Research on feature selection algorithm of unknown wireless protocol based on mutual information

Zhiguo Liu¹*, Qiaoyu Yao¹, Meilin Guo² and Xue Li Zha¹

¹Communication and Network Laboratory, Dalian University, Dalian, Liaoning Province, 116622, China
²Australian national university advanced computing (honors), Shandong University (Weihai), Weihai, Shandong Province, 264209, China
*Zhiguo Liu’s e-mail: liuzhiguo@dlu.edu.cn

Abstract. Aiming at the problem of unknown wireless protocol identification methods that focus on the application layer, but the wireless network application layer data is difficult to capture, and the existing method research stays at the signal level and the identification effect is poor, this paper proposes an identification method for feature selection of wireless protocol data from the data link layer. This paper extracts the protocol features of the data frame set by improving the AC algorithm and the association rule algorithm, using mutual information theory, and combining the definition of information correlation and redundancy with the normalization idea, further screening and verification of frame features are carried out to achieve effective extraction of protocol features, effectively improve the accuracy of feature sets, and thereby improve the efficiency and accuracy of protocol recognition.

1. Introduction
With the rapid development of computer network technology, the network equipment connected to the Internet has become diversified, and wireless network has become a very important network access mode in contemporary society with its advantages of mobility and flexibility. In today's high information technology, it is becoming more and more important to ensure the security of wireless network transmission.

Traditional network security technology research focuses on the application layer [1-4], and can solve daily network problems. However, there are no good solutions to the deep-seated network problems such as network supervision, differentiated services and information countermeasures. Obtaining the data information of network traffic can bring new ideas to solve these problems to a certain extent, and the feature extraction of network data is the core of such problems [5]. Ermanet[6] and others proposed to use unidirectional statistical features to promote protocol identification in the network core, and put forward several supervised classification methods for real-time protocols that only use the first few packet. When the classification accuracy of protocol labels is high, it can achieve efficient protocol identification, but it can do nothing for protocols without labels. McGregor[7] et al. put forward that the protocol is divided into several clusters, and the application clusters of each cluster are manually marked by EM algorithm. Hei[8] proposed CFI(Combined frequency Items) algorithm to solve the problem that the number of elements in the feature candidate set increases exponentially with the increase of time and the length of frequent items in AC(Aho-Corasick) algorithm. This algorithm is a combination and optimization method of AC algorithm and Apriori...
algorithm, which uses AC algorithm to generate frequent byte items and matches frequent items according to Apriori algorithm to obtain protocol features. Wang Yong et al. [9] combined automata with association rules, and proposed to use automatic state machine to display the transition process of protocol state. This method uses protocol creation language automata to represent protocol keywords, which has high flexibility and accuracy, but consumes too much resources. At present, these researches are aimed at wired networks, but for wireless networks, application layer protocols are difficult to capture.

For the research on wireless communication protocol identification, there are relatively few published materials at home and abroad. There is a big problem with the currently used identification methods: the subject of research stays at the signal level and attempts to perform protocol identification by collecting some parameters of the wireless signal or extracting a certain item of data in the protocol. This will result in a single extraction feature, large signal parameter recognition errors in the early stage, and low protocol recognition accuracy. Wen Aixia [10] proposed a frame feature selection algorithm based on improved mutual information to extract features from bitstream unknown protocol frames. The method is effective but the feature contains relatively single information. Zheng Jie [11] clusters the original data, then analyzes different clusters, and finds the address field as the protocol feature, but it also has the problem of single feature. For example, modulation recognition based on single feature extraction, blind recognition of convolutional codes based on Walsh_Hadmard transform, modulation recognition based on second-order cyclic statistics, and modulation recognition using linear classifier [12-14], etc.

In response to the above problems, this paper proposes a recognition method for feature selection of wireless protocol data from the data link layer. The method uses PIFE (Protocol Initial Feature Extraction, PIFE), an initial feature set extraction method that improves the AC algorithm and the association rule algorithm. On this basis, feature selection is applied to protocol recognition, further feature selection is performed on the initial rough feature set, and refined feature information is extracted. This algorithm is an unsupervised feature selection algorithm, which can effectively improve the accuracy of the feature set, thereby improving the accuracy of protocol recognition.

2. Protocol feature selection
In the actual communication process, there are some fixed fields in the intra-frame format of communication data, such as address field, identification field, control field and other identification bits. These identification bits have certain relevance in statistical data, and in the actual communication process, the fixed sequence content within the frame will remain unchanged for a period of time. In other words, these markers will show a certain frequency over a period of time. Therefore, it can be used as a feature to perform feature extraction on the bit stream, and further unknown protocol identification can be completed.

