Comput Syst Sci & Eng (2020) 2: 113–121


International Journal of
Computer Systems
Science & Engineering

# Research on E-Commerce Transaction Payment System Based on C4.5 Decision Tree Data Mining Algorithm

**Bing Xu**[1]*, **Darong Huang**[2]† **and Bo Mi**[3]‡

[1]*School of Economics and Management, Chongqing Jiaotong University, Chongqing, China*
[2,3]*College of Information Science and Engineering, Chongqing Jiaotong University, Chongqing, P.R. China*

In this paper, according to the information classification algorithm in data mining, data in the network payment system of e-commerce is mined, forming an effective evaluation of the security of the network payment system. Firstly, the method of network security risk prediction is discussed. Secondly, according to the characteristics of network payment system, the system security index system is analyzed in detail, and the specific application process of the C4.5 Classification Algorithm in security evaluation is discussed. Finally, the data mining process is designed in detail and the corresponding code established. In this paper, data mining theory is applied to the network payment security evaluation system, and an algorithm system is constructed to evaluate the network payment security. The algorithm system realizes the effective evaluation and judgment of the network payment system security as well as warning of potential network security problems, effectively changing the previous way of network security management, and ensures the security and stability of the network payment system is maximized.

Keywords: E-Commerce Transactions; C4.5 Decision Tree; Data Mining; Payment System

## 1. INTRODUCTION

Security evaluation plays a very important role in network system security and is an important addition to a firewall and other security measures. Security evaluation can complete the protection of a network system without affecting the performance index of the network system.

There are four sections in this paper. Firstly, the methods of the current e-commerce payment security evaluation and its disadvantages are analyzed. Secondly, the standards of payment system security evaluation in e-commerce are discussed, and the algorithm of decision analysis in the standard is analyzed. Finally, the algorithm to build a data mining model is used, and the construction of the system model is completed. Overall, when regarding security assessment, some researchers have in-depth research results and a large number of practical applications. At present, some researchers have gradually applied data mining, immune algorithm, agent technology and other technologies to the field of security assessment, among which the combination of data mining inthe field of security assessment has produced more effective research results [1–3].

Wenke Lee research group of Columbia University first applied data mining to the security evaluation system, and achieved important research results through a large number of experiments. The research group mainly tried to apply the classification algorithm, sequence mining algorithm and

*xbing66@aliyun.com
†darong.huang@163.com
‡bo.mi@163.com

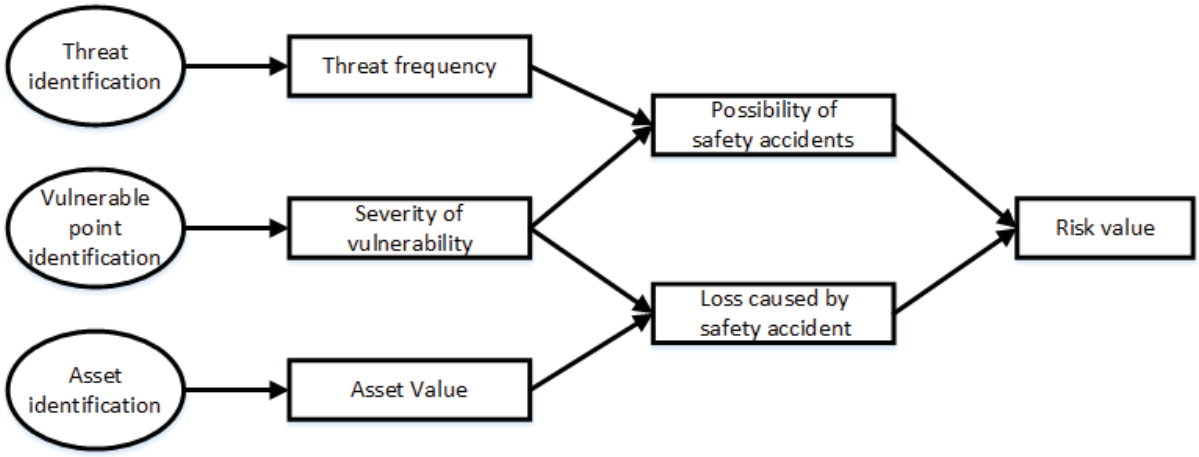


**Figure 1** Payment Network Risk Security Assessment Schematic.

association data mining algorithm of data mining to the security evaluation information system, through the data source based on the network and host to carry out data mining detection. At the same time, the data mining laboratory of Columbia University actively tried a clustering algorithm to detect intrusion anomalies, and achieved positive research results. The forest research group of the University of New Mexico studies the application of a short sequence matching algorithm in the process of security evaluation, and realizes anomaly detection based on the short sequence matching algorithm [4]. The research on the application of CIDF is very extensive, W. Lee uses the method of data mining to process audit data, then proposes and implements a multi-level security evaluation system based on CIDF, this improves the overall accuracy and scalability of the security evaluation system [5–9].

