





3-бөлім**Раздел 3****Section 3****Информатика****Информатика****Computer
Science**

IRSTI 81.93.29

DOI: <https://doi.org/10.26577/JMMCS.2022.v114.i2.09>

Narbayeva S.M.* , Tapееva S.K. , Turarbek A. , Zhunusbayeva S. 
Al-Farabi Kazakh National University, Kazakhstan, Almaty
E-mail: *narbaevasalta777@gmail.com

A MACHINE LEARNING MODEL BASED ON HETEROGENEOUS DATA

Big data is widely used in many areas of business. The information between organizations is systematically reproduced and processed by data, and the collected data differs significantly in attributes. By composing heterogeneous data sets, they complement each other, therefore, data exchange between organizations is necessary. In a machine learning collaborative learning process based on heterogeneous data, the current schema has many challenges, including efficiency, security, and availability in real-world situations. In this paper, we propose a secure SVM learning mechanism based on the consortium blockchain and a threshold homomorphic encryption algorithm. By implementing the consortium's blockchain, it is possible to build a decentralized data exchange platform, and also to develop a secure algorithm for the support-vector machine classifier based on threshold homomorphic encryption.

Key words: Blockchain, heterogeneous data, SVM, secure scheme.

С.М. Нарбаева*, С.К. Тәпеева, А. Тұрарбек, С. Жұнусбаева
Әл-Фараби атындағы Қазақ ұлттық университеті, Қазақстан, Алматы қ.
E-mail: narbaevasalta777@gmail.com

Гетерогенді деректерге негізделген машиналық оқыту моделі

Үлкен деректер бизнестің көптеген салаларында кеңінен қолданыс тапқан. Ұйымдар арасындағы ақпараттарды жүйелі түрде шығаруға және мәліметтерді өңдейді, жинаған мәліметтер атрибуттарда айтарлықтай ерекшеленеді. Гетерогенді мәліметтер жиынтығын құра отырып, олар бір-бірін толықтырады, сондықтан ұйымдар арасында мәліметтер алмасу қажет. Гетерогенді деректерге негізделген машиналық оқытуды бірлесіп оқыту процесінде қазіргі сызба көптеген мәселелерге ие, соның ішінде нақты жағдайларда тиімділік, қауіпсіздік және қол жетімділік. Бұл мақалада біз консорциум блокчейніне және шекті гомоморфты шифрлау алгоритміне негізделген қауіпсіз ТВӘ оқыту механизмін ұсынамыз. Консорциум блокчейнін енгізу арқылы орталықтандырылмаған деректер алмасу платформасын құруға, сонымен қатар шекті гомоморфты шифрлау негізінде тірек-векторлар әдісі классификаторының қауіпсіз алгоритмін құруға болады.

Түйін сөздер: Блокчейн, гетерогенді деректер, ТВӘ, қауіпсіздік сызбасы.

С.М. Нарбаева*, С.К. Тәпеева, А. Турарбек, С. Жұнусбаева
Казахский национальный университет имени аль-Фараби, Казахстан, г.Алматы
E-mail: narbaevasalta777@gmail.com

Модель машинного обучения, основанная на гетерогенных данных

Большие данные широко используются во многих областях бизнеса. Информация между организациями систематически воспроизводится и обрабатывается данными, а собранные данные существенно различаются по атрибутам. Составляя разнородные наборы данных, они дополняют друг друга, следовательно, необходим обмен данными между организациями.

В процессе совместного обучения машинного обучения на основе разнородных данных текущая схема имеет множество проблем, включая эффективность, безопасность и доступность в реальных ситуациях. В этой статье мы предлагаем безопасный механизм обучения МОВ, основанный на блокчейне консорциума и пороговом гомоморфном алгоритме шифрования. Путем внедрения блокчейна консорциума можно построить децентрализованную платформу обмена данными, а также разработать безопасный алгоритм классификатора опорно-векторные машины на основе порогового гомоморфного шифрования.

Ключевые слова: Блокчейн, гетерогенные данные, МОВ, безопасная схема.

1 Introduction

Intelligent automotive technology is developing very rapidly, and recent advances suggest that autonomous car navigation will be possible in the near future. One of the new trends of protecting data is the blockchain technology that today is used in different areas. In this regard, we believe that absolutely all vehicles will have a full-fledged on-board computer with the ability to install secure applications with access to navigation and other sensors in reading mode. Therefore, the implementation of blockchain solutions will be quite affordable without additional hardware modifications [1].