The process of feature selection of unknown wireless protocol based on mutual information is shown in figure 1:

![Figure 1. Protocol feature selection process.](image)

2.1. Initial feature set extraction method based on improved AC algorithm and association rule algorithm
Before the feature extraction of the unknown protocol, the intercepted bitstream data is preprocessed by statistical methods. In the statistical process, affected by the huge difference in the length of the protocol fixed sequence and the consumption of memory and time, it is necessary to select an appropriate sequence length for processing, and then perform splicing processing through association rules.
When using multi-pattern matching algorithm to extract frequent sequences, if the extracted sequence length is too long, it will occupy a lot of space in the process of building dictionary tree. Therefore, it is necessary to analyze the known protocol format, determine the appropriate extraction length $n$, and then establish a full binary tree with depth $n$ as dictionary tree, and define each node as a state.

After the dictionary tree is constructed, save the status information of each leaf node in the form of an array (appearance times $t$, appearance position $l$) to save space. Because the unknown protocol data in the form of bit stream only contains two elements "0" and "1", and the dictionary tree is a full binary tree, there is no mismatch in the matching process. Because the traditional AC algorithm is not applicable because of its time efficiency defect, it is necessary to improve the matching speed of AC algorithm. The improved AC algorithm abandons the state mismatch transfer function, establishes the jump function and adds auxiliary bits to count the occurrence times of each state in the bit stream data, so that the occurrence times and positions of the data sequence can be counted only by one scan. The calculation formula of the state jump function is as follows:

$$\text{new\_station} = \text{station} \times 2 + 2^{\text{new\_bit}}$$  \hspace{1cm} (1)

Where $\text{new\_bit}$ is the value of the next bit to be read, $\text{station}$ is the current node, and $\text{new\_station}$ is the next node to be jumped.

In the algorithm, the calculation formula of the threshold $L$ for measuring whether the frequent requirements are met is as follows:

$$L = \text{count}_{\text{arg}} \times \lambda = (m - n + 1) \times (2^n)^{-1} \times \lambda$$  \hspace{1cm} (2)

In which $\text{count}_{\text{arg}}$ is the expected number of times of sequence $T$ to be counted in the bitstream, and $m - n + 1$ represents the number of sequences with length $n$ in the bitstream with length $m$. $\lambda$ is the weight factor, which can be set manually after analyzing the known protocol. This setting is to control data redundancy, that is, to control the number of pseudo frequent sequences, while ensuring that all feature sequences are extracted as much as possible.

After obtaining short frequent sequences by dictionary tree statistics, it is necessary to further analyze their internal relations and splice short frequent sequences into long frequent sequences to obtain the initial feature set. Two frequent fields exist before and after the location or contain the relationship, and calculate the support degree $\text{sup}(f)$ of the frequent fields after connection:

$$\text{sup}(f) = \frac{H(f)}{\min(H(f_i), H(f_j))}$$  \hspace{1cm} (3)

where $f_i$ and $f_j$ are two fields that can be connected, $f$ is the long frequent field after $f_i$ and $f_j$ are connected, and $H(f)$ is the frequency of $f$ field. If the two fields meet the splicing conditions, the spliced new fields replace the original two fields, and the frequent itemsets are updated to obtain the initial feature set $F$.

The flow of obtaining the initial feature set $F$ is shown in the following figure 2:
Figure 2. Initial feature set acquisition process.

Since different feature fields have a front-back relationship in the intra-frame position, according to this relationship, the occurrence probability of feature fields based on a certain position difference is mined, and this occurrence probability is stored in the feature matrix $M$. The feature fields in the initial feature set $F$ are arranged according to the position in the frame from small to large to avoid invalid calculation in the process of feature extraction. $F_i$ represents the $i$-th feature in the initial feature set.

$$M = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1k} & \cdots & P_{18} \\ P_{21} & P_{22} & \cdots & P_{2k} & \cdots & P_{28} \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ P_{k1} & P_{k2} & \cdots & P_{kk} & \cdots & P_{k8} \end{bmatrix}$$

$P_{ij}$ represents the occurrence probability of the $i$-th feature and the $j$-th feature in the initial feature set based on the position difference $D$, and $k$ is the number of features in the initial feature set.