Li Guangxia takes the research of security evaluation technology based on the data mining method as the core concept [10], discusses how to apply the clustering algorithm in the data mining method in a security evaluation, and then attempts and proposes a nearest neighbor priority algorithm based on the idea of "similar" and the shortest distance algorithm. The algorithm finds the nearest neighbor in the s-neighborhood of each point, adopts the principle of block reading (nearest neighbor search algorithm SNN), and uses the data in the KDD cup 1999 data set to test the efficiency of the algorithm. Han Zongfen, Liu Ke, Jinhai and Guo Li used data mining technology for collaborative security assessment, and proposed a collaborative intrusion rule generation algorithm based on data mining [11–15]. Using distributed collaborative security assessment technology based on data mining, they can effectively detect collaborative intrusion, and also have the detection ability for other unknown cooperative attack modes. Li Qinghua, Tong Jianhua, Meng Zhonglou, Zhang Wei, by analyzing the role of data mining technology in the intrusion feature search, put forward a feature mining model applied in the mixed mode security evaluation system based on network and host [16–22]. Liang Tiezhu and Li Jiancheng (April 2002) analyzed how the artificial intelligence method was applied to the security evaluation system, discussed the main related technical problems [23], and illustrated the effect of the clustering method on enhancing the classification performance of network connection through this case.

This paper primarily uses data mining technology to realize the network payment security evaluation management system. Based on the actual needs of network payment security management, the data mining algorithm is used to accurately, quickly and dynamically judge the security situation of the network payment system, and to grasp the attacks on the network payment system. Information is categorised to avoid potential network payment errors caused by network security problems, causing losses to enterprises and consumers. Therefore, the research in this paper has a strong practical application value. In addition, there are only a few research papers on the management system for the evaluation of network payment system security. Since the network payment platform security evaluation system differs significantly from other products, it is difficult to compare directly to other products.

The security assessment management system is applied to the security management process of the network payment system. The research of this paper is based on the actual needs of network payment companies. According to the standards and requirements of network payment system security assessments, the C4.5 decision tree algorithm is applied to the network payment system security assessment management process to ensure the security of the network payment system. The research in this paper has a strong practical application background and a large practical value.

## 2. E-COMMERCE PAYMENT SECURITY EVALUATION MODEL

### 2.1 E-commerce Payment Security Evaluation System

The network security risk assessment is to identify, evaluate and comprehensively analyze the risks existing in the payment network environment. It mainly involves the basic elements of users' assets, threats and vulnerabilities.

Considering the aspect of threat identification,three typical attack types are selected, DoS attacks, ARP attacks and replay attacks, as the main indicators to judge the attack situation received by the payment gateway when conducting an

**Table 1** Security Assessment of Online Payment System.

| Project | Grade | Standard | |
|---|---|---|---|
| DoSattack | High risk | Attack frequency x $\geqslant$ 100/s | High risk=A Medium risk=B Low risk=C |
| | Medium risk | Attack frequency 100>x $\geq$ 10/s | |
| | Low risk | Attack frequency x < 10/s | |
| ARPattack | High risk | Attack frequency x $\geq$ 300/s | |
| | Medium risk | Attack frequency 300>x $\geq$ 100/s | |
| | Low risk | Attack frequency x < 100/s | |
| Replay attack | High risk | Attack frequency x $\geq$ 50/s | |
| | Medium risk | Attack frequency 50>x $\geq$ 10/s | |
| | Low risk | Attack frequency x < 10 / s | |
| Error severity | High risk | Judge according to the system log judgment criterion of Server 2003 | |
| | Medium risk | | |
| | Low risk | | |
| Type of safety accident | High risk | Judge according to the system log judgment criterion of Server 2003 | |
| | Medium risk | | |
| | Low risk | | |
| Gateway error time | High risk | Repair in 60 minutes | |
| | Medium risk | Repair in 30 minutes | |
| | Low risk | Repair in 5 minutes | |
| Routing attack | High risk | Attack frequency x $\geq$ 100/s | |
| | Medium risk | Attack frequency 100>x $\geq$ 10/s | |
| | Low risk | Attack frequency x < 10/s | |
| Server supply | High risk | Attack frequency x $\geq$ 100/s | |
| | Medium risk | Attack frequency 100>x $\geq$ 10/s | |
| | Low risk | Attack frequency x < 10/s | |
| Gateway attack | High risk | Attack frequency x $\geq$ 100/s | |
| | Medium risk | Attack frequency 100>x $\geq$ 10/s | |
| | Low risk | Attack frequency x < 10/s | |
| Host attack | High risk | Attack frequency x $\geq$ 100/s | |
| | Medium risk | Attack frequency 100>x $\geq$ 10/s | |

