decryption. Therefore, only encryption times are listed for SHA.

4. Conclusion

With the rapid development of network technology, electronic commerce has transformed traditional trade practices, facilitating daily activities while introducing security challenges. To address these concerns, this paper aims to assess the correlation degree of data, particularly within the realm of e-commerce. The study utilizes correlation analysis based on customer shopping records to analyze the relationships between various indicators of customer data. This approach aims to safeguard customer information by identifying patterns, anomalies, and potential security risks. Additionally, several secure multilateral computing mechanisms are implemented to prevent network attacks effectively. Access to the network is restricted to authorized analysts only, who can perform complex correlation analysis on the data. Furthermore, the paper introduces some public key encryption methods, such as nonparametric statistics, to enhance security. Basic principles are discussed, laying a theoretical foundation for future algorithm design. In the context of big data, the paper constructs a data-based data segment storage and statistical dependency method for passwords. By leveraging advanced encryption methods, the paper aims to strengthen the overall security posture of e-commerce systems and ensure the privacy of user data. In terms of encryption and decryption performance, AES typically performs better than DES and ECDH; the choice of algorithm is based on several variables, such as security needs, compatibility, and the particular use case. The AES algorithm is employed to provide authorization and authentication for users. Through statistical analysis of data security measures, the protection of users' personal information from leakage is ensured. These efforts are aimed at bolstering user privacy and maintaining the confidentiality of sensitive data transmitted over e-commerce networks.

5. Declarations

5.1. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

5.2. Funding

The author received no financial support for the research, authorship, and/or publication of this article.

5.3. Institutional Review Board Statement

Not applicable.

5.4. Informed Consent Statement

Not applicable.

5.5. Declaration of Competing Interest

The author declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

6. References

[1] Yuniar, A. D. (2024). Thin privacy boundaries: proximity and accessibility of E-commerce privacy policy in young consumers of Indonesia. International Journal of Social Economics, 1-11. doi:10.1108/IJSE-11-2022-0740.

[2] Tao, S., Liu, Y., & Sun, C. (2024). Understanding information sensitivity perceptions and its impact on information privacy concerns in e-commerce services: Insights from China. Computers and Security, 138, 103646. doi:10.1016/j.cose.2023.103646.

[3] Tao, S., Liu, Y., & Sun, C. (2024). Examining the inconsistent effect of privacy control on privacy concerns in e-commerce services: The moderating role of privacy experience and risk propensity. Computers and Security, 140. doi:10.1016/j.cose.2024.103794.

[4] Mutambik, I., Lee, J., Almuqrin, A., Zhang, J. Z., & Homadi, A. (2023). The Growth of Social Commerce: How It Is Affected by Users' Privacy Concerns. Journal of Theoretical and Applied Electronic Commerce Research, 18(1), 725–743. doi:10.3390/jtaer18010037.

[5] Paulson, G. (2024). Assessing data phishing risks associated with unencrypted apps on smartphones with non-parametric test and random forest model: Insights from Kuwait phishing scam calls. Journal of Engineering Research (Kuwait), 1-7. doi:10.1016/j.jer.2023.09.017.

[6] Verdier, H., Laurent, F., Cassé, A., Vestergaard, C. L., Specht, C. G., & Masson, J. B. (2023). Simulation-based inference for nonparametric statistical comparison of biomolecule dynamics. PLoS Computational Biology, 19(2), 1010088. doi:10.1371/journal.pcbi.1010088.

[7] Xu, Z., Yang, Y., Gao, X., & Hu, M. (2023). DCFF-MTAD: A Multivariate Time-Series Anomaly Detection Model Based on Dual-Channel Feature Fusion. Sensors, 23(8), 3910. doi:10.3390/s23083910.

[8] Liu, R., & Wang, E. (2023). Blockchain and mobile client privacy protection in e-commerce consumer shopping tendency identification application. Soft Computing, 27(9), 6019–6031. doi:10.1007/s00500-023-08099-8.

[9] Srivastava, S., & Shobhna Jeet, D. R. (2023). E-Commerce and Privacy Issues. Russian Law Journal, XI(5), 2170-2175.

[10] Shen, H., Wu, G., Xia, Z., Susilo, W., & Zhang, M. (2023). A Privacy-Preserving and Verifiable Statistical Analysis Scheme for an E-Commerce Platform. IEEE Transactions on Information Forensics and Security, 18, 2637–2652. doi:10.1109/TIFS.2023.3269669.

[11] Ren, X. Y., Zhang, P., & Zhou, Y. Q. (2019). Distinct model on privacy protection of dynamic data publication. Cluster Computing, 22, 15127-15136. doi:10.1007/s10586-018-2506-3.

[12] Sharma, K. (2010). An Evaluation of Consumer Privacy Protection in E-Commerce Websites: A Comparative Study of Six E-Stores: Part II. EDPACS, 42(2), 1-19. doi:10.1080/07366981.2010.526040.

[13] Bresson, G., & Logossah, K. (2011). Crowding-out effects of cruise tourism on stay-over tourism in the Caribbean: Non-parametric panel data evidence. Tourism Economics, 17(1), 127-158. doi:10.5367/te.2011.0028.

[14] Wu, Y., Wang, W., Toll, M., Alkhoury, W., Sauter, M., & Kolditz, O. (2011). Development of a 3D groundwater model based on scarce data: The Wadi Kafrein catchment/Jordan. Environmental Earth Sciences, 64, 771-785. doi:10.1007/s12665-010-0898-3.

[15] Yonghui, Z., Su, L., & Phillips, P. C. (2011). Testing for Common Trends in Semiparametric Panel Data Models with Fixed Effects. Cowles Foundation Discussion Paper No. 1832. doi:10.2139/ssrn.1951892.

[16] Yang, C. (2011). Analysis on protection of e-commerce consumer network privacy. Procedia Engineering, 15, 5519-5524. doi:10.1016/j.proeng.2011.08.1024.

[17] Diwandari, S., & Hidayat, A. T. (2021, March). Comparison of classification performance based on dynamic mining of user interest navigation pattern in e-commerce websites. In Journal of Physics: Conference Series: IOP Publishing, 1844(1), 012025. doi:10.1088/1742-6596/1844/1/012025.

[18] Wang, X., Dai, H. N., & Zhang, K. (2019). Secure and flexible economic data sharing protocol based on ID-based dynamic exclusive broadcast encryption in economic system. Future Generation Computer Systems, 99, 177–185. doi:10.1016/j.future.2018.11.013.

[19] Sreedevi, E. P., Kattumannil, S. K., & Dewan, I. (2021). A non-parametric test for independence of time to failure and cause of failure for discrete competing risks data. Statistics, 55(5), 1107-1122. doi:10.1080/02331888.2021.1975712.

[20] Serrano, E., Such, J. M., Botía, J. A., & García-Fornes, A. (2014). Strategies for avoiding preference profiling in agent-based e-commerce environments. Applied Intelligence, 40(1), 127–142. doi:10.1007/s10489-013-0448-2.

[21] Qiuyang, G., Qilian, N., Xiangzhao, M., & Zhijiao, Y. (2019). Dynamic social privacy protection based on graph mode partition in complex social network. Personal and Ubiquitous Computing, 23, 511-519. doi:10.1007/s00779-019-01249-6.

[22] Ding, H., Peng, C., Tian, Y., & Xiang, S. (2019). A risk adaptive access control model based on Markov for big data in the cloud. International Journal of High-Performance Computing and Networking, 13(4), 464-475. doi:10.1504/IJHPCN.2019.099269.