The development of cloud computing and edge computing has led to a proliferation of data, such as the large amount of data generated everyday in vehicular social networks, which can be used to optimize the security, convenience and entertainment of applications in vehicular social networks [2]. Effective data analysis methods need to be used in such scenarios, among which machine learning and deep learning are particularly important [3]. Among the commonly used machine learning methods, support vector machine (SVM) model has significant advantages in performance and robustness, so it has a wide range of applications [4].

Take the vehicular social network as an example. There are various organizations within transport networks, such as a vehicle manufacturer, a vehicle management agency, and a provider of vehicle social media application services. These entities have different data sources, and differences in the data sources cause the data to complement each other in terms of attributes [5]. We call the scenario data heterogeneous data.

However, for a single organization, its dataset cannot cover the multidimension, which has great limitations in the use process. Especially in the training process of SVM classifier, the classification effect of the final model is highly correlated with the quality of the data set, so it is difficult for a single organization to train an ideal classifier through its own data. Therefore, it is necessary to share heterogeneous data among multiple institutions. Through data sharing, a dataset covering multiple attributes can be combined to improve the effectiveness of the classifier. From another perspective, the dataset obtained after the fusion of these heterogeneous data can be vertically partitioned into sub-data sets provided by each unit according to the attributes. However, in the process of data sharing, data privacy is facing serious challenges. First of all, the heterogeneous data to be shared contains users' privacy information. With the increasing attention of the government and individuals to users' privacy issues, more and more regulations restrict the sharing of users' data by enterprises. As a result, direct data sharing is subject to increasingly stringent regulations. In addition, for the data owners, the high value of heterogeneous data is mainly reflected in the privacy of the data, that is, the data is only owned by itself, or a small number of institutions. So

if the data is shared directly, it becomes less private and less valuable and data owners are unwilling to reduce the value of their data [6].

For a long time, privacy disclosure issues raised in diverse scenarios has been highly concerned [7]. Among those scenarios, many researches pay attention to train a machine learning classifier securely over both horizontally and vertically partitioned datasets. Many existing solutions adopt secure multi-party computation (SMC) to prevent privacy disclosure. Firstly, in those schemes, how to balance security and efficiency issues still faces big challenges [8]. Then, one or more aided servers are essential with the assumption that they are trusted or semi-trusted during the training process. Obviously, in a real-world scenario, it is impractical to provide such aided servers for the participants. To deal with the two challenges of applying the privacy protection scheme to real-world scenarios, we propose an efficient and secure SVM classifier training scheme based on consortium blockchain where no third party is introduced [9].

In this paper, we propose a security SVM training mechanism based on consortium blockchain for multi-source heterogeneous data sharing scenario, which solves the above two problems. First, because the differential privacy protection scheme introduces noise to the training results and the training process is not secure, we adopt the scheme based on homomorphic encryption [10,11].

We introduce block chain to establish a decentralized data sharing platform for sharing secret data. When each participant shares data, they simply upload the data to the data sharing platform. The access control and permission mechanism of the consortium blockchain fully ensures the unknowability of the external data and the openness and transparency of the internal data.

We propose a SVM training scheme that contributes more secure and efficient heterogeneous data sharing. First, an open, reliable and transparent data sharing platform was built based on blockchain technology. The operation of the platform does not rely on trusted third parties. The data on the platform is visible to members in the blockchain and not visible to the outside. After that, most of the training work was completed locally by each participant based on clear text data. We introduce threshold homomorphic encryption scheme to ensure a data privacy protection scheme in a decentralized environment. All data that needs to be shared can be fully protected by this scheme and maintain its homomorphic property. Our scheme guarantees a controllable degree of privacy protection by setting the size of the threshold. A large number of experiments based on real datasets prove the feasibility and efficiency of the scheme.

2 Secure Machine Learning over Heterogeneous Data

Consider a dataset D is combined with several participants who have its own dataset $D^p \in A, B, \dots, N$, where x_i^p represent the i -th instance in D^p , and y_i is shared as a data label between all related i -th instance x_i^X . When training a SVM classifier, we define w as the model parameters, Δ_t as gradient in the t iteration, λ as the learning rate. Meanwhile, we assume that $[[m]]$ as the encryption of message m under Paillier. Table 1 shows the notations used in this paper.