e-commerce payment. In terms of vulnerability, this paper selects the indicators of payment system error severity, security event type, and system error time to judge the vulnerability of payment system. In terms of user asset identification, this paper classifies network threats into attack types such as router attacks, server attacks, gateway attacks, and host feeds according to the devices in the transaction payment system. By judging these attacks, the security of payment system can be determined and described.

## 2.2 Quantitative Standard of E-Commerce Network Payment Security Evaluation System

After determining the specific electronic payment security evaluation system, the indicators need to be furthered quantified. According to the international common standards, a specific quantitative indicator system has been designed as shown in Table 1.

According to the categories in Table 1, when the project meets any of the risk standards in Table 1, it means that the payment system is in a low risk state, which can be represented as case C. When the payment system is in a medium risk state, it means that the payment system is currently being attacked or is vulnerable, and the key problems of the gateway should be checked as soon as possible, which can be represented by case B. When the payment system is in a high-risk state, it means that although the current payment business may not be affected, according to the current attack data records the system has shown a high risk trend, and the system should be actively upgraded and checked to avoid a large system security risk in the future and affect the payment business. High-risk situations are represented by case A.

# 3. C4.5 DECISION TREE ALGORITHM

## 3.1 The Concept of the Decision Tree Algorithm

The decision tree algorithm is one of the broad-based inductive reasoning algorithms. It can handle the classification and prediction problems of categorical or continuous variables. It can express the model with graphs and if-then rules, and has a high readability. By continuously dividing information, the decision tree model has the largest difference when using dependent variables. The ultimate goal is to classify information into different organizations or branches, and establish the strongest classification on the value of dependent variables.

Decision trees are a supervised learning method that produces a tree structure similar to a flow chart. The method for processing the information of the decision tree algorithm is:establish classification decisions and standards, and then get predictions and analysis of new information. The leaf node is used as the last node of the decision tree to display the classification of the classification results. The internal nodes correspond to the variable tests. The node branches represent the test output and the variable data as variables. After classifying the data, the variable values are tested in the branch, and the paths between all nodes of the decision tree represent a class of classification rules.

In the field of data mining, the decision tree algorithm is popular due to the following three characteristics:

(1) The decision tree model representation method is extremely simple and easy to grasp and use. Usually, the decision tree model is represented by rules and graphs.

(2) The decision tree model can handle two types of variables: one is a categorical variable; the other is a continuous variable. When making split variable selection, we usually use the variable with the maximum information gain. The decision tree model represents the importance of variables. The importance here is relative.

(3) The decision tree model can solve problems with a large information set. As there is no relationship between the decision tree size and the database, each calculation is quite small. The decision tree can hold many variables, that is to say, no matter how many variables; the decision tree can be built.

## 3.2 Principle of C 4.5 Decision Tree Algorithm

Using the C4.5 decision tree algorithm, information gain is widely used. Initially, the information gain attributes are checked and the largest attribute found in order to build anew node, and so on, the branch category is determined by the information gain. Through the construction of layers of nodes, a complete decision tree is finally formed. There are many paths on the decision tree, and the classification rules are different for different paths. The completed decision tree represents a classification model. Compared with the traditional classification model, a decision tree has great advantages, including intuitive graphics and a simple operation method, which makes it easy to understand and master for users.

On the nodes above the decision tree, the information gain is used to determine the test variables. The information gain is used to classify the training samples, reflect the ability of the training samples, select the variable with the highest information gain, and use it as the segmentation reason of the nodes. According to different data, information gain will minimize the amount of data in order to reflect the instability and inaccuracy of data.