The occurrence probability of different feature fields in the same frame based on the position difference $D$ is analyzed. When analyzing the relation of occurrence probability between feature field $X$ and feature field $Y$, only calculate the occurrence probability of feature field $X$ and feature field $Y$ in the same frame at the distance $D$, immediately $dis(X), dis(Y) \in frame$; Furthermore, when the first appearance position $dis(X) > dis(Y)$ of the feature field $X$ and the feature field $Y$ occurs, $Y$ is determined to be the leading field, and only when $dis(X) > dis(Y)$ occurs, and the occurrence probability between the feature fields $X$ and $Y$ based on the position difference $D$ is calculated and analyzed only at certain times.

The occurrence probability is calculated by formula (4):

$$conf = \frac{P(dis(X) - dis(Y) = D)}{\min(P(X), P(Y))} > p_{\min}$$

s.t. $X, Y \in frame_i$

$X \in F_i, Y \in F_j$

$dis(X) > dis(Y)$

(4)

In the above formula, $conf$ represents confidence; $X, Y$ are two different characteristic fields; $dis(X), dis(Y)$ represent the appearance position of the feature field; $p_{\min}$ is the validity determination
threshold. When the confidence degree $\text{conf} > p_{\text{min}}$, the correlation is effective, which shows that the probability of occurrence of feature fields $X$ and $Y$ is $P(\text{dis}(X) - \text{dis}(Y) = D)$ when the position difference is $D$; Otherwise, when $\text{dis}(X) \leq \text{dis}(Y)$, set $P(\text{dis}(X) - \text{dis}(Y) = D) = -1$, and the occurrence probability is invalid. After obtaining the effective probability relationship, the features in the feature matrix will be screened by the feature selection algorithm based on improved mutual information, so as to further remove redundant features and improve the accuracy of feature extraction, and the output results can be used as the features of unknown protocols.

2.2. Frame feature selection method based on improved mutual information

Feature selection is based on the initial feature set obtained in the feature extraction stage, using the mutual information correlation theory in information theory, and analyzing the correlation and redundancy of features according to the idea of normalization. The main purpose is to further screen and verify the preliminarily extracted features. The frame feature selection algorithm based on bit stream mainly aims at the identification of unknown wireless protocols, and adopts unsupervised concept. Based on the feature selection algorithm based on mutual information proposed by Battiti [15], the adaptability and advancement are improved in combination with the characteristics of wireless protocol data, and an improved mutual information frame feature selection algorithm is proposed.

Firstly, considering that there is no clear target variable for unknown protocol identification, the algorithm takes the correlation and redundancy between the whole features as the feature selection and evaluation criteria, and gives the following definition based on this:

**Definition 1 Correlation degree:** Expressed by the average mutual information of feature $f_i$ and the entire feature set, as shown in formula (5):

$$\text{Re}l(f_i) = \frac{1}{n} \sum_{j=1}^{n} I(f_i, f_j) = \frac{1}{n} \left( H(f_i) + \sum_{1 \leq j \leq n, j \neq i} I(f_i, f_j) \right)$$  \hspace{1cm} (5)

In the above formula, $H(f_i)$ represents the entropy of the feature $f_i$, and represents the amount of information that the feature $f_i$ has. The larger the value of $H(f_i)$, the more favorable it is for clustering. $\sum_{1 \leq j \leq n, j \neq i} I(f_i, f_j)$ represents the mutual information between feature $f_i$ and other features, and the larger the value, the smaller the contribution value of other features to the subsequent clustering algorithm.

**Definition 2 Conditional correlation:** the calculation of conditional correlation of a selected feature $f_g$ to an unselected feature $f_i$ is expressed by formula (6):

$$\text{Re}l(f_g \mid f_i) = \frac{H(f_g \mid f_i)}{H(f_g)} \cdot \text{Re}l(f_g)$$ \hspace{1cm} (6)

In the above formula, $\text{Re}l(f_g \mid f_i)$ represents the mutual information between feature $f_g$ and the whole feature set, and $H(f_g)$ represents the entropy of features $f_g$, that is, the amount of information contained.