Table 1. Notations

Notations	Description
D^A	The dataset of participant
d^A	The dimension of dataset D^A
x_i^A	The i -th data instance of dataset D
y_i	Number

3 System Model

We divide our system into three components based on their relationship with the data. As shown in Fig. 1, they are data device (DD), data provider (DP) and blockchain service platform (BSP).

- **Data Device:** Refer to devices capable of generating data, including sensors, mobile devices, and so on. Because the data directly collected from these devices contains high-value information, these data are collected, processed, and then used for data analysis.

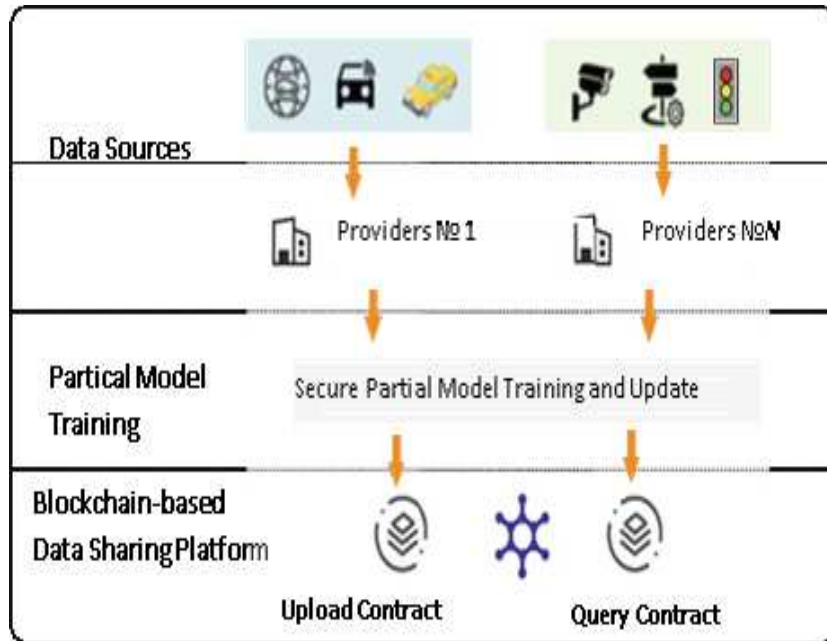


Figure 1: Overview of secure SVM training scheme over heterogeneous data

- **Data Provider:** The equipment that generates the data is collected, stored, and used by different parties. These participants are called participants and act as data providers in our solution. Due to the different equipment, the collected data is different, and due to the different data processing methods, the available data after processing has different

attributes and complement each other. In addition to serving as data providers, these participants also act as model trainers to train machine learning models in collaboration. According to the scheme in this paper, most of the training work is done locally in participant.

- **Blockchain Service Platform:** This is a service platform that runs on the consortium blockchain. On one hand, it provides a transparent data sharing platform distributed in participant, allowing participant to retrieve all the data recorded in the BSP. At the same time, no one captured the data recorded on the BSP for changes. On the other hand, BSP has strong security protections, making data outside of participant invisible to entities. In addition, communication data between the BSP and participant is also encrypted, preventing data leakage.

4 Threat Model

In the scheme, there is only one role of the data provider. We treat participants honest but curious when it comes to the security model, that is, all participants are curious about the data of other participants, but they will execute the scheme according to the rules. In addition, due to the large number of interactions between participant and BSP, potential threats in the interaction process are also considered.

- **Known Ciphertext Model.** BSP is a common and transparent data sharing platform for all participants. The data shared by each participant is visible to other participants. These data include the dense intermediate value and the decrypted calculation results.
- **Known Background Model.** We assume that multiple participants can conspire and collaborate to analyze shared data. Compared with the above threat model, this model can obtain more information. Under the above system model and threat model, we established the following three system design goals to meet the system's requirements for security, accuracy and performance.
- **Data privacy is fully protected.** Under the two threat models, during the entire training process, the privacy of the original data and the shared intermediate value will not be leaked, and the participants cannot infer valuable information from the shared data. Second, the data in the data sharing platform is guaranteed to be invisible to the outside world.
- **High accuracy of training results.** Generally speaking, the introduction of privacy protection schemes may introduce noise into the calculation process and cause inaccurate calculation results. Our design goal is to obtain a classifier that is not significantly different from conventional training conditions.
- **Low training overhead.** Similarly, the introduction of privacy protection schemes will increase training overhead. These overheads are mainly caused by additional computing operations such as encryption and decryption, and additional communication overhead. Therefore, our solution needs to ensure low training overhead while ensuring security