If the probability of occurrence of a certain type of event is $P$, let the number of information after the occurrence of the event be $I(P)$. If $P = 1$, then $I(P) = 0$, since any event will appear, the event appears. No information will be supplied. Alternatively, if the probability of occurrence of a certain type of event is very small and there is a great instability, then the amount of information generated after the occurrence of this event will be large, so $I(P)$ is in a decreasing state and develops as a class of decreasing function. The information set is represented by $S$, the category variable is represented by $E$, and the number of categorical variables is represented by m. The information set is divided into m subsets by the variable $E$, representing samples of values in the information set. The chances of the m kinds of probabilities corresponding to them appear, so the information amount of the i-th result is $I(P)$. The average information of the classified samples is represented by entropy;a measurement standard used to measure uncertain variables and can also measure the purity of the information set. The functional formula of entropy is shown in Equation 1.

$$I(S_1, S_2, \cdots, S_m) = \sum_{k=1}^{n} \rho_i (log_2(\rho_i)) \tag{1}$$

The variable classification is used to train the information set ability, and the variable detection can be performed by increasing the amount of data. After using the algorithm on the data, a node is finally formed, this factor is marked, and the data of all kinds of samples are obtained.

# 4. THE ESTABLISHMENT OF DATA MINING MODEL

## 4.1 Establishment of Decision Tree Model

Before the establishment of the decision tree model, the training samples are selected and confirmed, the trained samples can then participate in the creation of the decision tree and shape the relevant attributes of the decision tree. The relevant information of the training sample data set is shown in Table 2:

The training sample data is filtered first;any missing or incomplete data is deleted to ensure the integrity of the sample characteristics of the retained sample data. The index safety registration is marked according to the above index quantification standard.

The above data, as sample data,after processing provides sample training for the whole data mining system. The sample data can effectively train the sensitive value of the decision tree model. It is a foundation for the effective prediction of the future level of network payment security assessment. The results after generalization are shown in Table 3.

From the above analysis, it is shown that the structure of the decision tree is in fact tree shaped, and each node has an optional attribute to complete the segmentation. All the branches are a small part of the segmentation; therefore leaf nodes can be used to represent a distribution. The elastic algorithm is often used in order to get the decision tree, using the method of processing from top to bottom. At first, all information data exists on the root node, and its attributes are category fields. All records are then divided and transferred by attributes. Here, the method of selecting an attribute is a certain measure. If the information classification on a node is completely consistent, segmentation will not be implemented and a leaf node will be formed. Attribute selection is therefore the basis of decision tree construction. If all the node attributes are selected, a complete decision tree will be created.

**Table 2** Sample Data Set.

| Time Interval | DoS Attack | ARP Attack | Replay Attack | …… | Host Attack | Classification |
|---|---|---|---|---|---|---|
| 1 | 120 | 89 | 24 | …… | 52 | A |
| 2 | 180 | 92 | 22 | …… | 21 | A |
| 3 | 5 | 24 | 32 | …… | 9 | C |
| 4 | 6 | 72 | 12 | …… | 9 | C |
| 5 | 24 | 110 | 24 | …… | 8 | B |
| 6 | 26 | 162 | 41 | …… | 10 | B |
| 7 | 31 | 142 | 12 | …… | 11 | B |
| 8 | 25 | 310 | 75 | …… | 10 | A |
| …… | …… | …… | …… | …… | …… | …… |
| 95 | 150 | 240 | 65 | …… | 4 | A |
| 96 | 15 | 150 | 42 | …… | 9 | B |
| 97 | 18 | 224 | 22 | …… | 9 | B |
| 98 | 21 | 192 | 42 | …… | 12 | B |
| 99 | 32 | 99 | 37 | …… | 19 | C |
| 100 | 64 | 143 | 42 | …… | 42 | A |

**Table 3** Sample Data Set After Generalization.

| Time Interval | DoS Attack | ARP Attack | Replay Attack | …… | Host Attack | Classification |
|---|---|---|---|---|---|---|
| 1 | A | C | B | …… | B | A |
| 2 | A | C | B | …… | B | A |
| 3 | C | C | B | …… | C | C |
| 4 | C | C | B | …… | C | C |
| 5 | B | B | B | …… | C | B |
| 6 | B | B | B | …… | B | B |
| 7 | B | B | B | …… | B | B |
| 8 | B | A | A | …… | B | A |
| …… | …… | …… | …… | …… | …… | …… |
| 95 | A | B | A | …… | C | A |
| 96 | B | B | B | …… | C | B |
| 97 | B | B | B | …… | C | B |
| 98 | B | B | B | …… | B | B |
| 99 | C | C | B | …… | B | C |
| 100 | B | B | B | …… | B | A |

In this paper, the C4.5 algorithm is selected for data set processing. Different labels have different attribute values. One label corresponds to two attribute values respectively. Therefore, two classifications are realized. $C_1$ denotes a high risk class, where there are nine data samples; there are five samples in the low risk class denoted by $C_2$. According to data mining technology, the information and data of classification requirements are obtained. The process is as follows:

$$I(9,5) = -9/14\log_2 9/14 - 5/14\log_2 5/14 = 0.94$$

Next, from the DoS attack values, the entropy of each attribute is calculated.