**Definition 3 Redundancy:** Redundancy is one of the important indexes to evaluate the relationship between features in the initial feature set. In terms of quantification, redundancy can be expressed by the difference between the above-mentioned correlation and conditional correlation, and its calculation formula is shown in (7):

$$\text{Re}d(f_i, f_g) = \text{Re}l(f_i) - \text{Re}l(f_g \mid f_i)$$  \hspace{1cm} (7)

Secondly, Battiti proposed a feature selection algorithm called mutual information (MIFS), which
not only needs class target variables $c$ as prior knowledge, but also needs to adjust $\beta$ parameters through experience. The MIFS algorithm is expressed by the formula (8):

$$I(f_i, c) - \beta \sum_{f_j \in S} I(f_i, f_j)$$  \hspace{1cm} (8)

The self-defined parameter $\beta$ in the algorithm is usually related to the redundancy between the selected feature and the candidate feature. If $\beta$ is too large, the related features will be eliminated excessively, while if $\beta$ is too small, redundant features will be selected excessively. Therefore, the value of $\beta$ should be considered comprehensively in the trade-off between correlation and redundancy. To solve this problem, Hanchuan and Fuhui [16] proposed MRMR algorithm according to the maximum correlation and minimum redundancy criteria. The MRMR algorithm can be expressed as:

$$I(f_i, c) - \frac{1}{|S|} \sum_{f_j \in S} I(f_i, f_j)$$  \hspace{1cm} (9)

MRMR eliminates user-defined parameters, but because the initial feature set may contain multiple protocols, that is, multi-category feature selection, only eliminating user-defined feature parameters cannot completely solve the problem. When multi-sample data is included, both sides of the minus sign cannot be guaranteed to remain within the range of $[0,1]$. Considering that the intercepted bit stream may contain multiple communication protocols, this paper proposes to limit both sides of the minus sign by normalization.

According to the definition of formula (5), it can be seen that $\text{Rel}(f_i)$ considers the mutual information value between the whole, and its value range is shown in formula (10):

$$0 \leq \text{Rel}(f_i) \leq \frac{1}{|S|} \sum_{f_j \in F} H(f_j)$$  \hspace{1cm} (10)

According to formula (5) and formula (6), we can know:

$$\text{Red}(f_i, f_g) \leq \frac{\text{Rel}(f_i) + \text{Rel}(f_g)}{2}$$  \hspace{1cm} (11)

Based on Formula (9) and Formula (11), we adapt and normalize the MIFS selection formula as follows:

$$\text{WNMIFS} = \sum_{f_j \in F} \frac{|S|}{H(f_j)} \text{Rel}(f_j) - \sum_{f_j \in F, f_k \in U} \frac{2 \ast \text{Red}(f_i, f_g) \ast \text{Rel}(f_i) + \text{Rel}(f_g)}{\text{Rel}(f_i) + \text{Rel}(f_g) \ast (|S|)^{-1}}$$  \hspace{1cm} (12)

Where $F$ represents the initial feature set; $|S|$ represents the number of features that do not belong to the target set; $H(f_j)$ represents the entropy of feature $f_j$, that is, the amount of information contained; $\text{Rel}(f_i)$ represents the mutual information between feature $f_i$ and the whole feature set; $U$ represents the selected feature set; $\text{Red}(f_i, f_g)$ represents the redundancy between the unselected feature $f_i$ and the selected feature $f_g$. By adapting and normalizing the MIFS selection formula, the feature with maximum correlation and minimum redundancy can be selected in the feature selection process.

According to the above analysis, the wireless protocol frame feature selection algorithm based on improved mutual information is as follows:

| Input: initial feature set $F$ |
|--------------------------------|
| Output: target collection $S$ |
Step1: Initialize $F$ as an initial feature set containing all features ($n$), $S$ is an empty set;

Step2: $\forall f_i \in F$; Calculate the average mutual information of $f_i$, that is, correlation $r$;

Step3: Find the largest $\text{Rel}(f_i)$ and move it from $F$ set to $S$ set;

Step4: Calculate the \textit{WNMIFS} of the remaining features, and select the feature with the maximum value to join the feature set;

1. For all feature pairs $(f_i, f_j)$, satisfies $f_i \in F, f_j \in S$ and $I(f_i, f_j)$ is calculated
2. Select the feature with the largest mutual information value as the next feature

\[ \text{WNMIFS} = \sum_{f_i \in F} |S| H(f_i) \text{Rel}(f_i) - \frac{1}{|S|} \sum_{f_i \in F, f_j \in S} 2 \times \text{dRe}(f_i, f_j) \times \text{Rel}(f_i) + \text{Rel}(f_j) \]

$setF \leftarrow F \setminus \{f_i\}; S \leftarrow \{f_i\}$

Step5: Judging whether \textit{WNMIFS} reaches the standard or not, moving the feature from the original set to the target set if it reaches the standard, and returning to Step4, otherwise executing Step 6;

Step6: $S$ is the selected feature, and the result is output.