5 Secure SVM Training Scheme over Heterogeneous Datasets

In this section, in order to clearly introduce the work of each participant in the training process, we assume that three participants participate in the SVM model training. The respective training sets are complementary in attributes. As shown in Fig. 3, the entire training process mainly includes three parts: local training, gradient update judgment, and model update. In these three steps, two data sharing and one decryption operation are involved. Finally, after multiple iterations, each participant gets its own partial model and uploads it to the blockchain to form a complete model together.

The data privacy protection method of this solution is based on a threshold homomorphic encryption algorithm. Before training the model, a pair of public and private keys needs to be generated for each participant. The public key is the same and the private key is different. Through the secret sharing scheme combined with the existing threshold key management scheme, such a key pair is negotiated and distributed. In addition, the three participants join the consortium blockchain data sharing platform as nodes, and they need to pass identity authentication before joining. Finally, all participants need to initialize the model parameters and preprocess the data set, including unified labeling and sample order.

6 Local Training Process

In order to ensure the efficient training of the model, this solution puts most of the work locally on three participants. During one iteration, all training work can be done locally before the gradient update judgment. This section will introduce how each participant can be trained locally based on its own heterogeneous data. SVM optimization algorithm based on stochastic gradient descent (SGD) is easy to perform. SVM based on stochastic gradient descent can be expressed in the following form:

$$f(w) = \frac{1}{2}w^T w + C \sum_{i=1}^m \max(0, 1 - y_i w^T x_i) \quad (1)$$

Algorithm 1. SVM based on SGD

Require: Training set D , learning rate λ , maxIters T .

Ensure: Trained model w^* .

- 1: **for** $t = 1$ to T **do**
- 2: Select it from D randomly.
- 3: Update $\Delta t + 1$ by Eq. (1).
- 4: Update $w t + 1$ by Eq. (4).
- 5: **end for**
- 6: return w^* .

The right part of the equation is the hinge-loss function, where C is the misclassification penalty and we take $\frac{1}{m}$ as its value.

At each iteration, we use Eq. (2) to calculate the gradient.