For Low riskDoS attacks:

$$s_{11} = 2, s_{21} = 3, I(s_{11}, s_{21}) = -2/5\log_2 2/5 - 3/5\log_2 3/5 = 0.971$$

For Medium risk DoS attacks:

$$s_{12} = 2, s_{22} = 3, I(s_{12}, s_{22}) = -4/4\log_2 4 = 0$$

For High risk DoS attacks:

$$s_{13} = 2, s_{23} = 3, I(s_{13}, s_{23}) = -3/5\log_2 3/5 - 2/5\log_2 3/5 = 0.971$$

According to the data mining formula, the $E$ and $Gain$ values are calculated.

$$E = 5/14I(s_{11}, s_{21}) + 4/14I(s_{12}, 2_{22}) + 5/14$$
$$I(s_{13}, 2_{23}) = 0.694$$
$$Gain = I(s_1, s_2) - E = 0.246$$

The same method is used to calculate the gain value of the other indicators. After comparing the indicators, the greatest value is gain (DoS attack) therefore a node is established to complete the test attribute DoS attack. At the same time, a separate branch is introduced after each value, and the sample data is divided according to this method. The branch, “moderate risk DoS attack” is the same class that will face other risk

types, so nodes can be established. Usually, 'yes' is used to express it. Using this method, nodes are created of related branch samples, and leaf nodes are established. The related workflow is shown in Figure 2 below:

After the decision tree is generated, the branches of the decision tree need to be pruned. Before the decision tree is created, there are outliers in the selected data, which leads to abnormal conditions in each branch structure. Therefore, it is necessary to reduce the decision tree, which is called pruning. In the process of pruning, the same metric is chosen to prune the abnormal branches. Therefore, the decision tree classification will be established to ensure the accuracy of the model.

## 4.2 Research and Analysis of Data Mining Model

In this paper, data mining technology is used to build the relevant model, when the sample exists, the decision tree is created. The payment security status of an electronic payment network environment is effectively evaluated. At the same time, the establishment of a data mining model can provide the necessary decision support and prediction function for enterprise decision makers.

In addition, after the classification model is established, the correctness of the model should be evaluated. The evaluation object is the test data, which comes from the database and has no relationship with the sample data. For example, in the sample data, the risk factors and probability are the same. The accuracy of the data mining model built in this paper can be proved to meet the requirements, and can be used for the security assessment management of a payment system. Furthermore, it can be judged whether the correct rate of the model meets the requirements. The network payment system can use the data mining model to classify the categories, and provide relevant security measures according to different data with different security levels. In the process of the experiment, the data classification model established in the paper is selected. In order to ensure the feasibility and accuracy of the experiment, the K-fold cross-check test method is selected. The K-fold cross-check method essentially refers to dividing the experimental sample into K parts, each one being the same, and then performing an accuracy test on each group. The accuracy obtained by the test is compared with all the samples, the comparison data is found, and the accuracy rate under the model is defined. The detailed detection process is shown in Table 4 below:

The test results of the data mining algorithm constructed in this paper show an obvious rise. When the number of samples is 100, due to the small number of test samples, the data mining algorithm cannot get effective data for training so its correct recognition rate is only 80%. This is lower than a manual safety assessment. When the number of samples is 600, the correct recognition rate of the data mining algorithm is 92%, exceeding the manual inspection recognition rate. When the number of training samples is 800, the accuracy of data mining algorithm is 95%. It can be seen that the algorithm constructed in this paper has obviously exceeded the efficiency of artificial security assessment.

The use of decision tree model to build a network payment security rating theory model has been successfully completed.

# 5. THE DESIGN OF NETWORK PAYMENT SYSTEM SECURITY EVALUATION MODULE

The network payment system security evaluation module is the core module of the system. Through this module, the measured network security evaluation data can be effectively divided. Further judgments on the security evaluation of the network payment system can be made. In the process of using the module, the critical probability value to determine the security evaluation classification of the network payment system can be found. The system automatically judges the security evaluation of the network payment system. Based on the data mining,using the C4.5 decision tree algorithm, the future network payment system security evaluation is predicted and judged, and the current network payment business security situation is clearly displayed to enterprises. It plays an important decision-making and auxiliary function for enterprises to reasonably grasp the security status of the network payment system and formulate targeted security strategies.