The frame feature selection algorithm based on bitstream is to select the final feature after obtaining the initial feature set, and its algorithm flowchart is show in figure 3:

![Feature selection flow chart](image)

Figure 3. Feature selection flow chart.

3. Simulation verification

3.1. Verification of feature selection

The accuracy, false recognition rate and time complexity of feature selection are evaluated by
simulation. The definition formula of accuracy is as follows:

$$R = \frac{\text{Count}(\text{True feature})}{\text{Count}(\text{allTrue feature})} \times 100\% \tag{13}$$

$R$ represents the proportion of correctly selected protocol features to total protocol features, $\text{Count}(\text{True feature})$ represents correctly selected features, and $\text{Count}(\text{allTrue feature})$ represents total correct features.

$$RP = \frac{\text{Count}(\text{False feature})}{\text{Count}(\text{all feature})} \times 100\% \tag{14}$$

The misidentification rate $RP$ represents the proportion of erroneous features to the total protocol features during feature selection, $\text{Count}(\text{False feature})$ represents the wrongly selected features, and $\text{Count}(\text{all feature})$ represents the total number of features in the initial feature set.

In order to verify the recognition rate of the algorithm, HDLC, Ethernet and 802.11 are used as simulation experimental data. The accuracy of the algorithm proposed in this paper will be tested under different data redundancy levels, and the experimental results will be compared with the protocol formats of known protocols. The accuracy and misidentification rate of feature selection are shown in table 1 and table 2 respectively. Experimental results show that the proposed algorithm has higher accuracy and lower false recognition rate.

| Table 1. Recognition accuracy. | Recognition accuracy P |
|--------------------------------|-------------------------|
| data redundancy               | HDLC        | Ethernet | 802.11 |
| 12.5%                         | 100%        | 100%     | 100%   |
| 25%                           | 100%        | 100%     | 88.9%  |
| 37.5%                         | 90%         | 88.9%    | 90%    |
| 50%                           | 88.9%       | 91.6%    | 91.6%  |
| 62.5%                         | 84.6%       | 83.6%    | 92.3%  |
| 75%                           | 84.2%       | 83.3%    | 85.7%  |
| 87.5%                         | 83.3%       | 82.8%    | 81.2%  |
| 100%                          | 83.3%       | 82.1%    | 87.5%  |

| Table 2. Misrecognition Rate. | Misrecognition rate RP |
|------------------------------|-------------------------|
| data redundancy              | HDLC        | Ethernet | 802.11 |
| 12.5%                        | 0.00%       | 0.00%    | 0.00%  |
| 25%                          | 0.00%       | 0.00%    | 8.88%  |
| 37.5%                        | 7.27%       | 8.07%    | 7.27%  |
| 50%                          | 7.40%       | 5.60%    | 5.60%  |
| 62.5%                        | 9.48%       | 10.09%   | 4.74%  |
| 75%                          | 9.03%       | 9.54%    | 8.17%  |
| 87.5%                        | 8.91%       | 9.17%    | 10.03% |
| 100%                         | 8.35%       | 8.95%    | 6.25%  |

The core of the wireless protocol frame feature selection algorithm based on improved mutual information is to consider the redundancy effect between different protocol features when the
extracted features include multiple protocol features. In order to verify the improvement effect, the simulation verified the accuracy comparison of the algorithm in this paper, the NMIFS algorithm and the LIS algorithm. The test results are as follows:

It can be seen from figure 4 that with the increase of data redundancy, the accuracy of feature selection of WNMIFS algorithm proposed in this paper is higher than that of existing algorithms, and the decrease of accuracy is less affected by the increase of redundancy than that of existing algorithms.

On the basis of calculating the correlation, the algorithm in this paper further calculates the redundancy between data, and the increase of time complexity is inevitable. The data based on protocol features is not huge. Based on careful analysis of experimental calculation formulas, this algorithm sacrifices some storage space to save the intermediate results of calculation. The growth rate of time consumption of the algorithm is shown in figure 5. From the simulation results, the time consumption of the proposed algorithm is not obviously increased compared with that of the original algorithm.

3.2. Verification of initial feature set extraction
Experiments evaluate the quality of initial feature set extraction by simulating the time complexity of initial feature set extraction, as shown in figure 6.

It can be seen from the graph that the time complexity of the initial feature set extraction method proposed in this paper is less than that of the existing traditional methods.