$$\Delta_t = \lambda w_t - I[(w x_i < 1)] x_i y_i \quad (2)$$

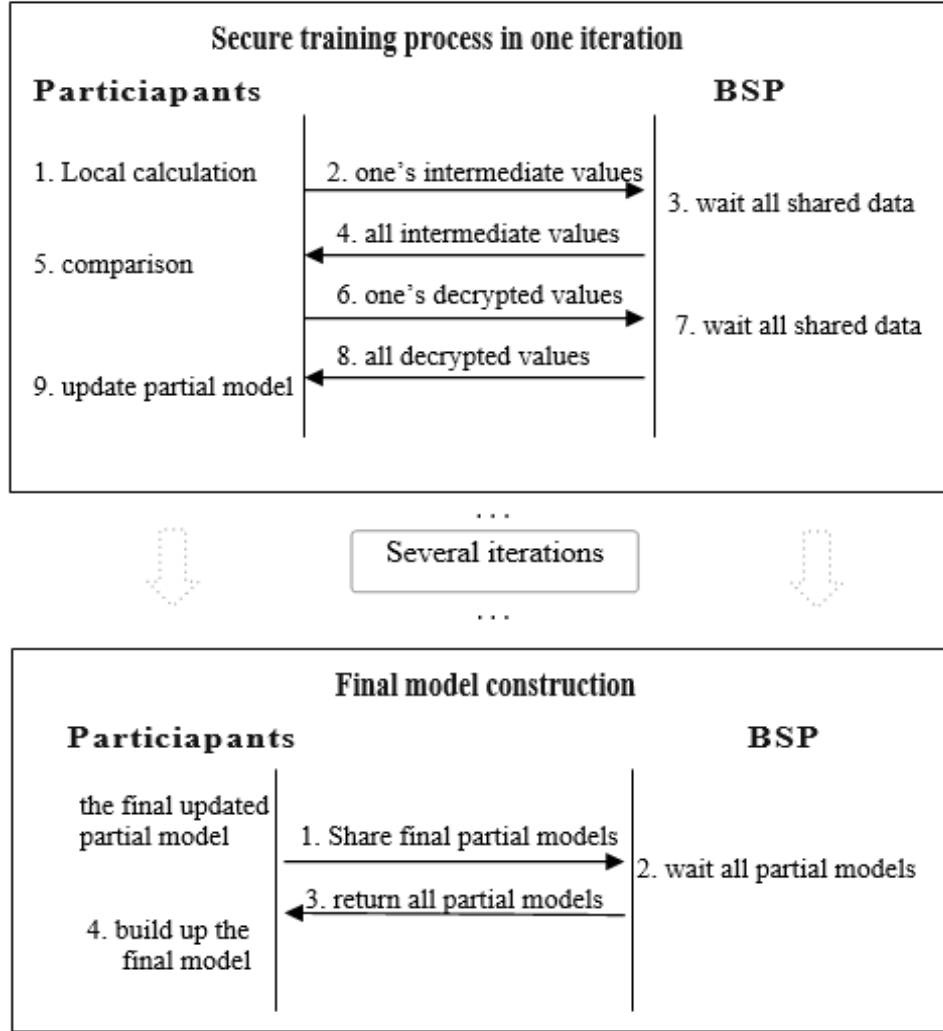


Figure 2: Workflow of secure training over heterogeneous datasets

If $I[(wx_i) < 1]$ is true which means $(wx_i < 1)$, $I[(wx_i) < 1] = 1$; Otherwise, $I[(wx_i) < 1] = 0$. Then we can update the w by Eq. (4).

$$w(t+1) = w_t - \lambda \Delta_t \quad (3)$$

Through one iteration of the training process over several heterogeneous datasets, only when calculating I , data exchange between multiple participants is required. The rest of the training operations are performed locally. We represent wx_i by a in the following sections.

$$I = \begin{cases} 1 & y_1(w^A x_i^A + w^B x_i^B + w^C x_i^C) < 1 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

The complete algorithm is described in Algorithm 2.

The three participants need to share the calculated median value to the BSP during the training model. This solution treats the shared data with a threshold homomorphic encryption

scheme to ensure data security and ensure that the gradient can be calculated correctly. To judge how to update the gradients, here we use additive homomorphic encryption to construct Eqs. (5), (6) and (7)

Algorithm 2. Partial model training process

Require: Training set D^A, D^B, D^C , learning rate λ , maxIters T .

Ensure: Trained model w^* .

1: All participants perform the following operations simultaneously. Take participant A to describe in detail.

2: **for** $t = 1$ to T **do**

3: Select it i_t randomly.

4: Calculate $y = \sum_{i=1}^{-d^A} w x_i$

5: Cooperate with other participants to judge how to update gradient by Eq. (5).

6: Update Δ_{t+1} by Eq. (2).

7: Update w_{t+1} by Eq. (4).

8: **end for**

9: Get several partial model parameters and combine them.

10: **return** w_{t+1} .

$$[[a]] = [[\sum_{i=1}^n a^i]] = \prod_{i=1}^n [[a^i]] \quad (5)$$

$$[[r_2]] = [[\sum_{i=1}^n r_2^i]] = \prod_{i=1}^n [[r_2^i]] \quad (6)$$

$$[[ar_1 + r_2]] = [[ar_1]][[r_2]] = [[\sum_{i=1}^{r_1} a]][[r_2]] = \prod_{i=1}^{r_1} [[a]][[r_2]] = [[a]]^{r_1} [[r_2]] \quad (7)$$

In order to determine the update method of the gradient, the method adopted in this solution is to compare the encrypted calculation result with the constant 1. In Algorithm 3, the security comparison algorithm in three participant scenarios is introduced in detail. It is obvious that for integer a , if $(ar_1 + r_2) > (r_1 + r_3)$, we can derive that $a > 1$, otherwise $a < 1$.