The specific implementation code is as follows:

```
NODE* C4. 5( MATRIX *matrix, NODE* parent, UINT target, UINT state)
/* Routine to build a decision tree, based on Quinlan's C4. 5algorithm.
*/
{
    NEGENTROPY negentropy_struct:
    NODE *node;
    UINT n_vars = matrix->bwidth, n_sanples = matrix->height, i,j, split;
    REAL **data = matrix->data;
    REAL best_threshold, min_negentropy,_negentropy;
    /* A1locate memory for this node */
    node = (NODE*) malloc (sizeof (NODE));
    if (!node)
      err_ exit (_ _FILE__, _ _LINE__) ;
      /* Set up links in decision tree */
      node->parent = parent; /* Set address of parent node */
      if (parent != NULL) /* parent to child: not relevant for root node/*
      {
      /*Pass address of this node to the parent node */
      if (state == ON)
parent->on = node;
      else
        if (state 8# 0FF)
          parent->off = node;
      }
       min_negentropy = 1.0;
       for (i=0; i<n, vars: i++)
       {
       for (j=0; j<n_ samples: j++)
        if (i != target)
```

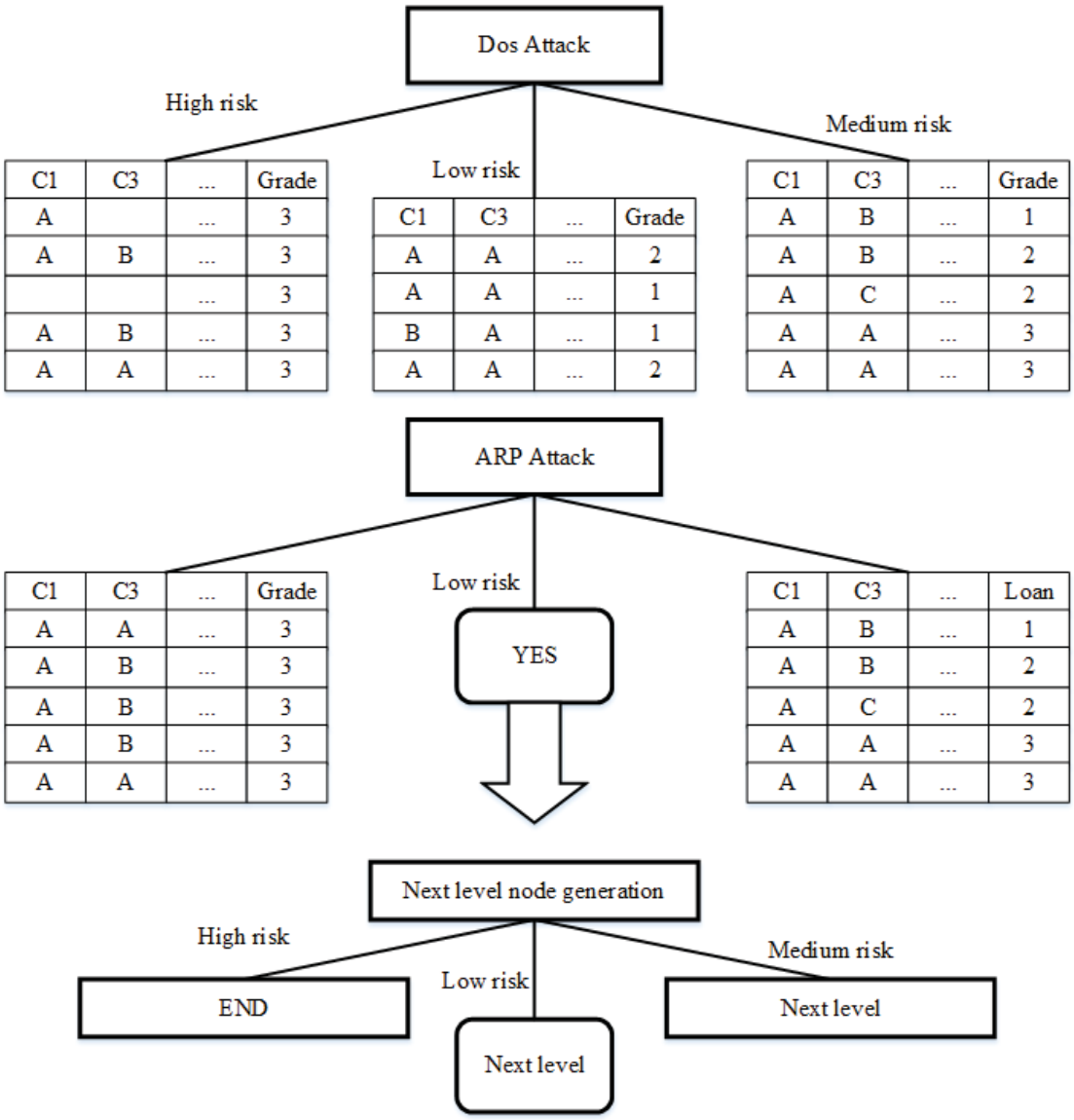


**Figure 2** Schematic Diagram of Decision Tree Generation Process.

**Table 4** Comparison of Detection Accuracy of K-fold Cross-check Test Method in Dangerous Pass Data Mining.

| Sample Size | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 |
|---|---|---|---|---|---|---|---|---|
| K | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| Number of Samples Per Segment | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
| Overall Accuracy | 80% | 85% | 89% | 91% | 91% | 92% | 94% | 95% |

```
        {
          /* Set trial values for this node…. */
          node->idx = i;
          node->threshold = data[j][i];
          /* …and calculate the negentropy of this partition */
          Negentropy_struct = negentropy (data, n_samples,
node,target)
          _negentropy = negentropy_struct. ne;
          /* If this negentropy is lower than any other, retain
the
            index and threshold for future use*/
          if (_negentropy < min_negentropy)
           {
            min_negentropy = _negentropy;
            split= i;