4. Conclusion
Based on the booming development of wireless access equipment at this stage, the rapid development
of wireless communication technology has led to the continuous increase of wireless communication protocols, which has brought new challenges to network security. However, the traditional protocol identification methods are concentrated on the network application layer, and the existing wireless protocol research is concentrated on the signal layer, which cannot effectively deal with the existing network security challenges. This paper uses the idea of improved AC cutting algorithm and association rules to segment and splice the data link layer protocol to extract the initial feature set to reduce the algorithm time consumption and improve the algorithm efficiency. In feature selection, this paper adopts mutual information correlation theory, combines the information correlation degree of frame features, redundancy correlation definition and normalization ideas, and further screens and verifies to extract the features of unknown protocol data, and realizes the data link layer and application More in-depth analysis of layer data. The experimental data proves that the method can effectively screen the feature set and obtain a more reasonable feature set. At the same time, compared with the original algorithm, the time growth rate of the algorithm is kept at a low level, and it has a certain degree of advancement and reliability. The next step is to further analyze the protocol features obtained through this article, accurately extract the protocol format features based on a large number of data frames and protocol features, and further analyze the semantics contained in the features to achieve the purpose of identifying unknown protocols.

References:

[1] Wang, Y., Li, X., Meng, J. (2011) Biprominer: Automatic mining of binary protocol features. In: Parallel and Distributed Computing, Applications and Technologies (PDCAT). Gwangju. pp. 179-184.

[2] Sen, S., Spatscheck, O., Wang, D. (2004) Accurate, Scalable in-network Identification of P2P Traffic Using Application Signatures. In: Proceedings of the 13th international conference on World Wide Web. New York. pp. 512-521.

[3] Zander, S., Nguyen, T., Armitage, G. (2005) Self-Learning IP traffic classification based on statistical flow characteristics. In: International Conference on Passive and Active Network Measurement. Boston. pp. 325-328.

[4] Yin, S.S., Yang, M., Song, J. (2012) General Application Layer Protocol Analysis Technology Based on Regular Expressions. Information Technology., 02: 99-101.

[5] Cai, X., Zhang, R., Wang, B. (2010) Machine Learning and keyword-matching integrated Protocol Identification. In: IEEE International Conference on Broadband Network and Multimedia Technology. Beijing. pp. 164-169.

[6] Krishna, R.S.B., Aramudhan, M. (2014) Feature selection based on information theory for pattern classification. In: International Conference on Control, Instrumentation, Communication and Computational Technologies. Institute of Electrical and Electronics Engineers Inc. pp. 1233-1236.

[7] Nhaila, H., Merzouqi, M. (2015) Hyperspectral images classification and Dimensionality Reduction using Homogeneity feature and mutual information. In: Intelligent Systems and Computer Vision. Morocco. pp. 1-5.

[8] Hei, X.H., Bai, B.B., Wang, Y.C. (2019) Feature Extraction Optimization for Bitstream Communication Protocol Format Reverse Analysis. In: Trust Security and Privacy in Computing and Communications. Rotorua. pp. 662-669.

[9] Wang, Y., Zhang, N., Wu, Y. (2013) Protocol Specification Inference Based on Keywords Identification. In: Advanced Data Mining and Applications. Zhejiang. pp. 443-454.

[10] Wen, A.X. (2015) Research on Discovery Technology of Unknown Protocol Features in Bitstream Data.https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201601&filename=1015710383.nh.

[11] Zheng, J., Li, J.P. (2017) Feature identification of unknown protocol. In: 13th International Computer Conference on Wavelet Active Media Technology and Information Processing. Chengdu. pp. 147-149.
[12] Xin, Z., Ying, W.U. (2011) Signal modulation recognition based on KPCA and LDA. Systems Engineering & Electronics., 33(7): 1611-1616.
[13] Jian, L. (2019) Blind Reconstruction of Convolutional Code Based on Partitioned Walsh-Hadamard Transform. Journal of Electronics and Information Technology., 41: 2047-2054.
[14] Moore, A., Zuev, D., Crogan, M. (2013) Discriminators for use in flow-based classification. https://www.doc88.com/p-7405021573624.html?s=rel&id=1.
[15] Battiti, R. (1994) Using mutual information for selecting features in supervised neural net learning. Neural Networks IEEE Transactions on., 5(4): 537-550.
[16] Peng, H.C., Long, F.H. (2005) Feature Selection Based on Mutual Information. In: Criterion of Max Dependency, Max-Relevance and Min-Redundancy IEEE Transactions of Pattern Analysis and Machine Intelligence. pp. 1226-1238.