Algorithm 3. Privacy-preserving gradient update judge**Input A:** $[[a^i]]$ from participant i .**Input B:** $r_1^i, [[r_2^i]], r_3^i$ from participant i .**Ensure:** $a > 1$ or $a < 1$.1: Each participant i picks three positive integers r_1^i, r_2^i, r_3^i , where $|r_3^i - r_2^i| < r_1^i$, and encrypts r_2^i to get $[[r_2^i]]$.2: Each participant i uploads $[[ai]], r_1^i, [[r_2^i]], r_3^i$.3: Each participant i downloads all the other participants' $[[ai]], r_1^i, [[r_2^i]], r_3^i$.4: Each participant i calculates $[[a]], [[r2]]$ by Eq. (5) and Eq. (6), and calculates r_1 and r_3 where $r_1 = \sum_{i=1}^n r_1^i$ and $r_3 = \sum_{i=1}^n r_3^i$.5: Each participant i calculates $[[ar1 + r2]]$ by Eq. (7).6: Each participant i decrypts $[[ar1 + r2]]$ by sub-private key SK^i and uploads it to *BSP*.7: Each participant i downloads all other decrypted values from participants to recover $(ar1 + r2)$, and compares $(ar1 + r2)$ with $(r1 + r3)$.8: If $(ar1 + r2) > (r1 + r3)$, $a > 1$; Else $a < 1$.9: return $a > 1$ or $a < 1$.**7 Data Sharing on BSP and Security Analysis**

Participant relies on BSPs to securely calculate intermediate values. BSP simplifies complex point-to-point communication between participants. Participant completes data on-chain and data query by calling smart contracts. During the iteration process, each participant uploads data twice: calculating the intermediate value (IV) and the decrypted value (DV), respectively. These two data are also read twice.

1. The Format of IVs

Iteration Round: When multiple data providers train the model collaboratively, some data needs to be exchanged in each iteration. Therefore, in order to represent the data exchanged in each round and to distinguish it from other rounds of data, a field is required to indicate the training round. Iteration Round is maintained by smart contracts.

DP ID: A field that identifies the owner of the data. When a node calls a contract to upload data, its address will be automatically recorded in this field.

Training Intermediate Value: The intermediate value of the encrypted state during model training. The values provided by each participant will be summed and compared to the magnitude of 1 in the encrypted state.

r1: An unencrypted random positive integer which is used to compare.

r2: An encrypted random positive integer which is used to compare.

r3: An unencrypted random positive integer which is used to compare. Random Positive Integer: It is generated randomly by each participant and its value is between 1 and m , the sum of which determines the data instances selected in the next iteration.

2. The Format of DVs

Iteration Round: Similar function described in IVs. DP ID: Similar function described

in IVs. Decrypted Value: Each participant decrypts the result obtained based on his own private key. By combining all these values, each participant can obtain the final decryption result.

The definition of computing security for a secure multiparty computing protocol is given below.

Definition 1 *The multi-party computation protocol with n participants under the cryptography model is considered to be computation security, if for any attacker A , there exists a corresponding simulator S in the ideal model interacting with A , and satisfying the following conditions:*

(1) *The running time of S is the polynomial of A 's running time.*

(2) *For any input set, the $n+1$ outputs produced by the multi-party computation protocol are computationally indistinguishable from the $n+1$ outputs produced by the ideal model.*

We conducted a security analysis based on the above idea. Thus we acquire the information which an attacker can get from the ideal model and the real protocol. Then we compare them and prove they are indistinguishable. In this scheme, n participants are involved to share their encrypted intermediate values to calculate

$$F : F([a]^1, \dots, [a]^n, 1, r_1^1, [[r_2^1]], r_3^1, \dots, r_1^n, [[r_2^n]], r_3^n).$$

Assume that the attacker has corrupted a set of participants $A = P_i1, \dots, P_i|A|$. Then all the data the attacker obtained in the ideal model is the output F of the participants and the input:

$$([a]^{i1}, \dots, [a]^{i|A|}, 1, r_1^{i1}, [[r_2^{i1}]], r_3^{i1}, \dots, r_1^{i|A|}, [[r_2^{i|A|}]], r_3^{i|A|}).$$

We construct a simulator S that simulates all the data the attacker gets in the real model based on the data the attacker obtained in the ideal model. Firstly, we analyze the information that the attacker can get in the real protocol.

Input Phase. Since all the participants share their encrypted input:

$$([a]^1, \dots, [a]^n, 1, r_1^1, [[r_2^1]], r_3^1, \dots, r_1^n, [[r_2^n]], r_3^n)$$

The attacker is able to get all of them. Especially for the corrupted participants, the attacker also gets $a^{i1}, \dots, a^{i|A|}, r_2^1, \dots, r_2^{|A|}$.

Computation Phase. At each step of the calculation phase, the attacker obtains data $[[x + y]]$ based on $[[x]]$ and $[[y]]$.