```

```
            best_threshold = data[j][i];
            }
          }/*if (i != target)*/
          }/*for (j=0; j<n_samples; j++)*/
          }/*for (i=0; i<n_ vars; i++)*/
        /* Save the combination of best attribute and threshold
value */
          node->idx = split;
          node->threshold = best_ threshold;
        if (negentropy_struct. status != INACTIVE)
          {
            node->on = node->off= NUL;
            node->idx = negentropy_struct. status;
          }
```

```
        else
      {
        node->on = C4.5(matrix,node, target, ON) ;
        node->off =C4.5(matrix,node,target, OFF) ;
      }
      return node;
}
```

## 6. CONCLUSIONS

In this paper, the C4.5 decision tree algorithm in data mining is used to subdivide the network payment security evaluation system. In the research process of the system, this paper first makes an in-depth study and analysis on the criteria of the security evaluation of the network payment system, then selects the C4.5 decision tree algorithm to calculate the system security evaluation index system, and transforms the system security evaluation system into a computer calculation model. Then the computer algorithm model is applied to the subsequent system design and development process. The data flow of system security evaluation is designed by the data mining algorithm. Through the design of the system, the practical judgment of the network payment security based on the network payment security detection data is realized. It effectively divides the network payment security into low risk, medium risk and high risk to achieve the goal of fine management of network payment system security. This plays a reference role and value for the network payment enterprises to reasonably determine the security of the payment network and to formulate targeted security control measures for different security early warnings.

## ACKNOWLEDGEMENTS

The paper is supported by the Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJQN201900739), Dynamic Option Pricing Mechanism Study on Highway Revenue Rights from Heterogeneity Perspective.

## REFERENCES

1. Wenke, Lee, Stolfo, S.J., Chan, P.K., Eskin, E., Wei Fan, Miller, M., Hershkop, S., Junxin Zhang. Real Time Data Mining-based Intrusion Detection[P]. DARPA Information Survivability Conference & Exposition II, 2001. DISCEX '01. Proceedings, 2001.
2. Wenke, Lee, Salvatore, J., Stolfo, Kui, W. Mok. Adaptive Intrusion Detection: A Data Mining Approach[J]. Artificial Intelligence Review, 2000,14(6).
3. Wenke, Lee, Stolfo, S.J., Mok, K.W. A Data Mining Framework for Building Intrusion Detection Models[P]. Security and Privacy, 1999. Proceedings of the 1999 IEEE Symposium on,1999.
4. Conway, K.D., Fitzpatrick, J.M. The Customer Relationship Revolution—A Methodology for Creating Golden Customers[EB/OL], http://www.eloyaltyco.com.
5. Subba, B., Biswas, S., Karmakar, S. A Game Theory Based Multi Layered Intrusion Detection Framework for Wireless Sensor Networks[J]. International Journal of Wireless Information Networks, 2018.
6. José Francisco, Colom, Gil, D., Mora, H., et al. Scheduling Framework for Distributed Intrusion Detection Systems Over Heterogeneous Network Architectures[J]. Journal of Network and Computer Applications, 2018, 108.
7. Arshad, J., Azad, M.A., Mahmoud Abdellatif, M., et al. COLIDE: A Collaborative Intrusion Detection Framework for Internet of Things[J]. IET Networks, 2019, 8(1):3–14.
8. Zhang, H., Yu, X., Ren, P., et al. Deep Adversarial Learning in Intrusion Detection: A Data Augmentation Enhanced Framework[J]. 2019.
9. Dawoud, A., Shahristani, S., Raun, C. [IEEE 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA) - Krakow, Poland (2018.5.16–2018.5.18)] 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA) - A Deep Learning Framework to Enhance Software Defined Networks Security[J]. 2018:709–714.
10. Li, Guangxia, Cui, Zhe. Application of Data Mining in Performance Pay Management of Public Institutions [J]. Journal of Shijiazhuang Vocational and Technical College, 2014, 26 (04): 11–13
11. Li, Guangxia, Zhang,Siliang, Cui, Zhe. Research on Association Rule Discovery Method [J]. Software guide, 2014,13 (04): 14–16
12. Li, Guangxia, Zhu, Feng, Zhang, Siliang, Cui, Zhe. Research on Multi Decision Tree Fusion Based on Genetic Algorithm [J]. Coal Technology, 2011, 30 (12): 130–132
13. Fu, Yue, Ma, Guofu, Xu, Jiwei, Li, Guangxia. Research on MDTF of Multi Decision Tree Fusion Model [J]. Computer Engineering and Design, 2008 (13): 3391–3393
14. Li, Guangxia, Cui, Zhe, Wang, Zhanfeng. Research on Randomized Processing Methods in Privacy Protection [J]. Fujian Computer, 2008 (07): 102 + 120
15. Liu, Cuijuan, Li, Yuanyuan, Li, Guangxia. Evaluation Model of Data Mining Results Based on Software Measurement [J]. Shanxi Electronic Technology, 2008 (01): 76–78
16. Li, Guangxia, Zhu, Feng, Wang, Zhanfeng, Cui, Zhe. Research and Application of Intrusion Detection Based on Data Mining [J]. Journal of Shijiazhuang Vocational and Technical College, 2007 (06): 33–36
17. Han, Zongfen, Liu, Ke, Jin, Hai, Guo, Li. Distributed Cooperative Intrusion Detection Based on Data Mining [J]. Journal of Huazhong University of Science and Technology (NATURAL SCIENCE EDITION), 2002 (07): 33–35
18. Hong, Liping, Li, Qinghua, Yang, Yanmei, Guan, Qingjuan. Analysis of Library Personalized Service Under Data Mining Technology [J]. Fujian Computer, 2017, 33 (04): 68 +72
19. Zhao, Feng, Li, Qinghua, Zhao, Yanbin. A Sequential Pattern Mining Algorithm Based on Bayesian Method [J]. Computer Engineering, 2006 (14): 17–19
20. Xiong, Jiajun, Zhang, Li, Li, Qinghua. Coding and Detection of Data Mining Patterns in Intrusion Detection [J]. Computer Application and Software, 2005 (11): 13–15 +97
21. Ruan, Youlin, Li, Qinghua, Liu, Gan. Fast Mining and Updating Algorithm of Maximum Frequent Pattern [J]. Computer Engineering and Application, 2005 (24): 23–26 +143
22. Li, Qinghua, Zhao, Yanxi, Jiang, Shengyi. Protocol Analysis and Detection Model Based on Data Mining [J]. Computer Engineering and Design, 2005 (07): 1701–1703 +1826
23. Liang, Tiezhu, Li, Deyi, Song, Yunxian. Sequential Pattern Incremental Mining Method Based on Project Location Index [J]. Journal of Xi'an Jiaotong University, 2002 (10): 1032–1036

**Bing Xu**, M.B.A., Lecturer of School of Economics and Management, Chongqing Jiaotong University, mainly engaged in teaching and research on international economy and trade.

Education: Bachelor degree in English Language and Literature, Huazhong Normal University; M.B.A., graduated from Auckland Institute of Studies, New Zealand.

**Darong Huang**, Ph.D., Professor, College of Information Science and Engineering, Chongqing Jiaotong University, Chongqing, P.R.China.

**Bo Mi**, Ph.D., Associate Professor, College of Information Science and Engineering, Chongqing Jiaotong University, Chongqing, P.R.China.