Output Phase. In the output phase, the attacker gets the result:

$$F([a]^1, \dots, [a]^n, 1, r_1^1, [[r_2^1]], r_3^1, \dots, r_1^n, [[r_2^n]], r_3^n).$$

Then we construct the simulator S of the polynomial time. S takes $a^{i1}, \dots, a^{i|A|}$ and r_2^1, \dots, r_2^n and F as the input. The following step S_0 simulates the information calculated based on the input.

Step S_0 . S generates the encrypted data $[[a]^{i1}, \dots, [a]^{i|A|}, [[r_2^{i1}]], \dots, [[r_2^{i|A|}]]$ and $[[F]]$ based on $a^{i1}, \dots, a^{i|A|}, r_2^1, \dots, r_2^n$ and F . Then, S can simulate the calculations based on those encrypted data such as $[[a^{i1} + a^{i2}]]$ and $[[r_2^{i1}, \dots, r_2^{i2}]]$.

Step S1. After step $S0$, we can simulate part of the calculated intermediate values which are defined as $[[a]]^{j^1}, \dots, [[a]]^{j^{|r|}}, [[r_2^{j^1}]], \dots, [[r_2^{j^{|r|}}]]$. Then for the remaining intermediate values that cannot be directly simulated by $S0$, S simulates them by selecting the random numbers to generate the corresponding ciphertext. According to the threshold cryptosystem's security, these simulations are successful.

Step S2 Based on steps $S0$ and $S1$, we can simulate all the values calculated in the computation phase.

Step S3. F is one of S 's input, so S can easily get a simulation of F . From the above simulation process, the information obtained by the attacker from the ideal model and the information obtained from the real model are computationally indistinguishable. You can prove the security of the solution.

8 Conclusion

In this section, we propose an effective and secure SVM training scheme that helps multiple data providers train SVM classifiers on vertically partitioned datasets. The target of this chapter is to combine consortium blockchain technology and threshold Paillier to create a decentralized and secure SVM training platform. To achieve high performance, most training operations are performed locally on raw data, so there are only a few intermediate values that need to be shared across platforms.

References

- [1] Narbayeva S., Bakibayev T., Abeshev K., Makarova I., Shubenkova K., Pashkevich A., "Blockchain Technology on the Way of Autonomous Vehicles Development", *Transportation Research Procedia* 44 (2020): 168-175.
- [2] Cheng N., Lyu F., Chen J., Xu W., Zhou H., Zhang S., Shen X.S., "Big data driven vehicular networks", *IEEE Netw.* 32(6) (2018): 160-167.
- [3] Fadlullah Z.M., Tang F., Mao B., Kato N., Akashi O., Inoue T., Mizutani K., "State-of-the-art deep learning: evolving machine intelligence toward tomorrow's intelligent network traffic control systems", *IEEE Commun. Surv. Tutorials* 19(4) (2017): 2432-2455.
- [4] Lv L., Zhang Y., Li Y., Xu K., Wang D., Wang W., Li M., Cao X., Liang Q., "Communication-aware container placement and reassignment in large-scale Internet data centers", *IEEE J. Sel. Areas Commun.* 37(3) (2019): 540-555.
- [5] Song D.X., Wagner D., Perrig A., "Practical techniques for searches on encrypted data", *IEEE S&P* (2000): 44-55.
- [6] Li H., Zhu L., Shen M., Gao F., Tao X., Liu S., "Blockchain-based data preservation system for medical data", *J. Med. Syst.* 42(8) (2018): 141.
- [7] Shen M., Ma B., Zhu L., Du X., Xu K., "Secure phrase search for intelligent processing of encrypted data in cloud-based IoT", *IEEE Internet Things J.* 6(2), (2019): 1998-2008.
- [8] Xu K., Yue H., Guo L., Guo Y., Fang Y., "Privacy-preserving machine learning algorithms for big data systems", *IEEE 35th International Conference on Distributed Computing Systems (IEEE, Piscataway)* (2015): 318-327.
- [9] Mohassel P., Rindal P., "ABY 3: a mixed protocol framework for machine learning", *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, ACM, New York* (2018): 35-52.
- [10] Katz J., Lindell Y., ", *Introduction to Modern Cryptography* (Chapman and Hall/CRC, Boca Raton, 2014)
- [11] Mangasarian O.L., Wolberg W.H., ", *Cancer diagnosis via linear programming* Technical report (University of Wisconsin-Madison Department of Computer Sciences, 1990).